MDPI

*Article*

# Research on Coupling Knowledge Embedding and Data-Driven Deep Learning Models for Runoff Prediction

Yanling Li, Junfang Wei *, Qianxing Sun and Chunyan Huang

School of Mathematics and Statistics, North China University of Water Resources and Electric Power, Zhengzhou 450046, China; liyanling@ncwu.edu.cn (Y.L.); x202210080830@stu.ncwu.edu.cn (Q.S.); huangchunyan@ncwu.edu.cn (C.H.)
* Correspondence: weijunfang@stu.ncwu.edu.cn

**Abstract:** Accurate runoff prediction is crucial for watershed water resource management, flood prevention, and hydropower station scheduling. Data-driven models have been increasingly applied to runoff prediction tasks and have achieved impressive results. However, existing data-driven methods may produce unreasonable predictions due to the lack of prior knowledge guidance. This study proposes a multivariate runoff prediction model that couples knowledge embedding with data-driven approaches, integrating information contained in runoff probability distributions as constraints into the data-driven model and optimizing the existing loss function with prior probability density functions (PDFs). Using the main stream in the Yellow River Basin with nine hydrological stations as an example, we selected runoff feature factors using the transfer entropy method, chose a temporal convolutional network (TCN) as the data-driven model, and optimized model parameters with the IPSO algorithm, studying univariate input models (TCN-UID), multivariable input models (TCN-MID), and the coupling model. The results indicate the following: (1) Among numerous influencing factors, precipitation, sunshine duration, and relative humidity are the key feature factors driving runoff occurrence; (2) the coupling model can effectively fit the extremes of runoff sequences, improving prediction accuracy in the training set by 6.9% and 4.7% compared to TCN-UID and TCN-MID, respectively, and by 5.7% and 2.8% in the test set. The coupling model established through knowledge embedding not only retains the advantages of data-driven models but also effectively addresses the poor prediction performance of data-driven models at extremes, thereby enhancing the accuracy of runoff predictions.

**Keywords:** knowledge embedding; data driven; transfer entropy; loss function; IPSO algorithm; runoff prediction

## 1. Introduction

Against the backdrop of global warming, the Yellow River Basin is facing a series of extreme climate issues, including floods and droughts [1]. River runoff is a key component of the hydrological cycle and has undergone substantial changes due to human activities [2]. The Lvovich method, proposed by Soviet hydrologist Mikhail Lvovich, integrates precipitation, evaporation, runoff, and groundwater to establish a water balance model. This model simulates water volume changes under different conditions, analyzing the spatiotemporal distribution characteristics of water resources. Using the Lvovich method for analysis, results indicate that the annual base flow at Changshui Station showed a highly significant decreasing trend from 2000 to 2019, while Longmen Town Station and Baima Temple Station exhibited a noticeable increasing trend in the spring [3]. Over the past 60 years, the number of tributaries in the Serbian Black Sea Basin has increased, while the number of tributaries in the Adriatic Sea Basin has decreased by 5.5%, and in the Aegean Sea Basin by 8.8% [4]. With the intensification of climate warming and human activities, global runoff, particularly in arid regions, has significantly decreased [5]. This has had

a severe impact on the development and utilization of water resources as well as on the ecological environment [6]. Water scarcity has become a major challenge faced by many countries. It is estimated that by 2030, global water demand will reach 160% of the total available water resources [7]. By 2050, nearly 6 billion people will face water scarcity [5]. Human activities have altered the spatiotemporal distribution of water resources, resulting in complex ecological and socioeconomic consequences [2]. Global warming has led to reduced precipitation, increased evaporation, and intensified drought in the Yellow River Basin [8]. Research shows that since the 1980s, actual runoff in the Yellow River Basin has continuously declined due to surface acidification and a general decrease in basin precipitation [9]. Reduced runoff can lead to a series of environmental issues, such as the degradation of downstream aquatic ecosystems, increased flood threats and urban waterlogging, and shortages of water for domestic, industrial, and agricultural use [10]. However, the complex nonlinear and non-stationary characteristics of runoff time series often result in traditional linear time-series models providing unsatisfactory predictive accuracy. Additionally, the limited scope for improving model methods further exacerbates the challenge of runoff prediction. Given this, there is an urgent need to develop runoff prediction models that are both highly accurate and reliable to address these challenges and meet practical application needs.

Over the years, runoff prediction models have been mainly classified into two categories: process-driven models and data-driven models. Process-driven models are based on hydrological principles and rely on extensive hydro-meteorological data to simulate runoff processes and river channel evolution using mathematical models for hydrological forecasting [11]. Among them, commonly used process-driven models by scholars include the SWAT model, BASINS watershed modeling system [12], SWIM model [13], and others. The Xin'anjiang model is a popular conceptual hydrological model that has been widely used for rainfall-runoff simulation and prediction in China's humid and semi-humid regions. However, its application is limited in basins with limited data [14]. The TOPMODEL is a rainfall-runoff model that predicts the watershed's response to rainfall distribution by leveraging hydrological similarity between different points in the catchment. Compared to the Xin'anjiang model, TOPMODEL is equally efficient but requires calibration of fewer parameters, many of which have a physical basis. For flood simulation, TOPMODEL offers a simpler and more convenient parameter calibration process than the Xin'anjiang model [15]. Cai et al. [16] made appropriate modifications to the SWAT runoff module and calibrated key sensitive parameters. The improved SWAT model was able to better simulate monthly runoff variations. Introducing a dynamic river network method into the Distance Dynamic Model (DDD) can effectively enhance the model's accuracy in predicting flood peaks [17].

In runoff prediction models, data-driven approaches have become increasingly popular as they focus on capturing the nonlinear relationships between input and output parameters without considering the entire physical process [18]. These include models, such as Long Short-Term Memory (LSTM) [19], convolutional neural networks (CNNs) [20], Gated Recurrent Units (GRUs) [21], Support Vector Machines (SVMs) [22], Extreme Gradient Boosting (XGB) [23], and Bidirectional Long Short-Term Memory (BiLSTM) [24]. Guo et al. [25] combined physical mechanism models with Long Short-Term Memory (LSTM) networks and proposed 16 different runoff prediction model combination strategies. This approach effectively enhanced the daily runoff prediction performance of individual LSTM models. Yao et al. [26] proposed a hybrid model based on CNN-LSTM and GRU-ISSA to predict runoff volumes using historical meteorological and runoff data. Using the Bailongjiang Basin as a case study, the model achieved an RMSE of 2.17, significantly outperforming other benchmark models. Liu et al. [27] proposed a Least Squares Support Vector Machine (LS-SVM) model that corrects initial errors based on monthly runoff data from the Fuyu hydrological station. This model partially corrected the overestimation of actual runoff predictions. Chen et al. [28] proposed a Tree-State Long Short-Term Memory (treeLSTM) multilayer spatiotemporal model based on daily runoff data from the Hanjiang

Basin. This model integrates the temporal dependencies of historical hydrological data with the spatial correlations of hydrological variables, enhancing the physical interpretability of machine learning algorithms.

However, data-driven models heavily rely on training data and do not incorporate decades of accumulated empirical knowledge. Moreover, these models lack "common sense" or an understanding of the physical mechanisms in the real world, making them prone to generating unreasonable or unrealistic predictions. For example, Dong et al. [29] found that convolutional neural networks do not truly detect semantic objects (for instance, the model fails to learn the concept of birds when identifying them in images). Consequently, for certain artificially generated adversarial samples, such as specifically designed beer bottle images, despite their completely different visual appearance, neural networks may mistakenly classify them as birds. Similarly, in subsurface flow problems, groundwater and related data are used as input variables to establish a mapping relationship with neural networks. Based on this relationship and data, future groundwater volumes can be predicted. However, previous studies have shown that traditional data-driven neural networks perform poorly with noisy observations, which is inevitable in practice. Additionally, due to the absence of these time steps in the training data, the models cannot accurately predict future flow fields [30].

Despite the rich research achievements in previous studies, there are still many limitations in runoff prediction research. Process-driven models have too many parameters and require extensive hydrological, meteorological, and potential surface conditions. Due to the high cost of data collection, the applicability of these models is limited [31]. Data-driven models also require large amounts of data, and without sufficient data, the reliability of these models can be severely compromised [30]. Furthermore, data-driven models rely solely on large amounts of data and do not incorporate domain knowledge in their construction, which may result in unreasonable predictions. Consequently, data-driven models can be significantly affected by data noise, leading to completely incorrect results [32].

To overcome the aforementioned limitations, incorporating scientific knowledge or practical experience into deep learning models has become an emerging paradigm for many scientific problems. For example, Karpatne et al. [33] proposed a physical-guided neural network (PGNN) model, which incorporates physics-based loss into the learning objective function of the neural network to achieve scientifically consistent results. This method was then applied to the problem of lake temperature simulation. He et al. [34] proposed a theory-guided fully convolutional neural network (TgFCNN) model to solve the inverse problem of subsurface contaminant transport. TgFCNN can construct robust and reliable surrogate models with limited training and further be used for inverse modeling tasks, achieving good accuracy in estimating unknown contaminant source parameters and permeability fields. Chen et al. [35] developed a hard-constrained model within a theory-guided framework to ensure that the model outputs adhere to known governing equations. In 2019, Raissi [36] designed a method called Physics-Informed Neural Networks (PINNs), which incorporate nonlinear partial differential equations (PDEs) as regularization terms in the loss function. Wang et al. [30] proposed the Theory-Guided Neural Network (TgNN), which uses subsurface flow governing equations as domain knowledge to guide neural network predictions. Unlike purely data-driven models that can only interpolate within the range of training data, theory-guided models can extrapolate based on physical mechanisms, predicting data outside the range of training data and expanding the model's applicability.

To overcome the aforementioned limitations, this study proposes a runoff prediction model that couples knowledge embedding with data-driven approaches, aiming to establish a dual-driven model. This coupled model breaks the "black box" nature of neural networks, effectively addressing the poor prediction performance of data-driven models at runoff extremes, ensuring that runoff predictions conform to the probability density function of runoff, thereby improving model accuracy.

This study contributes from three main aspects:

(1) Using the transfer entropy method to select the characteristic factors of runoff and incorporating them as input variables ensure the reliability of the model's input variables.

(2) Using the Improved Particle Swarm Optimization (IPSO) algorithm for model parameter optimization to enhance computational efficiency of the model.

(3) Constructing a coupled knowledge-embedding and data-driven runoff prediction model, the performance of the coupled model improved by 5.7% and 2.8% on the training and testing sets, respectively, compared to traditional data-driven models.

## 2. Research Methodology

### 2.1. Research Area

The Yellow River originates from the Qinghai-Tibet Plateau and flows through nine provinces and autonomous regions: Qinghai, Sichuan, Gansu, Ningxia, Inner Mongolia, Shanxi, Shaanxi, Henan, and Shandong. It has a total length of 5464 km and a basin area of $79.5 \times 10^4$ km$^2$, covering 8% of China's land area [37]. The Yellow River basin features diverse terrain and landscapes, with mountainous terrain dominating the upper and middle reaches and plains dominating the middle and lower reaches. The terrain slopes from west to east, forming three distinct terraces. The first terrace is the Qinghai-Tibet Plateau, with an average elevation of over 4000 m. The second terrace is primarily the Loess Plateau, with relatively flat terrain and elevations ranging from 1000 to 2000 m. The third terrace is mainly the North China Plain, characterized by flat terrain [38]. The main stem of the Yellow River includes nine hydrological stations, such as Tangnaihai. In recent years, there has been a trend of decreasing runoff at hydrological stations in the Yellow River Basin [39]. The precipitation characteristics of the Yellow River basin include sparse rainfall in spring, abundant rainfall in summer, and cold, dry winters. Evaporation within the basin is significant, and annual rainfall distribution is uneven. It is one of the most water-scarce regions in China, and variations in runoff directly affect the watershed's water supply security [40]. Studies have shown that over the past 60 years, annual runoff in each sub-basin of the Yellow River has exhibited a significant downward trend, with an increase in the frequency and severity of extreme drought events [41]. The ecological environment of the basin is extremely fragile and highly sensitive to climate change and human activities. In recent years, the over-exploitation of water resources has led to a dramatic reduction in surface runoff, a continuous decline in groundwater levels, and a deterioration in water quality, among other ecological and environmental issues [42]. Therefore, research on runoff prediction in the Yellow River basin has attracted considerable attention. Improving the accuracy of runoff prediction within the basin is crucial for agricultural decision making, water resource management, and disaster assessment [43]. The study area is shown in Figure 1.

### 2.2. Data

This study focuses on the Yellow River basin and conducts runoff prediction research based on hydrological and meteorological data. The hydrological data include monthly runoff (R) data from January 1964 to December 2023, spanning 60 years, obtained from nine main hydrological stations along the main stem of the Yellow River. Meteorological data consist of monthly data from January 1964 to December 2023 for nine influencing factors: rainfall (RF), atmospheric pressure (AP), wind velocity (WV), temperature (T), temperature anomaly (TA), vapor pressure (VP), hours of sunshine (HS), relative humidity (RH), and sunshine percentage (SP). The data are sourced from the Yellow River Conservancy Commission of the Ministry of Water Resources (http://www.yrcc.gov.cn/, accessed on 1 January 2024) and the National Meteorological Information Center (NMIC) at the China Meteorological Administration (CMA) (https://data.cma.cn/, accessed on 1 January 2024). The data are reliable and valid, as detailed in Table 1.
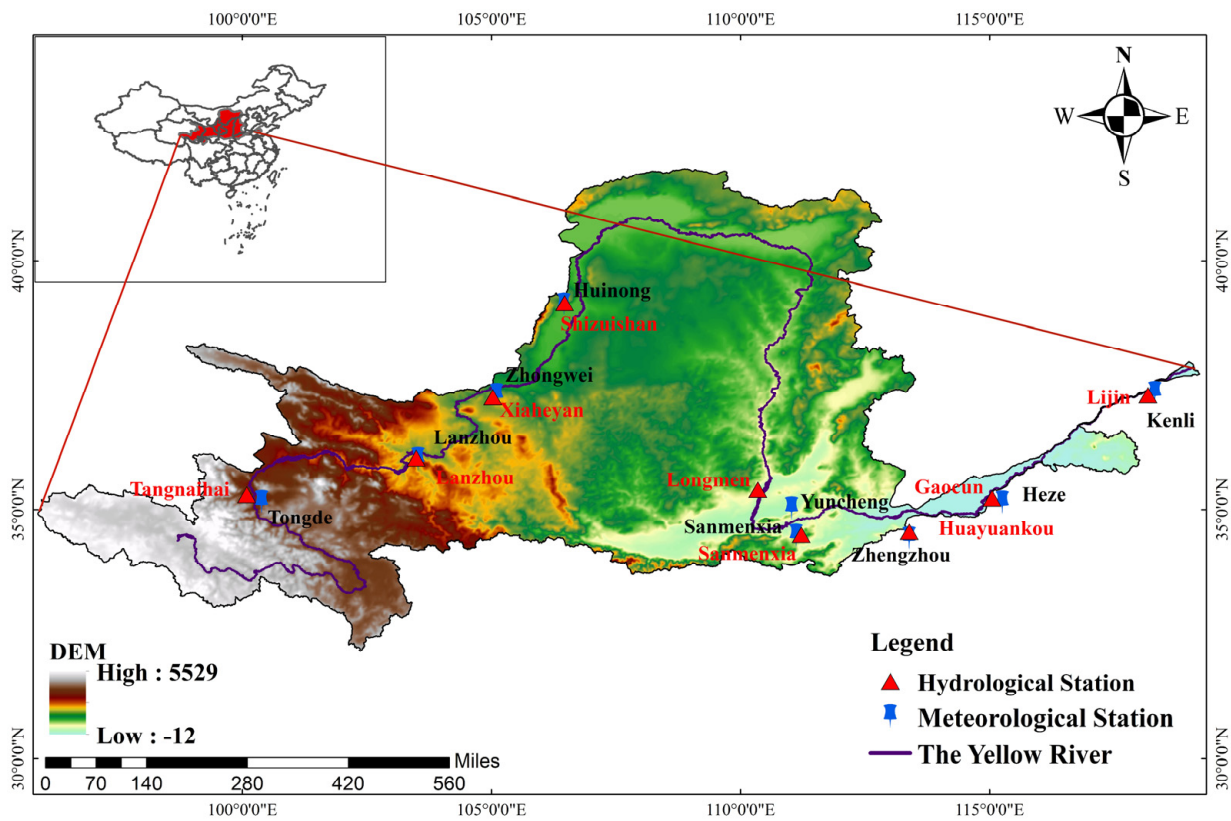
**Figure 1.** Study area. The red triangles represent the nine hydrological stations on the main stem of the Yellow River, while the blue markers indicate meteorological stations located spatially close to these hydrological stations.

**Table 1.** The datasets.

| Data | Data Type | Data Length | Data Source | Stations |
|---|---|---|---|---|
| Rainfall<br>Atmospheric pressure<br>Wind velocity<br>Temperature<br>Temperature anomaly<br>Vapor pressure<br>Hours of sunshine<br>Relative humidity<br>Sunshine percentage | Meteorological data | January 1964–December 2023 | The China Meteorological Data Service Centre (https://data.cma.cn/, accessed on 1 January 2024) | Tongde/Gaolan/Zhongwei/Huinong/Yuncheng/Sanmenxia/Zhengzhou/Heze/Kenli |
| Runoff | Hydrological data | January 1964–December 2023 | Yellow River Conservancy Commission of the Ministry of Water Resources (http://www.yrcc.gov.cn/, accessed on 1 January 2024) | Tangnaihai/Lanzhou/Xiaheyan/Shizuishan/Longmen/Sanmenxia/Huayuankou/Gaocun/Lijin |

## 3. Materials and Methods

The structure and flowchart of the model framework are shown in Figure 2. Sections 3.1 and 3.2, respectively, introduce the spatiotemporal characteristics of runoff, including standardized runoff index and cross-wavelet analysis. Sections 3.3–3.5 present transfer entropy, IPSO algorithm, and the coupled knowledge embedding with data-driven model, along with three metrics to evaluate the model's accuracy, reliability, and stability.
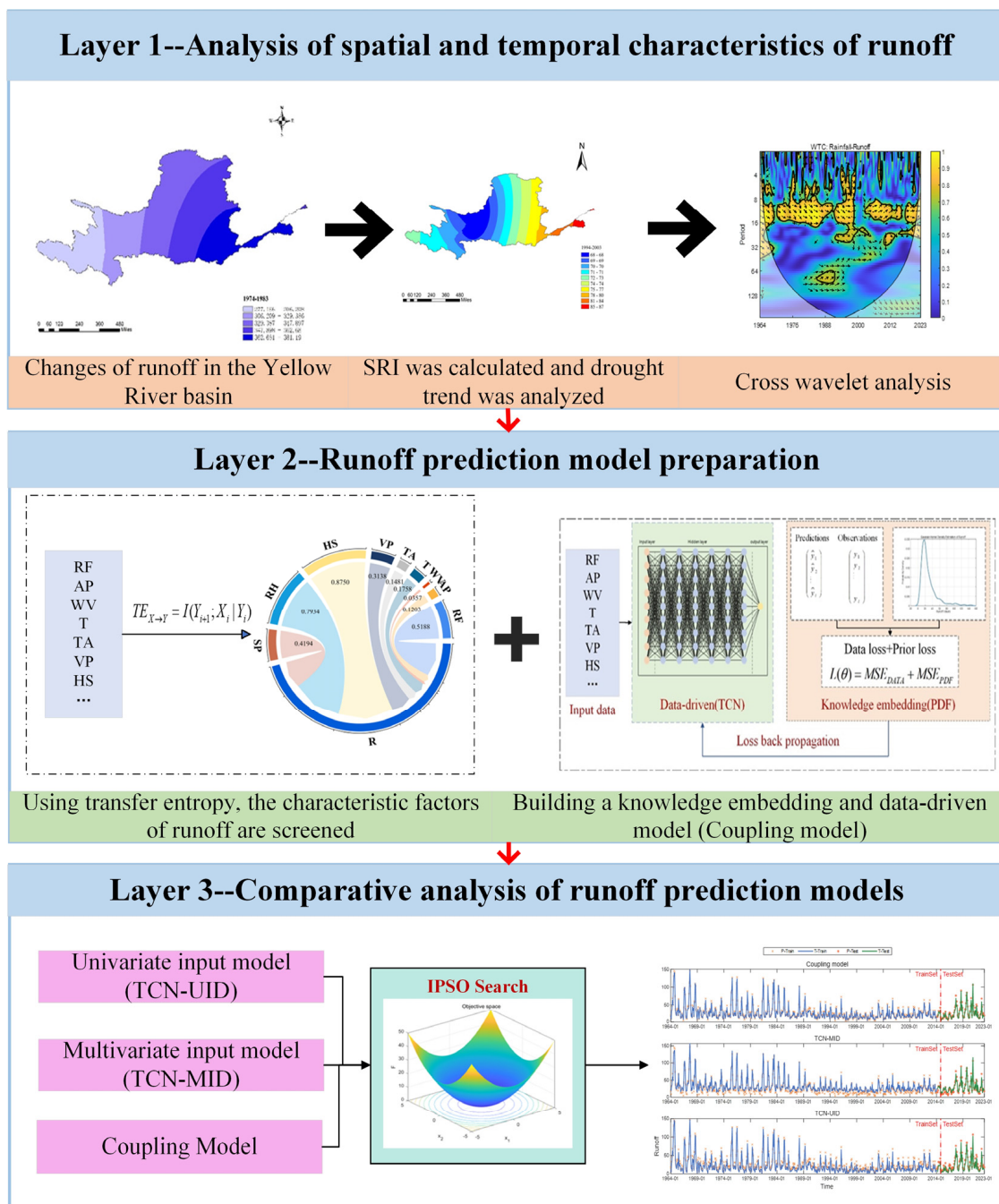
**Figure 2.** Flowchart of runoff prediction model structure.

*3.1. Calculation of the SRI*

The Standardized Precipitation Index (SPI), proposed by McKee in 1993 [44], describes natural precipitation using the Gamma function and normalizes it to quantify the severity of meteorological drought through precipitation probabilities. The calculation method for SPI is as follows:

(1) Let the precipitation amount be a random variable $x$ that follows a Gamma distribution, with $g(x)$ representing the probability density function.

$$g(x) = \frac{1}{\beta \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} (x > 0) \tag{1}$$

where $\alpha$ represents the shape factor, $\beta$ represents the scale factor, and $\Gamma(\alpha)$ represents the Gamma function. These parameters can be estimated using Maximum Likelihood Estimation (MLE):

$$\alpha = \frac{1 + \sqrt{1 + 4A/3}}{4A} \quad \beta = \frac{\bar{x}}{\alpha} \quad A = In(\bar{x}) - \frac{\sum\limits_{i=1}^{n} In(x_i)}{m} \tag{2}$$

where $\bar{x}$ represents the average precipitation amount, and $n$ denotes the length of the sequence.

(2) The probability density function of the precipitation amount $x$ is given by:

$$G(x) = \frac{1}{\beta^{\gamma}\Gamma(\gamma)} \int_{0}^{x} x^{\gamma-1} e^{-\frac{x}{\beta}} dx, x > 0 \tag{3}$$

When the precipitation amount is 0, the probability distribution is estimated as:

$$G(x = 0) = \frac{m}{n} \tag{4}$$

where $m$ represents the number of samples with zero precipitation, and $n$ denotes the total number of samples.

(3) By normalizing the Gamma distribution, the Standardized Precipitation Index (*SPI*) is obtained as follows:

$$SPI = S\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \tag{5}$$

where $t = \sqrt{\ln \frac{1}{G(x)^2}}$. When $G(x) > 0.5$, $G(x)$ is given by $1 - G(x)$, $S = 1$. When $G(x) \leq 0.5$, $S = -1$. $c_0 = 2.515517$, $c_1 = 0.802853$, $c_2 = 0.010328$, $d_1 = 1.432788$, $d_2 = 0.189269$, $d_3 = 0.001308$.

Since the calculation process and classification of the Standardized Runoff Index (SRI) are consistent with those of the Standardized Precipitation Index (SPI), the SRI can be obtained by replacing precipitation data with runoff data and describing drought severity using runoff probabilities. Based on the aforementioned equation, this study calculates the monthly SRI for each hydrological station and analyzes the hydrological drought conditions in the Yellow River Basin over the past 60 years, providing a data foundation for subsequent runoff prediction. The drought classification standards for SRI (Table 2) are established according to the "Meteorological Drought Classification" issued by the National Meteorological Center of the China Meteorological Administration.

**Table 2.** SRI-base drought classification.

| Type | Grade | SRI Value |
|:---:|:---:|:---:|
| 1 | No drought | $-0.5 < SRI$ |
| 2 | Light drought | $-1.0 < SRI \leq -0.5$ |
| 3 | Moderate drought | $-1.5 < SRI \leq -1.0$ |
| 4 | Severe drought | $-2.0 < SRI \leq -1.5$ |
| 5 | Extreme drought | $SRI \leq -2.0$ |

*3.2. Cross-Wavelet Transform (XWT)*

XWT decomposes time-series data into different components, analyzes the correlation between different time series based on continuous wavelet transform, and explores potential causes, reflecting the phase characteristics between sequences in both time and frequency

domains [45]. The cross-wavelet transform (XWT) of time series $x(t)$ and $y(t)$ is given by the following formula [46]:

$$W_{XY}(\alpha, \tau) = C_X(\alpha, \tau)C_Y^*(\alpha, \tau) \qquad (6)$$

where $C_X(\alpha, \tau)$ represents the wavelet transform coefficient of $x(t)$, and $C_Y^*(\alpha, \tau)$ represents the complex conjugate of the wavelet transform coefficient of $y(t)$.

Wavelet coherence (WTC) analyzes the dependence between two signals by examining the amplitude and phase information of time-series signals [46]. The calculation formula is as follows:

$$R^2(\alpha, \tau) = \frac{\left|S(\alpha^{-1}W_{XY}(\alpha, \tau)\right|^2}{S(\alpha^{-1}|W_X(\alpha, \tau)|^2) \cdot S(\alpha^{-1}|W_Y(\alpha, \tau|^2)} \qquad (7)$$

*3.3. Transfer Entropy Theory*

Transfer entropy (TE) reflects the degree of correlation between variables, measuring the causal relationship of information transfer in terms of magnitude and direction [40]. It is a method to measure causality in time series. The fundamental assumption of TE is that the cause of events in time precedes the effect. In fact, TE is an improvement upon Granger causality analysis. Granger causality requires that any variable in the system can be linearly expressed by lagged variables and error terms of the system. In contrast, transfer entropy makes no assumptions about the underlying structure of the data; thus, it can detect both linear and nonlinear causal relationships. However, calculating transfer entropy for high-dimensional data or long time series may increase computational time. Additionally, the estimation of transfer entropy depends on the sample size, and insufficient samples can affect the results.

3.3.1. Conditional Mutual Information

Given event $Z_K$, the mutual information obtained about event $y_j$ after knowing event $x_i$ is defined as:

$$I(y_j; x_i|z_k) = \log \frac{p(y_j|x_i, z_k)}{p(y_j|z_k)} \qquad (8)$$

To find the conditional average mutual information of $X$ on $Y$ given $Z$, we compute the expectation over variables $X$, $Y$, and $Z$:

$$
\begin{aligned}
I(Y; X|Z) &= E[I(y_j; x_i|z_k)] \\
&= \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{k=1}^{N} p(y_j, x_i, z_k) \log \frac{p(y_j|x_i, z_k)}{p(y_j|z_k)} \\
&= \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{k=1}^{N} p(y_j, x_i, z_k) \log p(y_j|x_i, z_k) - \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{k=1}^{N} p(y_j, x_i, z_k) \log p(y_j|z_k) \\
&= H(Y|Z) - H(Y|X, Z)
\end{aligned}
\qquad (9)
$$

3.3.2. Transfer Entropy

Given discrete variables $X_i$ and $Y_j, i = 1, 2, \cdots, N$, of equal length and interacting, the transfer entropy from $X$ to $Y$ reflects the information transfer from the past states of $X$ to $Y$, expressed as:

$$
\begin{aligned}
TE_{X \to Y} &= \sum p(Y_{i+1}, X_i, Y_i) \log \frac{p(Y_{i+1}|X_i, Y_i)}{p(Y_{i+1}|Y)_i} \\
&= \sum p(Y_{i+1}, X_i, Y_i) \log p(Y_{i+1}|X_i, Y_i) - \sum p(Y_{i+1}, X_i, Y_i) \log p(Y_{i+1}|Y_i) \\
&= H(Y_{i+1}|Y_i) - H(Y_{i+1}|X_i, Y_i)
\end{aligned}
\qquad (10)
$$

According to the definition of transfer entropy, $\log p(Y_{i+1}|Y_i)$ represents the probability of $Y_{i+1}$ occurring given the state of $Y_i$, thereby excluding the influence of $Y$ past states

on its future information, achieving the purpose of accurately measuring the information transmitted from $X$ to $Y$ by $TE$. Conditional mutual information measures the dependency between variables $X$ and $Y$, considering the information provided by variable $Z$. Therefore, based on the relationship between conditional mutual information and transfer entropy, this study proposes the following formula:

$$TE_{X \to Y} = I(Y_{i+1}; X_i | Y_i) \tag{11}$$

$TE_{X \to Y}$ represents the information transfer from $X$ to $Y$, indicating the extent of influence. If $TE_{X \to Y} > TE_{Y \to X}$, it means that the influence of variable $X$ on $Y$ is greater than the influence of $Y$ on $X$, thereby identifying $X$ as a driving factor for $Y$. By calculating the transfer entropy between meteorological data and runoff, and comparing the values of $TE_{X \to Y}$ and $TE_{Y \to X}$, meteorological driving factors related to runoff can be identified.

*3.4. Data-Driven Methods*

3.4.1. IPSO

Particle Swarm Optimization (PSO) is a stochastic optimization algorithm proposed by Eberhart and Kennedy in 1995. It simulates the collective behavior of birds, achieving optimal foraging behavior through information sharing among individuals. During the iterative process, particles move randomly according to certain rules, searching for the optimal solution within a specified range. When particles approach a local optimum, it induces the entire swarm to migrate towards that local optimum [40]. However, Particle Swarm Optimization (PSO) is prone to issues such as local optima and premature convergence. Therefore, this study introduces the concept of adaptive mutation and proposes an Improved Particle Swarm Optimization algorithm (IPSO). This algorithm aims to enhance the optimization capability of the swarm by improving the selection of classical weights and utilizing nonlinear weights, thereby reducing the randomness of particle position updates in the population as the number of iterations increases.

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \tanh(\frac{4\pi}{4t_{\max}}) \tag{12}$$

The tanh function constrains the weight $\omega$ to $[\omega_{\min}, \omega_{\max}]$. When the number of iterations is small, $\omega$ approaching $\omega_{\max}$ ensures that particles not only enhance the inheritance of initial velocity information at the beginning of iterations, thereby accelerating the search speed, but also maintain global search capability. As $t$ increases, $\omega$ decreases nonlinearly towards but does not equal $\omega_{\min}$, enhancing the flexibility of particles, ensuring the algorithm's local search capability. Moreover, particles in the later stages of iteration are more influenced by the global optimal position, aiding in the determination of the global optimum [40].

The specific steps of the IPSO optimization algorithm are as follows:

(1) Initialize the positions $X = (x_1, x_2, \cdots, x_D)$ and velocities $V = (v_1, v_2, \cdots, v_D)$ for all particles. The historical best position for each particle is $pbest = (p_1, p_2, \cdots, p_D)$, and for the swarm, it is $gbest = (g_1, g_2, \cdots, g_D)$.

(2) Calculate the fitness of each particle. If the current value is better than the particle's historical best value, update $pbest$. If the current value is better than the global historical best value, update $gbest$.

(3) Update the position and velocity of each particle using the following equations:

$$v_d^{t+1} = \omega v_d^t + c_1 r_1 (pbest_d^t + x_d^t) + c_2 r_2 (gbest_d^t - x_d^t) \tag{13}$$

$$x_d^{t+1} = x_d^t + v_d^{t+1} \tag{14}$$

where $\omega$ represents the inertia weight, typically initialized within the range $[0.4, 0.9]$, to maintain the particle's motion inertia and the ability to explore search space. $c_1$ and $c_2$, respectively, denote the individual learning factor and swarm learning factor,

with values in the range $[0, 4]$, used to balance the influence of individual and swarm experience information on the optimization process. $r_1$ and $r_2$ have values in the range $[0, 1]$, used to increase the randomness of the search process.

Function (13) consists of inertia, individual cognition, and social cognition. Inertia represents the particle's habit from previous iterations, inheriting its own velocity. Individual cognition represents the inheritance of past positions by the particle. Social cognition ensures shared information among particles, representing collective experience.

(4) Adaptively mutate particle positions and adjust the mutation probability *prob* based on the optimization search process. Additionally, when the D-dimensional position changes, the particle's position randomly varies within its range, as shown in the following formula:

$$prob = 0.5 \times \frac{t}{\text{max\_iter}} + 0.5 \tag{15}$$

$$x_d^{t+1} = r_3(\max(x_d) - \min(x_d)) + \min(x_d) \tag{16}$$

where *prob* represents the particle mutation probability, which decreases with increasing iteration count. max\_*iter* denotes the maximum number of iterations, $\max(x_d)$ and $\min(x_d)$ are, respectively, the maximum and minimum values of the D-dimensional vector. The range of values for $r_3$ is $[0, 1]$.

To prevent particles from getting stuck in local minima during the search process and to increase variability in selecting positions, a mutation factor (15) is introduced in the IPSO algorithm. This allows some particles to ignore the individual best positions selected in historical iterations and initialize randomly. As the number of iterations increases, the mutation probability for random particles decreases.

(5) If the iteration count reaches max\_*iter* or the global best fitness value is less than a specified value, terminate the process; otherwise, proceed to step (2). Validate the effectiveness of the improvement using the Sphere test function.

$$Sphere : f(x) = \sum_{i=1}^{n} x_i^2 \tag{17}$$

Apart from the selection of the inertia weight $\omega$, both algorithms are set with identical parameters. As seen from Figure 3, compared to the standard Particle Swarm Optimization algorithm, the IPSO algorithm, improved with inertia weight, demonstrates stronger optimization capabilities as the number of iterations increases, with a more stable trend.
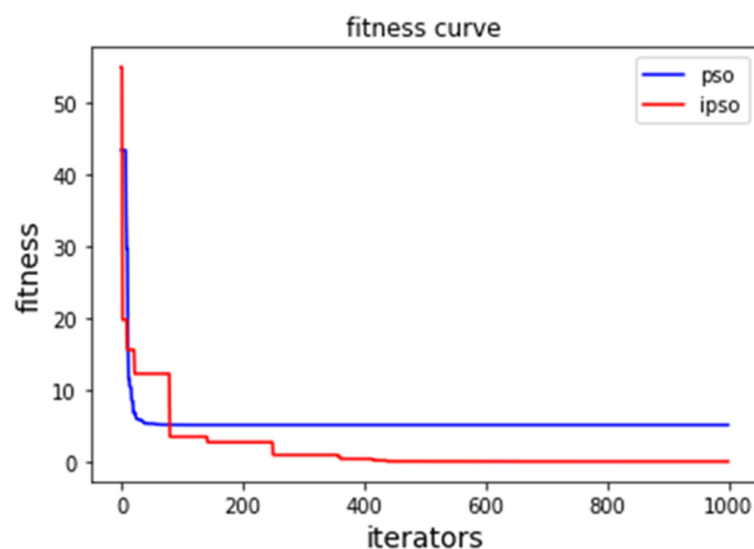


**Figure 3.** Comparison of iterative optimization results.

3.4.2. Temporal Convolutional Network

The Temporal Convolutional Network (TCN) is an emerging model proposed by Shao Jie Bai et al. [47] in 2018. It consists of modules such as causal convolution, dilated convolution, and residual connections. TCN is commonly used to address multidimensional time-series problems and offers advantages such as higher computational efficiency, more stable gradients, and smaller training memory footprint. A notable feature of TCN is that it does not incorporate information from future time steps, thereby avoiding data leakage issues.

(1) Causal convolution

As a time-series prediction model, TCN introduces causal convolution, allowing time-series problems to be transformed into predicting $y_1, y_2, \cdots, y_t$ using $x_1, x_2, \cdots, x_t$. As shown in Figure 4, for convolution kernel $F = (f_1, f_2, \cdots, f_k)$ and sequence $X = (x_1, x_2, \cdots x_t)$ the causal convolution at $x_t$ yields $(F \times X)_{x_t} = \sum_{i=1}^{k} f_i x_{t-k+i}$. Assuming the last two nodes of the input layer are $x_{t-1}, x_t$, the last node of the first hidden layer is $y_t$, and the convolution kernel is $F = (f_1, f_2)$, according to the formula, we have $y_t = f_1 x_{t-1} + f_2 x_t$. For causal convolution, the value at the previous layer at time t depends only on the values at the current layer at time $t$ and earlier, without extracting future data information. TCN employs a one-dimensional fully convolutional network with a stride of 1 and zero-padding size of $k - 1$ (where $k$ is the kernel size), ensuring that the input and output sizes of the model are equal.
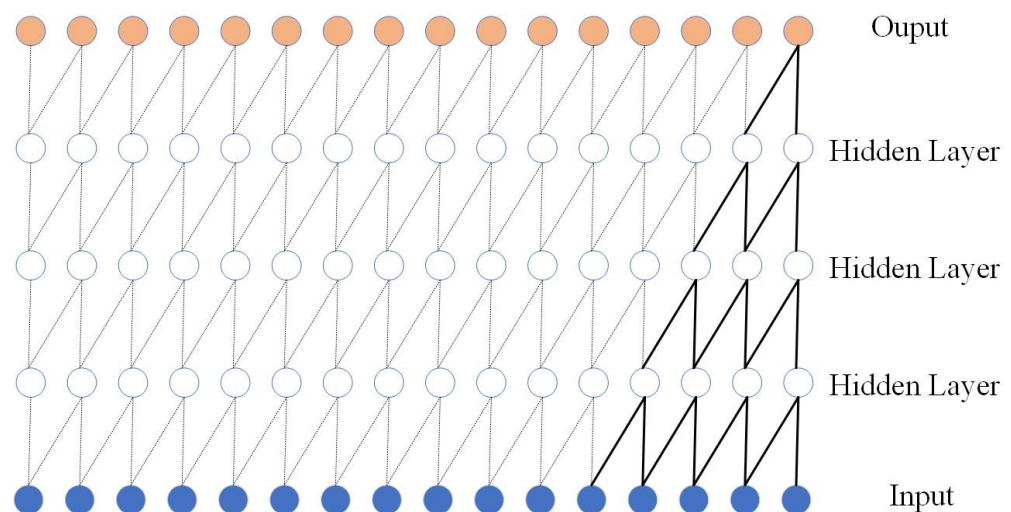
(2) Dilated convolution



**Figure 4.** Causal convolution.

Introducing dilated convolutions in TCN addresses the problem of limited receptive fields. As shown in Figure 5, dilated convolutions insert gaps into the receptive field of CNNs. The dilation factor $d$ increases exponentially by powers of 2, significantly enlarging the receptive field index of TCN. This allows shallow layers to capture larger receptive fields, and the output at the top layer can receive input information from a broader range [48]. The receptive field size of dilated convolution is $(k - 1) \times d + 1$, which can be expanded by increasing the kernel size $k$ or the dilation factor $d$.
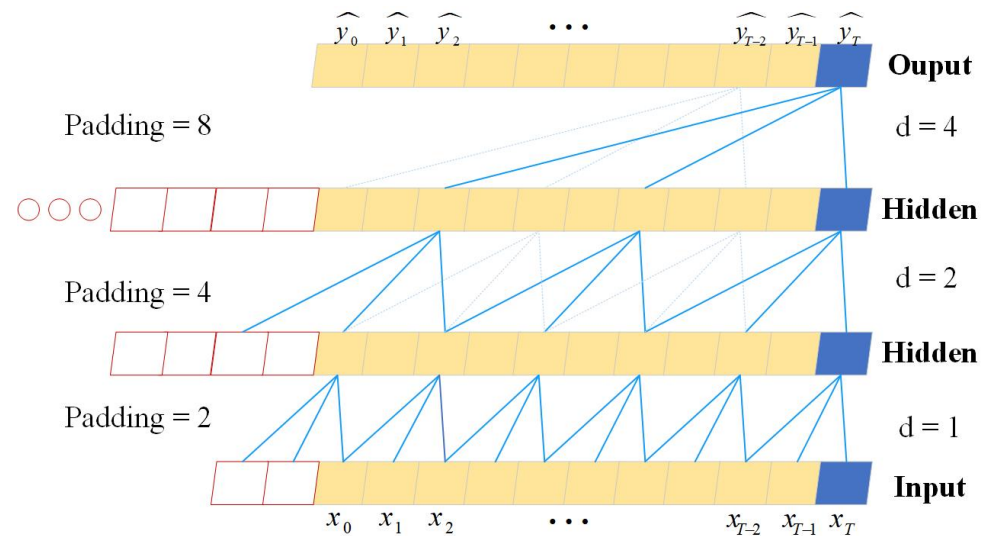
(3) Residual connection

**Figure 5.** Dilatational convolution.

When TCN introduces causal convolution and dilated convolution to expand the receptive field, the problem of gradient vanishing or gradient explosion can occur as the number of network layers increases [49]. To address this issue, this study introduces residual modules for analysis, as shown in Figure 6. Residual connections transmit information across layers by adding the input sequence *X* of the model to the output sequence *F*(*X*) of the convolution computation. The residual module consists of two sets of dilated convolution layers, weight normalization layers, ReLU activation functions, and Dropout layers [50]. Unlike the general ResNet model, TCN directly adds the input sequence *X* to the output sequence *F*(*X*) of the residual module while simultaneously processing *X* with convolution to ensure that *F*(*X*) and *X* have the same dimensions.
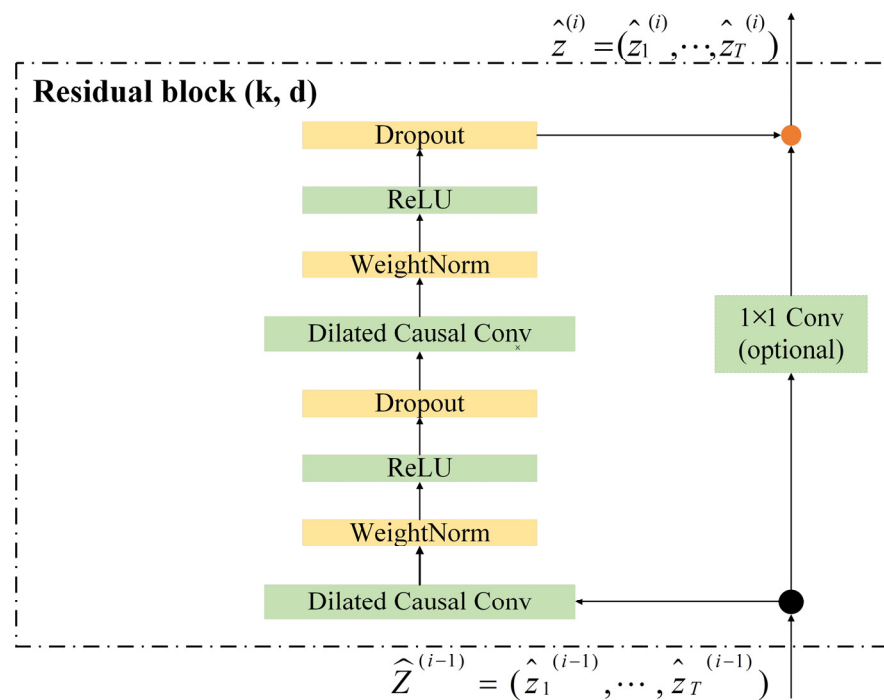


**Figure 6.** Residual module.

### 3.4.3. IPSO-TCN

The Improved Particle Swarm Optimization (IPSO) algorithm has efficient search and generalization capabilities and is easy to combine with machine learning models. However, due to the numerous parameters of TCN, complex debugging, and the lack of strict and accurate parameter selection methods, it is challenging to use in practical applications [51]. Therefore, the IPSO algorithm is combined with TCN for runoff prediction. Figure 7 shows the modeling process of the IPSO-TCN model.
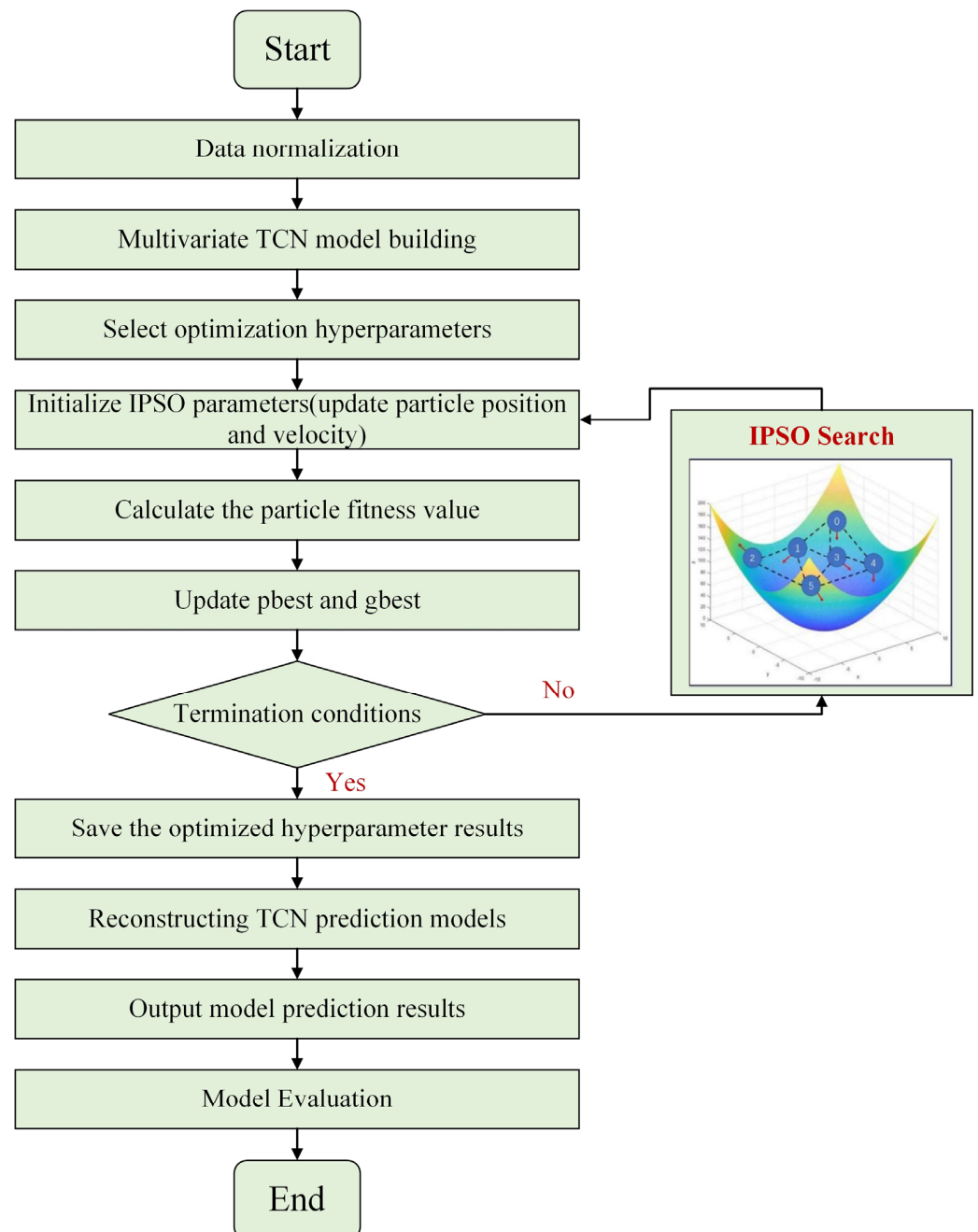


**Figure 7.** IPSO-TCN model flow.

As shown in Figure 7, the modeling steps of the IPSO-TCN model are as follows:

(1)     Standardized data of runoff driving factors.
(2)     Construction of a multi-input single-output TCN model.
(3)     Select hyperparameters to be optimized for the TCN model.

(4)  Initialize the positions and velocities of the particle swarm in the IPSO algorithm.
(5)  Calculate the fitness of the particle swarm.
(6)  Update the positions and velocities of the particle swarm.
(7)  Evaluate termination conditions; if not met, continue optimizing hyperparameters using the IPSO algorithm.
(8)  Reconstruct the TCN model using the computed optimal hyperparameters.
(9)  Output model prediction results and conduct model evaluation.

The IPSO-TCN model has numerous parameters. To enhance efficiency, we optimize the hyperparameters that significantly impact model performance. Based on prior experience, this study selects the number of convolution kernels, kernel size, Dropout factor, and InitialLearnRate as the parameters to be optimized, with $\omega_{max} = 0.9$, $\omega_{min} = 0.4$, $c_1 = 1.5$, $c_2 = 1.5$ and a maximum of 10 iterations. On the basis of the established TCN model, we set the optimization range for each parameter, randomly generating values between the lower and upper bounds of the parameters according to Formulas (13) and (16) to determine the initial positions and velocities of the particles and compute their fitness values. Through iterative and cyclic processes, the optimization results are eventually determined. Additionally, runoff influence factors are selected as input variables, with runoff as the output variable. Based on this, a multivariate IPSO-TCN runoff prediction model is established.

### 3.5. Coupled Knowledge Embedding and Data-Driven Runoff Prediction Model

#### 3.5.1. Knowledge Embedding

Knowledge embedding is an important method for integrating knowledge and data. Through knowledge embedding, barriers between knowledge and data can be eliminated, enabling the establishment of machine learning models with physical insights to improve model accuracy and robustness [52]. Knowledge embedding can be applied in many stages of modeling (Figure 8). For example, during data preprocessing, physical constraints, domain knowledge, and prior experience can be embedded, which often relates to feature engineering and data normalization. In the stage of model structure design, the network or topology structure of the model can be adjusted based on domain knowledge. Embed domain knowledge into the model optimization and adjustment process to construct a knowledge-embedded loss function.
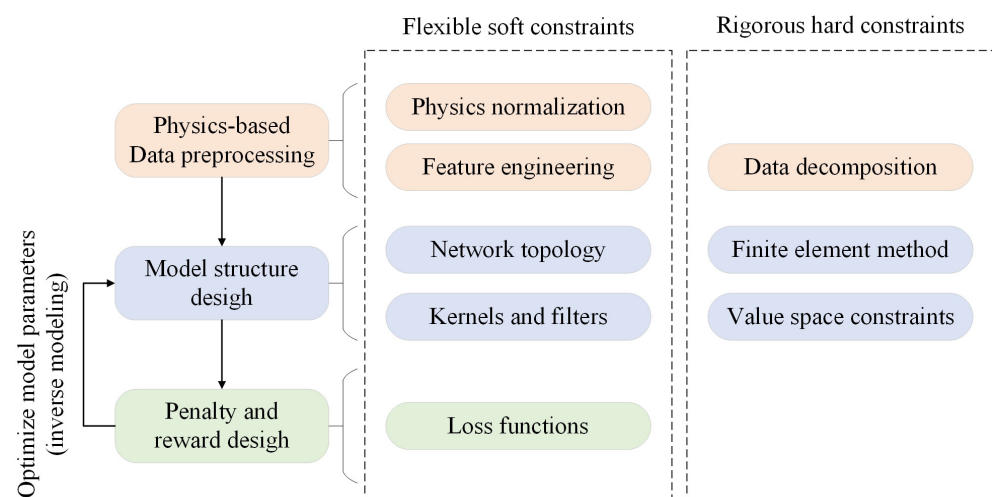


**Figure 8.** Diagram of the classification of knowledge-embedding algorithms.

However, currently, most runoff prediction models do not effectively utilize prior knowledge, experience, and physical mechanisms, greatly limiting the application of machine learning. Purely data-driven models not only require large amounts of data but may also produce predictions that violate physical mechanisms [36]. Integrating domain

knowledge into machine learning models has the potential to overcome barriers between data-driven and knowledge-driven models.

Knowledge embedding involves several steps:

(1)  This study obtained the probability density curve of runoff based on monthly discharge data from hydrological stations spanning from 1964 to 2023, using Gaussian kernel density estimation. The formula for Gaussian kernel density estimation is as follows:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

(18)

where $\hat{f}(x)$ represents the estimated density at point $x$, $n$ denotes the number of data points, $h$ is the bandwidth controlling the width of the kernel function, and $K$ is the kernel function. This study adopts the Gaussian function $K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$.

(2)  Incorporating the probability distribution information implied by runoff as prior knowledge, a custom loss function layer is defined to integrate these constraints into a data-driven model, establishing a coupled knowledge embedding and data-driven runoff prediction model. Therefore, the loss function of the coupled model can be reformulated as:

$$L(\theta) = MSE_{DATA} + MSE_{PDF}$$

(19)

$$MSE_{DATA} = \frac{1}{N_{DATA}}\sum_{i=1}^{N_{DATA}}|\hat{y}_i - y_i|^2$$

(20)

$$MSE_{PDF} = \frac{1}{N_{PDF}}\sum_{i=1}^{N_{PDF}}|\hat{y_i^*} - y_i^*|^2$$

(21)

The total loss of the coupled model is reconstructed as a combination of data loss and prior loss, as shown in Equation (19). In Equation (20), $\hat{y}_i$ and $y_i$ represent the predicted and true values, respectively. In Equation (21), $\hat{y_i^*}$ and $y_i^*$ are the probability density values corresponding to $\hat{y}_i$ and $y_i$, respectively. Embedding prior knowledge can assist in the training process of data-driven models, reducing the likelihood of the model making physically unrealistic predictions and thereby enhancing model performance.

### 3.5.2. Coupled Model

Due to the nonlinear and non-stationary characteristics of runoff sequences, incorporating prior knowledge embedded in these sequences as theoretical guidance for data-driven models can enhance runoff prediction and reduce the models' dependence on data. As shown in Figure 9, the main process of constructing a coupled knowledge-embedded and data-driven runoff prediction model can be simply represented in the following four steps:

(1)  Use transfer entropy to select feature factors.
(2)  Calculate the probability density function (PDF) of runoff based on Gaussian kernel density estimation and combine the runoff PDF with mean squared error (MSE) to reconstruct the loss function $L(\theta)$.
(3)  Use the selected feature factors as inputs for the IPSO-TCN model to predict runoff.
(4)  Train the data-driven model (IPSO-TCN) using the loss function embedded with runoff probability density values, continuously testing and validating the results.
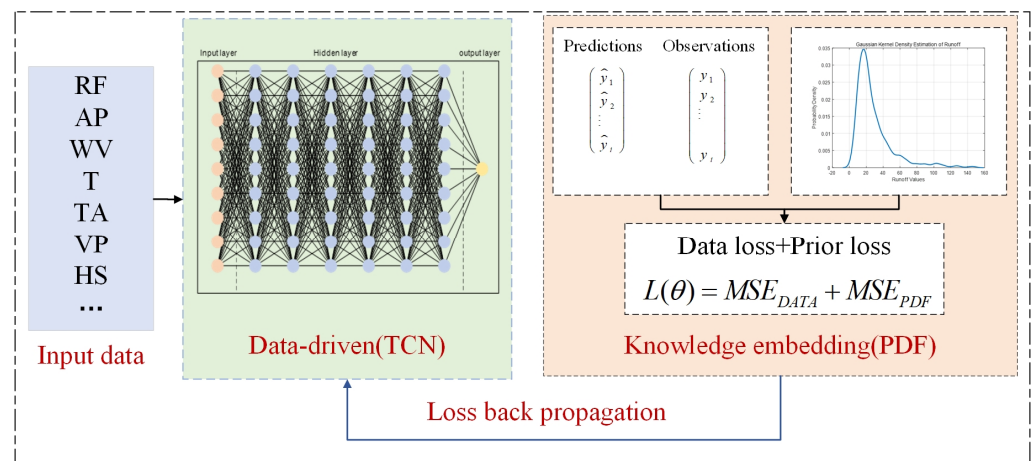
**Figure 9.** Coupled data-driven and knowledge-embedded architecture. In the data-driven model (TCN), the pink dots are the neurons in the input layer, the blue dots are the neurons in the hidden layer, and the yellow dots are the neurons in the output layer.

3.5.3. Evaluation Metrics

To evaluate the predictive performance of the model from different perspectives, this study selects Mean Absolute Error (*MAE*), Root Mean Squared Error (*RMSE*), and Nash–Sutcliffe Efficiency (*NSE*) as metrics to assess and compare the model's performance during training and testing. These efficiency criteria are defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{22}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|^2} \tag{23}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{24}$$

In the equation, $y_i$ represents the true value, $\hat{y}_i$ represents the predicted value, $\overline{y}_i$ is the mean of the predicted sequence, and $n$ is the length of the sequence.

## 4. Results

### 4.1. Spatiotemporal Analysis of Runoff

To study the runoff prediction model, we first analyze the spatial and temporal distribution characteristics of runoff. This study uses monthly runoff data from 1964 to 2023 (a total of 60 years) based on nine major hydrological stations in the Yellow River Basin to assess the spatial and temporal variations in runoff data, as shown in Figure 10. By calculating the Standardized Runoff Index (SRI) for the corresponding 60-month periods and applying the drought classification standards (Table 2), drought occurrences are identified when SRI is less than 0.5. The frequency of droughts over the past 60 years in the Yellow River Basin is obtained by accumulating the number of instances where SRI is less than 0.5, as shown in Table 3. The spatial and temporal variations in drought frequency are illustrated in Figure 11.
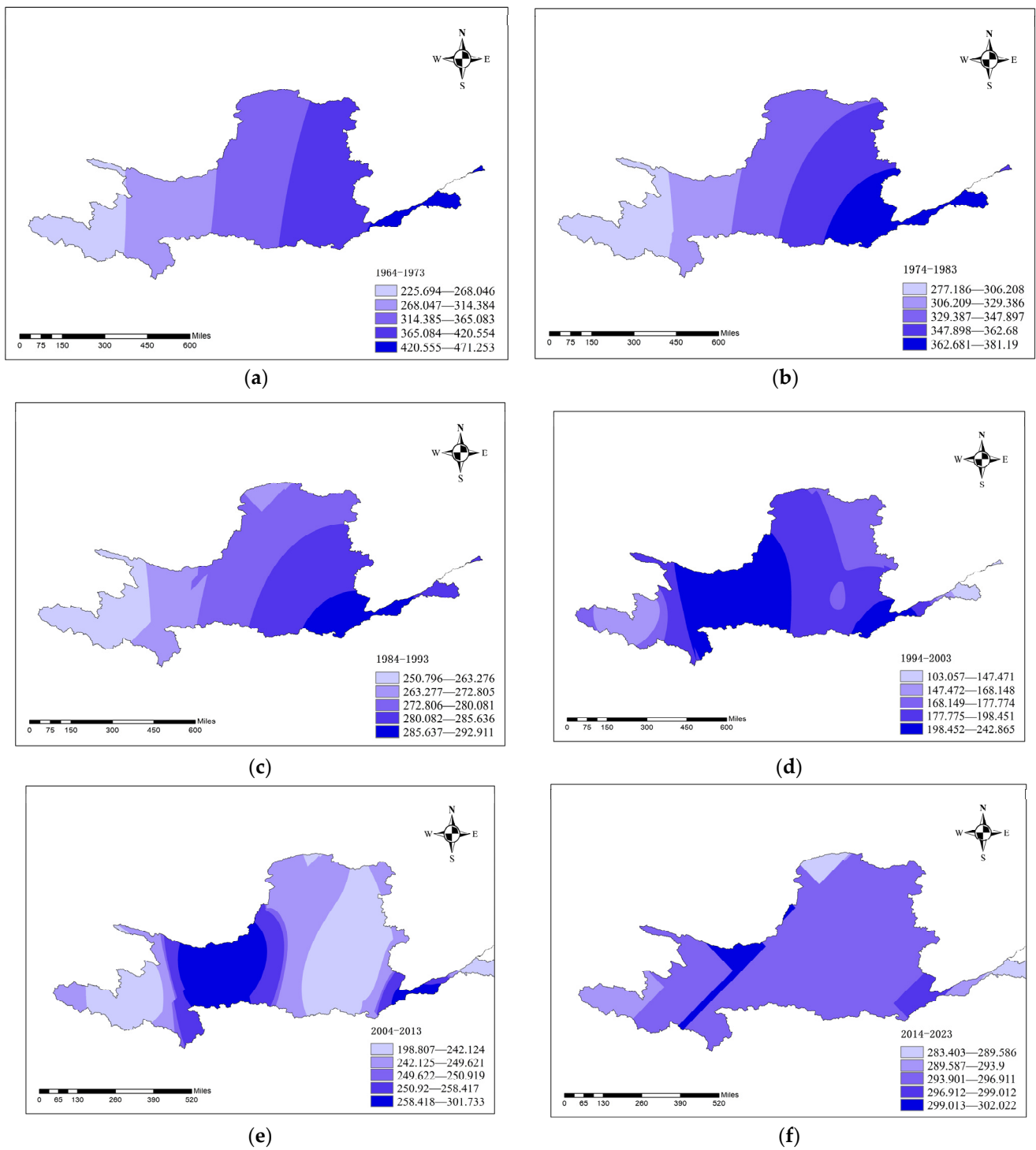
**Figure 10.** Runoff evolution trend. (**a**) is the spatial and temporal distribution of runoff in the Yellow River Basin during 1964–1973. (**b**) is the spatial and temporal distribution of runoff in the Yellow River Basin during 1974–1983. (**c**) is the spatial and temporal distribution of runoff in the Yellow River Basin during 1984–1993. (**d**) is the spatial and temporal distribution of runoff in the Yellow River Basin during 1994–2003. (**e**) is the spatial and temporal distribution of runoff in the Yellow River Basin from 2004 to 2013. (**f**) is the spatial and temporal distribution of runoff in the Yellow River Basin during 2014–2023.

**Table 3.** Drought frequency table.

| Station / Time | Tangnaihai | Lanzhou | Xiaheyan | Shizuishan |
|---|---|---|---|---|
| 1964–1973 | 36 | 57 | 50 | 44 |
| 1974–1983 | 22 | 23 | 19 | 18 |
| 1984–1993 | 42 | 42 | 42 | 35 |
| 1994–2003 | 73 | 63 | 70 | 68 |
| 2004–2013 | 36 | 18 | 24 | 28 |
| 2014–2023 | 27 | 29 | 23 | 29 |
| 1964–2023 | 236 | 232 | 228 | 222 |

| Station / Time | Longmen | Sanmenxia | Huayuankou | Gaocun | Lijin |
|---|---|---|---|---|---|
| 1964–1973 | 31 | 23 | 21 | 16 | 5 |
| 1974–1983 | 15 | 17 | 25 | 25 | 18 |
| 1984–1993 | 35 | 24 | 26 | 28 | 36 |
| 1994–2003 | 67 | 77 | 82 | 84 | 86 |
| 2004–2013 | 52 | 57 | 56 | 46 | 35 |
| 2014–2023 | 48 | 41 | 39 | 40 | 35 |
| 1964–2023 | 248 | 239 | 249 | 239 | 215 |

As shown by the runoff evolution trends in Figure 10, the runoff volume in the lower Yellow River was higher from 1964 to 1993, while the runoff volume in the middle reaches was higher from 1994 to 2013. The overall runoff volume in the Yellow River Basin has gradually decreased since 1984. The period from 1994 to 2003 marks the decade with the lowest runoff volume in the Yellow River Basin over the past 60 years.

As shown by the drought evolution trends in Figure 11, the drought conditions in the Yellow River Basin were relatively stable from 1964 to 1993, with the fewest drought occurrences in the decade from 1974 to 1983. Since 1994, the drought conditions in the Yellow River Basin have intensified, with an increase in drought occurrences, but starting from 2004, the drought conditions gradually returned to a stable state. Overall, the droughts in the Yellow River Basin over the past 60 years show a trend of shifting from the upper to the lower reaches. According to Table 3, Huayuankou station experienced the most drought occurrences, with 249 instances in the 60 years of monthly runoff data among the nine stations, making it the chosen example for studying the runoff prediction model.

*4.2. Cross-Wavelet Analysis*

The cross-wavelet transform technique is used to analyze the driving responses between meteorological drought and hydrological drought, providing a reference for setting the sliding window in subsequent runoff prediction models. The rainfall and runoff series are interdependent processes, mutually promoting and interacting with each other. Cross-wavelet analysis can more sensitively respond to the evolutionary characteristics between rainfall and runoff series.

To analyze the time–lag relationship between precipitation and runoff, cross-wavelet analysis is applied to meteorological and hydrological data spanning 60 years (1964–2023), totaling 720 months. Figure 12 presents the cross-wavelet power spectrum and coherence spectrum at the monthly scale for the Huayuankou hydrological station in the Yellow River Basin. The direction of the arrows indicates the phase relationship between meteorological drought and hydrological drought. An arrow rotating 30° signifies that the precipitation sequence leads or lags the runoff sequence by 1 month. An arrow pointing to the right indicates a strong positive correlation with the same phase changes between different sequences, while an arrow pointing to the left signifies a strong negative correlation with opposite phase changes. If the arrow points downward, it indicates an advance in changes, and if it points upward, it indicates a lag in changes.
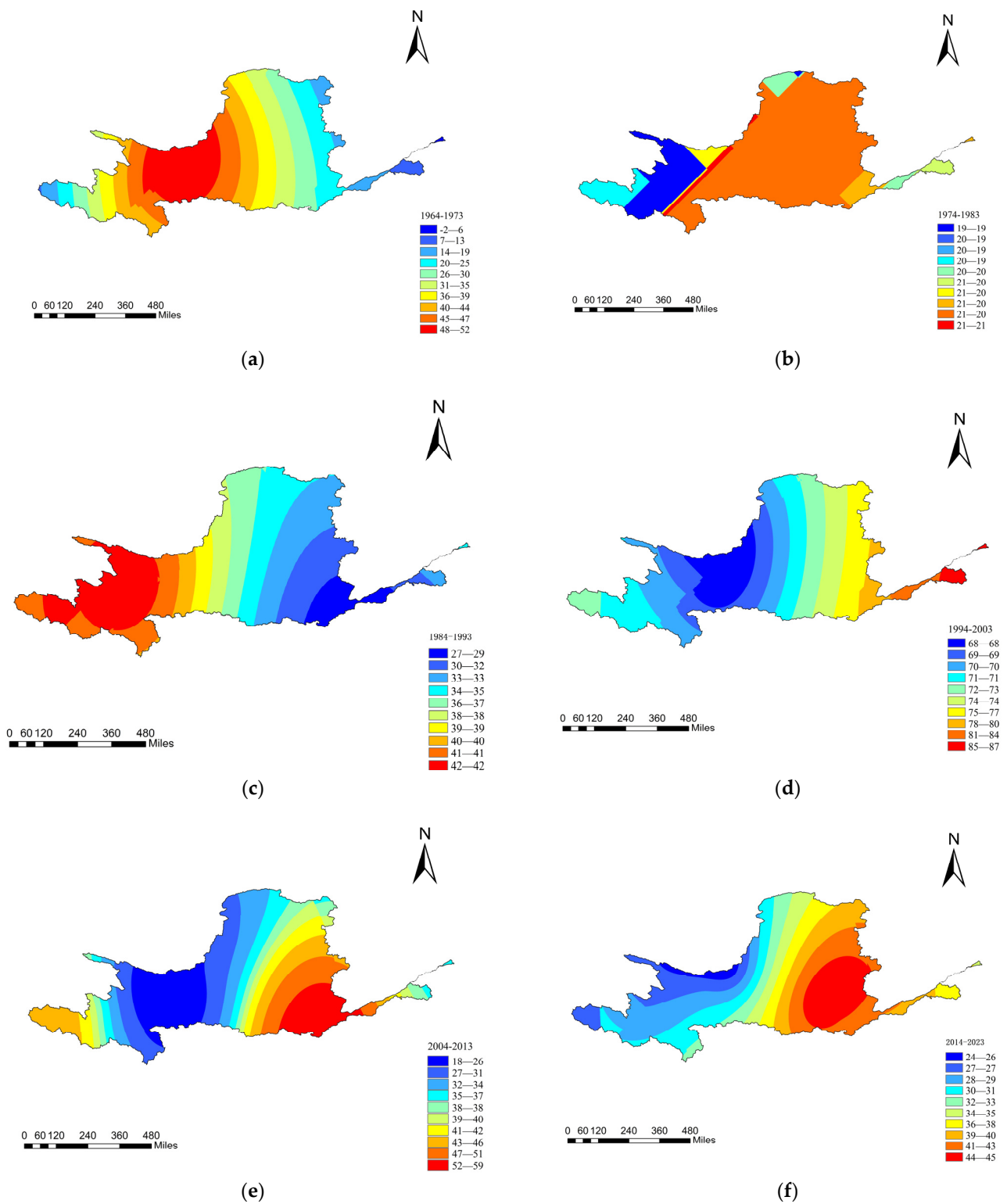
**Figure 11.** Drought evolution trend. (**a**) shows the spatial and temporal distribution of drought frequency in the Yellow River Basin during 1964–1973, and (**b**) shows the spatial and temporal distribution of drought frequency in the Yellow River Basin during 1974–1983. (**c**) shows the spatial and temporal distribution of drought frequency in the Yellow River Basin during 1984–1993, (**d**) shows the spatial and temporal distribution of drought frequency in the Yellow River Basin during 1994–2003, and (**e**) shows the spatial and temporal distribution of drought frequency in the Yellow River Basin during 2004–2013. (**f**) shows the spatial and temporal distribution of drought frequency in the Yellow River Basin during 2014–2023.
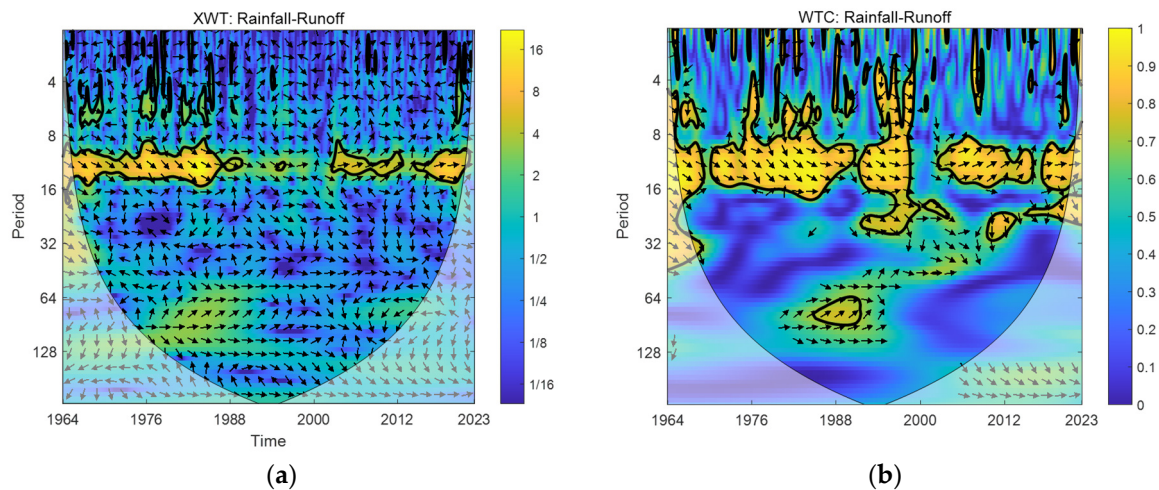
**Figure 12.** Cross-wavelet power spectrum and coherence spectrum. (**a**) is the cross wavelet power spectrum of precipitation and runoff, which mainly shows the periodicity of precipitation and runoff. (**b**) is the cross wavelet coherence spectrum of precipitation runoff, which mainly shows the time-lag relationship of precipitation runoff.

Based on the cross-wavelet power spectrum and coherence spectrum analysis shown in Figure 12a,b, the spectral energy at the Huayuankou station is mainly concentrated between 10 and 14 a during the periods of 1964–1988 and 2003–2023, indicating a strong positive correlation between precipitation and runoff. From the phase perspective, the precipitation–runoff relationship exhibits a positive phase, indicating a positive correlation. The phase difference between hydrological drought and meteorological drought ranges from 30° to 60°, meaning that hydrological drought lags behind meteorological drought by 1 to 2 months.

*4.3. Driving Factors Analysis*

Transfer entropy measures the reduction in the uncertainty of predicting runoff values by feature factors, excluding the influence of the past states of these factors. It is a model-free causal statistic representing the asymmetric information transfer between random variables [53]. $TE_{X \to Y}$ represents the information transfer from $X$ to $Y$, indicating the degree of influence. If $TE_{X \to Y} > TE_{Y \to X}$, it means the influence of variable $X$ on $Y$ is greater than that of $Y$ on $X$; thus, $X$ is considered the driving factor for $Y$.

Given the characteristics of transfer entropy mentioned above, this study collected data on nine factors related to runoff. These factors are rainfall (*RF*), atmospheric pressure (*AP*), wind velocity (*WV*), temperature (*T*), temperature anomaly (*TA*), vapor pressure (*VP*), hours of sunshine (*HS*), relative humidity (*RH*), and sunshine percentage (*SP*). Setting the variable $Y = \{Runoff\}$, $X = \{RF, AP, WV, T, TA, VP, HS, RH, SP\}$, the values of $TE_{X \to Y}$ and $TE_{Y \to X}$ were calculated. The results are shown in Table 4 and Figure 13.

Based on Table 4 and Figure 13, analysis of the transfer entropy calculations between $X = \{RF, AP, WV, T, TA, VP, HS, RH, SP\}$ and $Y = \{Runoff\}$ shows that the transfer entropy values from *R* to *AP*, *WV*, *T*, *TA*, *VP*, and *SP* are all smaller than those from these factors to *R*. This indicates that the direction of information transfer is from *R* to the influencing factors, suggesting they cannot be considered as causative factors. Furthermore, $TE_{RF \to R} = 0.5188$, $TE_{HS \to R} = 0.8750$, $TE_{RH \to R} = 0.7934$, the corresponding $TE_{R \to RF} = 0.4080$, $TE_{R \to HS} = 0.5540$, $TE_{R \to RH} = 0.5816$. The transfer entropy values of the former are all greater than those of the latter. This indicates that *RF*, *HS*, and *RH* are characteristic factors of runoff, namely, $R_{feature} = \{RF, HS, RH\}$.

**Table 4.** Transfer entropy value.

| Factors (X) | Runoff (R) (Y) | | Characteristic Factor |
|---|---|---|---|
| | $TE_{X \to Y}$ | $TE_{Y \to X}$ | |
| Rainfall (RF) | 0.5188 | 0.4080 | Yes |
| Atmospheric Pressure (AP) | 0.1203 | 0.2981 | No |
| Wind Velocity (WV) | 0.0357 | 0.1289 | No |
| Temperature (T) | 0.1758 | 0.4142 | No |
| Temperature Anomaly (TA) | 0.1481 | 0.2142 | No |
| Vapor Pressure (VP) | 0.3138 | 0.4204 | No |
| Hours of Sunshine (HS) | 0.8750 | 0.5540 | Yes |
| Relative Humidity (RH) | 0.7934 | 0.5816 | Yes |
| Sunshine Percentage (SP) | 0.4194 | 0.6072 | No |
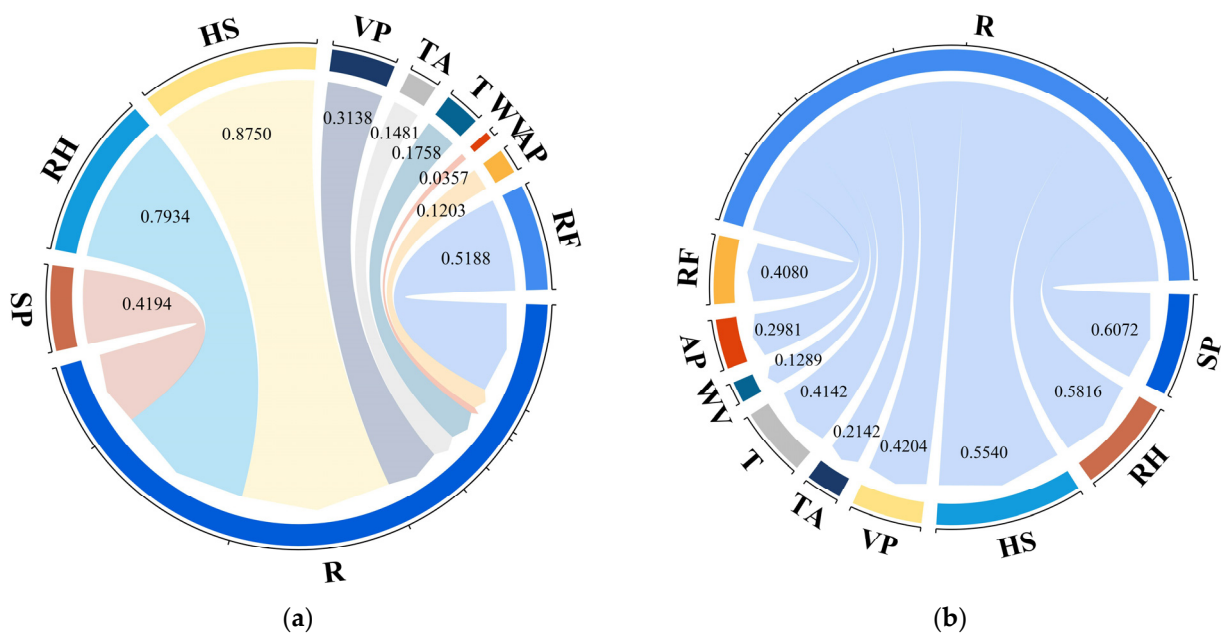


**Figure 13.** Information transfer relationship network. (**a**) is to calculate the degree of influence of meteorological factors on runoff, and (**b**) is to calculate the degree of influence of runoff on meteorological factors.

*4.4. Analysis of Coupled Model Predictions*

4.4.1. Model Parameter Configuration

Analysis from Section 4.3 reveals that using the transfer entropy method identifies RF, HS, and RH as factors related to runoff. This study investigates three runoff prediction models: the univariate input data model with runoff as the sole input variable (TCN-UID), the multivariate input data model with RF, HS, RH, and R as inputs (TCN-MID), and the coupled knowledge-embedding and data-driven runoff prediction model with RF, HS, RH, and R as inputs (coupling model). The datasets for all three models are divided into 85% training and 15% testing sets. Additionally, the computed runoff probability density function (19) is embedded into the data-driven TCN model to construct the coupled knowledge-embedding and data dual-driven runoff prediction model (coupling model). Following Section 3.2, the sliding window step size for all three models is set to 2. Section 3.4.1 guides the adaptive optimization of hyperparameters using IPSO, where parameters, such as number of convolutional kernels, kernel size, Dropout factor, and Initial Learn Rate, are optimized. Setting $\omega_{\max} = 0.9$, $\omega_{\min} = 0.4$, $c_1 = 1.5$, $c_2 = 1.5$, with a maximum iteration of 10, the optimization ranges and initial positions and velocities for each parameter are determined based on established models, calculating fitness values

iteratively to ultimately confirm optimization outcomes. The parameter configurations for all three models are consistent, detailed in Table 5.

**Table 5.** Parameters of three models.

| Parameters | Num Filters | Filter Size | Dropout Factor | Num Blocks |
|---|---|---|---|---|
| Value | 32 | 2 | 0.01 | 1 |
| Parameters | Optimizer | Initial Learn Rate | Max Epochs | Mini Batch Size |
| Value | Adam | 0.01 | 300 | 2 |

### 4.4.2. Analysis of Model Results

To comprehensively assess the predictive effectiveness of the coupled knowledge-embedding and data-driven model and underscore its applicability in practical applications, this study evaluated three models for runoff prediction within the watershed. These models include the TCN model with univariate input (TCN-UID), the TCN model with multivariate input (TCN-MID), and the coupled knowledge-embedding and data-driven runoff prediction model (coupling model). The prediction results and evaluation metrics for these different models are presented in Figures 14 and 15, as well as Table 6.
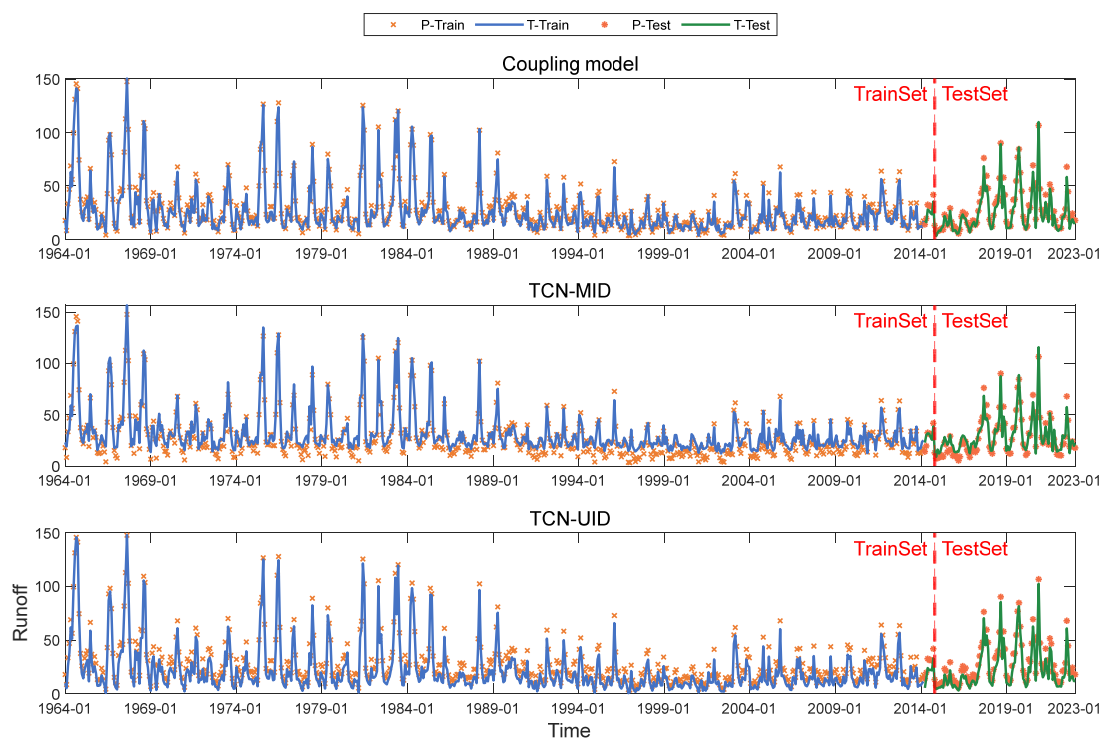


**Figure 14.** Prediction results of the three models.

**Table 6.** Model evaluation indicator values.

| Data Set | Models | $R^2$ | MAE | RMSE | NSE |
|---|---|---|---|---|---|
| Train | TCN-UID | 0.915 | 6.494 | 6.764 | 0.893 |
| | TCN-MID | 0.934 | 5.016 | 5.980 | 0.917 |
| | Coupling model | 0.978 | 2.814 | 3.459 | 0.962 |
| Test | TCN-UID | 0.892 | 6.249 | 6.523 | 0.841 |
| | TCN-MID | 0.917 | 4.767 | 5.727 | 0.907 |
| | Coupling model | 0.943 | 4.007 | 4.749 | 0.951 |

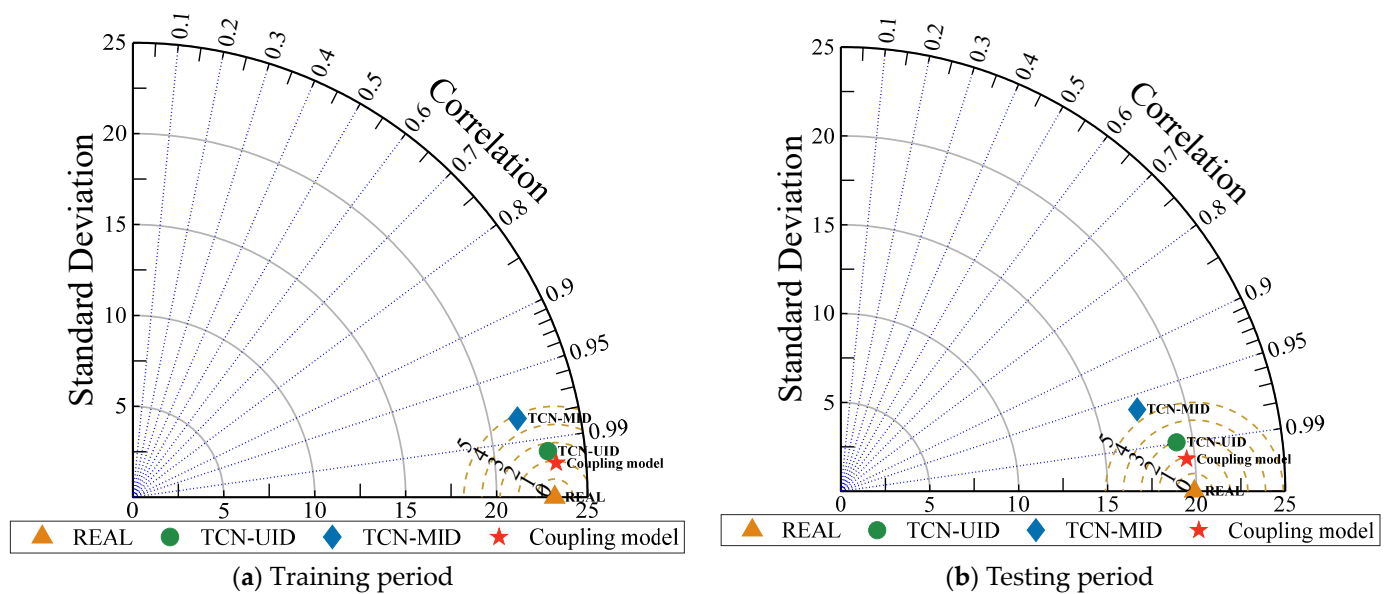(**a**) Training period                    (**b**) Testing period

**Figure 15.** Taylor diagrams of the four models.

From Figures 14 and 15, and Table 6, it is evident that we used data from January 1964 to December 2014 as the training set and data from January 2015 to December 2023 as the test set within the watershed. The datasets were input into TCN-UID, TCN-MID, and the coupling model for runoff prediction within the watershed. As shown in Figure 14, the coupling model exhibits the best-fit performance compared to TCN-UID and TCN-MID, both in the training and test sets. Even at extreme values, the model accurately predicts these extremes. Furthermore, as indicated by Figure 15 and Table 6, the coupling model closely approximates the true values in both training and test sets, showing the strongest correlation among the three models. Different evaluation metrics demonstrate the superior performance of the coupling model, highlighting its excellent model performance and strong stability.

When the model parameters are consistent, the TCN-MID model shows an improvement in performance of 2.08% and 2.8% in the training and validation sets, respectively, compared to the TCN-UID model. These results indicate that using R, RF, HS, and RH as input variables helps the model better capture the temporal dependence and periodic characteristics of runoff sequences. By incorporating multiple features (R, RF, HS, RH) as inputs, the model can learn the interaction relationships between different features, thereby capturing dependencies and complementary information among features. This enhances the model's sensitivity and adaptability to future data changes, improves its generalization capability, and ultimately ensures more reliable prediction results.

Similarly, by incorporating the runoff probability density function as prior knowledge and constructing a new loss function (19), embedded into the data-driven TCN model, the coupling model for runoff prediction is established with R, RF, HS, and RH as input variables and R as the output variable. Compared to the TCN-UID and TCN-MID models, the coupling model shows performance improvements of 6.9% and 4.7% in the training set, and 5.7% and 2.8% in the test set, respectively. This indicates that the coupling model exhibits high stability and fault tolerance, reducing the impact of external uncertainties on the model.

As shown in Figure 16, there are differences among the three models. The TCN-UID model demonstrates overall good predictive performance but performs poorly at extremely low values, especially during winter, where reduced runoff may lead to droughts. Therefore, there is room for further improvement of the TCN-UID model. The TCN-MID model shows better overall predictive performance compared to TCN-UID, mitigating the disadvantage of poor performance at extreme low values. However, its predictive

performance at extremely high values is less satisfactory, potentially causing flooding during periods of increased summer runoff. Therefore, the TCN-MID model struggles to effectively capture the nonlinear and non-stationary characteristics of runoff, resulting in discrepancies between predicted and actual values.



(**a**)



(**b**)
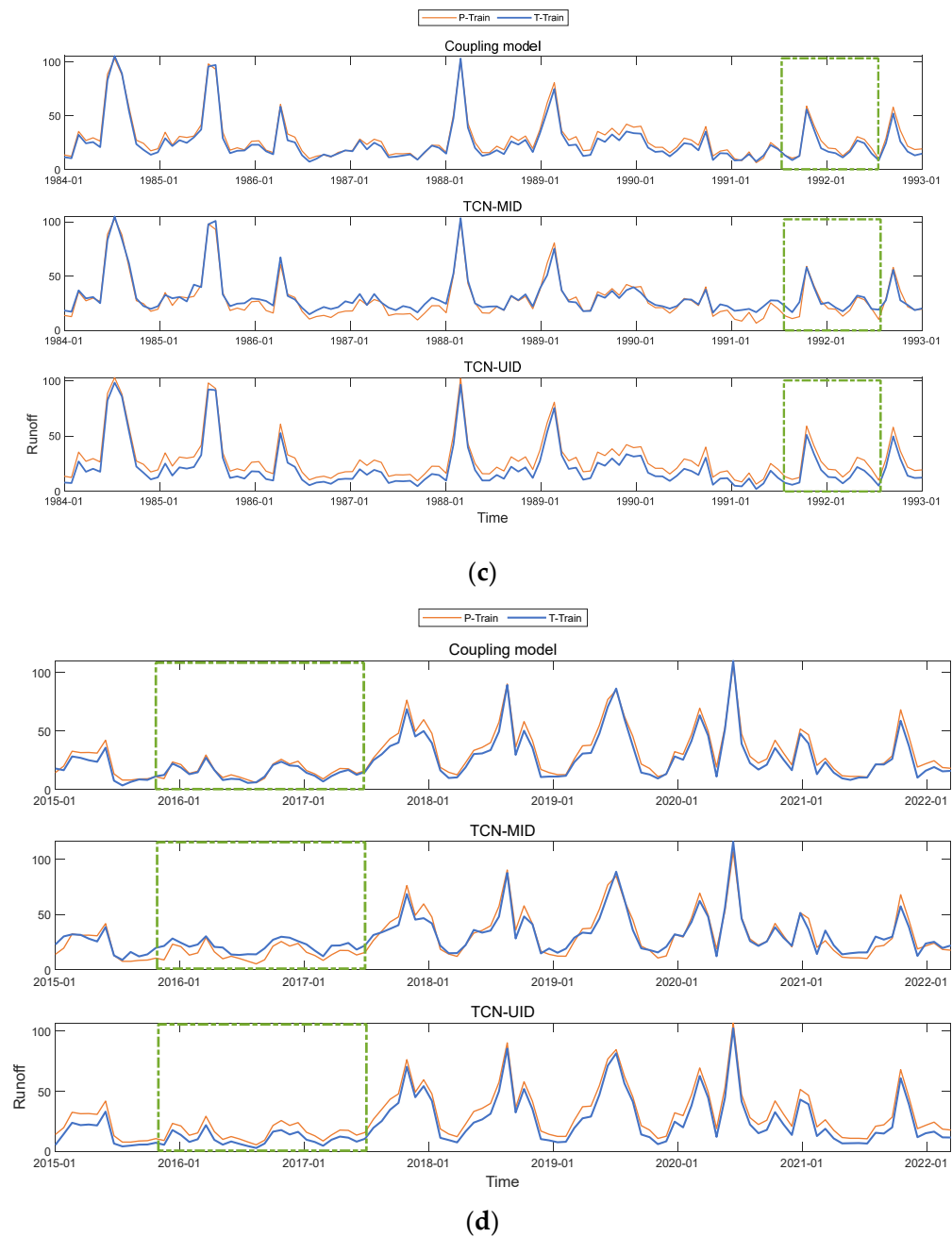
**Figure 16.** *Cont.*

**(c)**



**(d)**
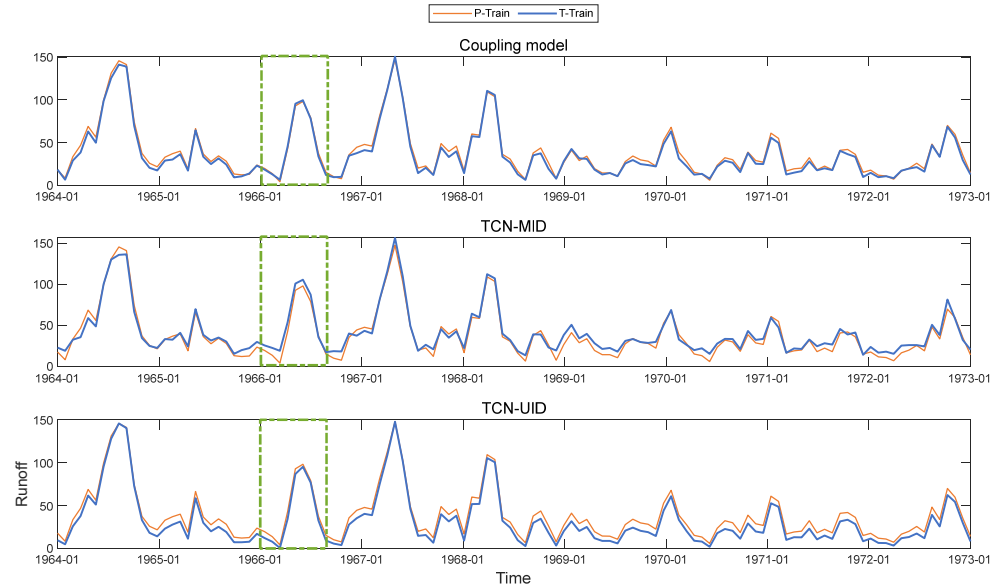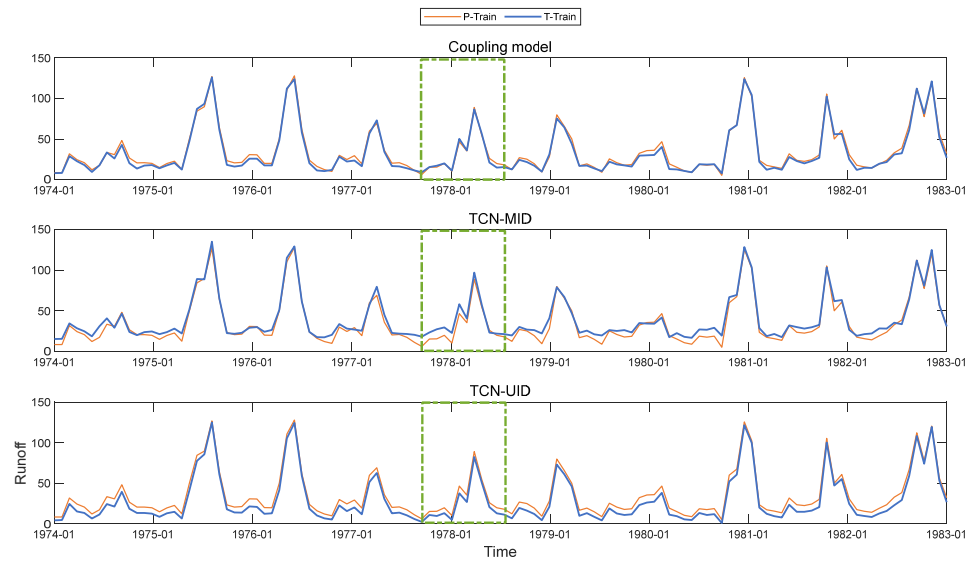
**Figure 16.** Comparison of prediction effect of three models. (**a**–**d**) shows that the TCN-UID model has poor prediction effect at the runoff minimum, and the TCN-MID model has poor prediction effect at the runoff maximum, and the coupling model has the best prediction effect, which can well adapt to the nonlinear and non-stationary nature of runoff.

In contrast, the coupling model accurately predicts these extreme values, demonstrating robustness in capturing the nonlinear and non-stationary characteristics of runoff sequences. It exhibits strong performance even at extreme values, indicating its ability to mitigate uncertainties and adapt well to varying runoff conditions.

## 5. Discussion

### 5.1. The Impact of Climate Change and Human Activities

From 1970 to 2018, drought in the Yellow River Basin began to rise, making it one of the most drought-affected basins in China [54]. The Yellow River Basin is significantly influenced by various factors, such as ecosystem conditions, agricultural and pastoral

development, energy resources, economic composition, population distribution, intensity of human activities, and runoff utilization patterns within different sections of the basin, making it a typical basin sensitive to environmental changes [55]. The region suffers from severe water shortages, highly uneven precipitation distribution, and frequent extreme weather events such as floods and droughts, which seriously hinder the socioeconomic development, ecological protection, and high-quality development of the Yellow River Basin. Additionally, global warming has intensified the El Niño-Southern Oscillation (ENSO) phenomenon, leading to highly uneven spatial and temporal precipitation distribution in the basin, increasing the likelihood of severe droughts [56]. Between 1994 and 2003, there was a significant decreasing trend in runoff in the Yellow River Basin, with the overall runoff volume gradually decreasing [57]. Therefore, runoff prediction plays a crucial role in ensuring water security in the Yellow River Basin [58].

Additionally, beyond climate change, human activities also impact runoff development. The continuous population growth and rapid economic development have led to increased industrial and agricultural production and water resource consumption [59]. This has resulted in over-exploitation of water resources and a series of ecological degradation phenomena, such as river flow interruption, reduced vegetation cover, and declining groundwater levels, which in turn affect runoff development in the Yellow River Basin [39].

*5.2. A Coupled Knowledge-Embedded and Data-Driven Runoff Prediction Model*

Using transfer entropy to identify runoff driving factors, results show that rainfall, sunshine duration, and relative humidity are the primary influencing factors, consistent with He's findings [60]. Employing multivariable input for runoff prediction alleviates the limitations of using a single variable, as noted by Lin [61]. The IPSO algorithm proposed in this study automatically optimizes hyperparameters, significantly saving time and enhancing model performance, aligning with Lin's findings. Jin [62] highlighted that due to global warming and human activities, runoff exhibits nonlinear and non-stationary characteristics, making it difficult for traditional models to adapt. Combining data-driven models with knowledge reduces severe data dependency. The results indicate that the coupled knowledge-embedded and data-driven runoff prediction model, which integrates runoff probability density as prior knowledge into the TCN training, accurately predicts runoff and addresses poor peak value prediction in deep learning models [63]. The coupled model considers time-series data and prior knowledge, capturing the nonlinear characteristics of runoff changes through deep learning. Furthermore, it breaks the "black box" nature of neural networks, with domain knowledge improving model accuracy, consistent with Wang's findings [64].

Despite the advantages of the IPSO-TCN model in handling nonlinear relationships and large datasets, we also acknowledge its limitations, particularly its high dependency on data quality and the complexity of model parameter tuning [65]. Overall, this study provides deeper insights into runoff prediction, which is expected to positively impact the development of more targeted drought management and response strategies.

*5.3. Advantages and Limitations*

Despite the positive results achieved by the coupled model in various aspects, it still faces several challenges. The model requires a substantial amount of data for training, which may limit its practical application, especially in remote or data-scarce regions [31]. Future research could incorporate other physical laws, control equations, or expert knowledge into the proposed model framework to address runoff prediction issues or other related engineering problems, such as wind power forecasting [30]. Overall, this study presents a novel and effective method for runoff prediction, offering new perspectives not only for runoff forecasting but also for predicting other factors such as temperature, humidity, and wind speed.

With the advancement of science and technology and societal progress, climate factors no longer solely influence runoff occurrences; human activities are increasingly playing a

significant role in runoff variations [66]. Future research aims to include analyses of human impact factors, such as land use, the South-to-North Water Diversion Project, and reservoir dams, to explore ways to achieve a win–win situation for both social development and ecological protection [67].

## 6. Conclusions

Against the backdrop of intensified climate change and human activities, runoff patterns have undergone significant changes, posing major challenges to the ecological environment protection and sustainable development of the Yellow River Basin. This study selected 60 years of monthly runoff (R) data from January 1964 to December 2023 at nine main hydrological stations along the main stream of the Yellow River, along with data from nine meteorological variables. Preliminary GIS analysis was conducted to analyze runoff variations in the Yellow River basin over the past 60 years, followed by the calculation of the Standardized Runoff Index (SRI). The hydrological station with the highest frequency of droughts was chosen as the primary research focus. The transfer entropy method was employed to select runoff's influencing factors from the nine meteorological variables. The Improved Particle Swarm Optimization (IPSO) method was used for model parameter optimization. The sliding window step was determined through cross-wavelet transform. Experimental comparisons among the TCN-UID, TCN-MID, and coupling model were conducted, and the research findings are as follows:

(1) From 1964 to 1983, the overall runoff in the Yellow River basin remained stable, but it gradually decreased starting from 1984. The period from 1994 to 2003 had the lowest runoff in nearly 60 years. Over this period, droughts in the Yellow River basin showed a trend of shifting from upstream to downstream.

(2) The primary cycle of drought in the Yellow River during the study period was 10–14 months, with hydrological drought lagging behind meteorological drought by 2 months.

(3) RF (rainfall), HS (hours of sunshine), and RH (relative humidity) are the three main driving factors of runoff.

(4) Using IPSO for model parameter optimization improved the model's prediction accuracy. In model evaluation metrics, the coupling model outperformed the TCN-UID and TCN-MID models in terms of MAE (Mean Absolute Error), RMSE (Root Mean Square Error), and Nash–Sutcliffe Efficiency Coefficient (NSE), effectively capturing the nonlinear and non-stationary characteristics of runoff sequences.

(5) By constructing a loss function based on the runoff probability density function, a knowledge-embedded and data-driven runoff prediction model was established. This approach breaks the traditional reliance on data and eliminates barriers between knowledge and data. Compared to the data-driven model (TCN-MID), the coupling model shows performance improvements of 6.9% and 4.7% on the training set and 5.7% and 2.8% on the test set. The coupling model not only benefits from data-driven advantages but also effectively addresses the issue of poor prediction performance at extreme values, enhancing the accuracy of runoff predictions.

The IPSO algorithm designed in this study enables automatic hyperparameter optimization, significantly reducing time costs and enhancing model prediction accuracy. The IPSO algorithm is highly portable and can be integrated with any deep learning method. Comparative experiments with TCN-UID, TCN-MID, and the coupling model demonstrate that not only does multivariable input improve model prediction accuracy, but it also strongly affirms the necessity of incorporating prior knowledge into data-driven models. These findings provide a basis for decision making in ecological protection and high-quality development in the Yellow River Basin. Considering the probabilistic density knowledge inherent in runoff, developing feasible data-driven models based on prior knowledge remains a crucial topic for future research.

## Abbreviations

This table lists all the abbreviations and their full forms used in this study.

| Abbreviation | Full Term |
| --- | --- |
| AP | atmospheric pressure |
| BiLSTM | Bidirectional Long Short-Term Memory |
| CMA | China Meteorological Administration |
| CNN | Convolutional Neural Networks |
| DDD | Distance Dynamic Model |
| GRU | Gated Recurrent Units |
| HS | Hours of sunshine |
| IPSO | Improved Particle Swarm Optimization |
| LS-SVM | Least Squares Support Vector Machine |
| LSTM | Long Short-Term Memory |
| MAE | Mean Absolute Error |
| MLE | Maximum Likelihood Estimation |
| NMIC | National Meteorological Information Center |
| NSE | Nash-Sutcliffe Efficiency |
| PDEs | partial differential equations |
| PDF | probability density function |
| PGNN | physical-guided neural network |
| PINN | Physics-Informed Neural Networks |
| PSO | Particle Swarm Optimization |
| R | runoff |
| RF | rainfall |
| RH | Relative humidity |
| RMSE | Root Mean Squared Error |
| SP | Sunshine percentage |
| SPI | Standardized Precipitation Index |
| SRI | Standardized Runoff Index |
| SVM | Support Vector Machines |
| T | Temperature |
| TA | Temperature anomaly |
| TCN | Temporal Convolutional Network |
| TCN-MID | multivariable input models |
| TCN-UID | univariate input models |
| TE | Transfer entropy |
| TgFCNN | theory-guided fully convolutional neural network |
| TgNN | Theory-Guided Neural Network |
| treeLSTM | Tree-State Long Short-Term Memory |
| WTC | Cross-wavelet transform |
| WV | Wind velocity |
| XGB | Extreme Gradient Boosting |
| XTC | Wavelet coherence |
| VP | Vapor pressure |

# References

1. Ni, Y.; Lv, X.; Yu, Z.; Wang, J.; Ma, L.; Zhang, Q. Intra-annual variation in the attribution of runoff evolution in the Yellow River source area. *Catena* **2023**, *225*, 107032. [CrossRef]
2. Yang, L.; Zhao, G.; Tian, P.; Mu, X.; Tian, X.; Feng, J.; Bai, Y. Runoff changes in the major river basins of China and their responses to potential driving forces. *J. Hydrol.* **2022**, *607*, 127536. [CrossRef]
3. Gan, R.; Xu, M.; Yang, F.; Zuo, Q.; Zhang, X. The assessment of baseflow separation method and baseflow characteristics in the Yiluo River basin, China. *Environ. Earth Sci.* **2022**, *81*, 323. [CrossRef]
4. Valjarević, A. GIS-Based Methods for Identifying River Networks Types and Changing River Basins. *Water Resour. Manag.* **2024**, *51*, 146–168. [CrossRef]
5. Wang, Q.; Shao, W.; Guan, Q.; Sun, Y.; Du, Q.; Zhang, E.; Yan, Y.; Yang, X. The structural equation modeling constructed for runoff change attribution analysis outperforms traditional methods. *J. Hydrol.* **2024**, *636*, 131317. [CrossRef]
6. Chen, T.; Zou, L.; Xia, J.; Liu, H.; Wang, F. Decomposing the impacts of climate change and human activities on runoff changes in the Yangtze River Basin: Insights from regional differences and spatial correlations of multiple factors. *J. Hydrol.* **2022**, *61*, 128649. [CrossRef]
7. Ricart, S.; Rico, A.M. Assessing technical and social driving factors of water reuse in agriculture: A review on risks, regulation and the yuck factor. *Agric. Water Manag.* **2019**, *217*, 426–439. [CrossRef]
8. Lian, Y.; Sun, M.; Wang, J.; Luan, Q.; Jiao, M.; Zhao, X.; Gao, X. Quantitative impacts of climate change and human activities on the runoff evolution process in the Yanhe River Basin. *Phys. Chem. Earth Parts A/B/C* **2021**, *122*, 102998. [CrossRef]
9. Zhao, G.; Kondolf, G.M.; Mu, X.; Han, M.; He, Z.; Rubin, Z.; Wang, F.; Gao, P.; Sun, W. Sediment yield reduction associated with land use changes and check dams in a catchment of the Loess Plateau, China. *Catena* **2017**, *148*, 126–137. [CrossRef]
10. Shi, H.; Wang, G. Impacts of climate change and hydraulic structures on runoff and sediment discharge in the middle Yellow River. *Hydrol. Process.* **2015**, *29*, 3236–3246. [CrossRef]
11. Wan, Y.; Konyha, K. A simple hydrologic model for rapid prediction of runoff from ungauged coastal catchments. *J. Hydrol.* **2015**, *528*, 571–583. [CrossRef]
12. Perkins, J.R.; Fraser, A.J.; Muxworthy, A.R.; Neumaier, M.; Schenk, O. Basin and petroleum systems modelling to characterise multi-source hydrocarbon generation: A case study on the inner Moray Firth, UK North Sea. *Mar. Pet. Geol.* **2023**, *151*, 106180. [CrossRef]
13. Furlani, S.; Antonioli, F. The swim-survey archive of the Mediterranean rocky coasts: Potentials and future perspectives. *Geomorphology* **2023**, *421*, 108529. [CrossRef]
14. Zang, S.; Li, Z.; Zhang, K.; Yao, C.; Liu, Z.; Wang, J.; Huang, Y.; Wang, S. Improving the flood prediction capability of the Xin'anjiang model by formulating a new physics-based routing framework and a key routing parameter estimation method. *J. Hydrol.* **2021**, *603*, 126867. [CrossRef]
15. Peng, D.; Zhijia, L.; Fan, X. Application of TOPMODEL in Buliu River Basin and comparison with Xin'anjiang model. *Water Sci. Eng.* **2008**, *1*, 25–32. [CrossRef]
16. Cai, Y.; Zhang, F.; Shi, J.; Carl Johnson, V.; Ahmed, Z.; Wang, J.; Wang, W. Enhancing SWAT model with modified method to improve Eco-hydrological simulation in arid region. *J. Clean. Prod.* **2023**, *403*, 136891. [CrossRef]
17. Muthanna, T.M.; Alfredsen, K.; Skaugen, T.; Tsegaw, A.T. A dynamic river network method for the prediction of floods using a parsimonious rainfall-runoff model. *Hydrol. Res.* **2020**, *51*, 146–168. [CrossRef]
18. Querales, M.; Salas, R.; Morales, Y.; Allende-Cid, H.; Rosas, H. A stacking neuro-fuzzy framework to forecast runoff from distributed meteorological stations. *Appl. Soft Comput.* **2022**, *118*, 108535. [CrossRef]
19. 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*, 130091. [CrossRef]
20. Xie, Y.; Sun, W.; Ren, M.; Chen, S.; Huang, Z.; Pan, X. Stacking ensemble learning models for daily runoff prediction using 1D and 2D CNNs. *Expert Syst. Appl.* **2023**, *21*, 119469. [CrossRef]
21. Gao, S.; Huang, Y.; Zhang, S.; Han, J.; Wang, G.; Zhang, M.; Lin, Q. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *J. Hydrol.* **2020**, *589*, 125188. [CrossRef]
22. Samantaray, S.; Sahoo, A.; Prakash Satapathy, D. Improving accuracy of SVM for monthly sediment load prediction using Harris hawks optimization. *Mater. Today Proc.* **2022**, *61*, 604–617. [CrossRef]
23. Ni, L.; Wang, D.; Wu, J.; Wang, Y.; Tao, Y.; Zhang, J.; Liu, J. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *J. Hydrol.* **2020**, *586*, 124901. [CrossRef]
24. Xiong, J.; Peng, T.; Tao, Z.; Zhang, C.; Song, S.; Nazir, M.S. A dual-scale deep learning model based on ELM-BiLSTM and improved reptile search algorithm for wind power prediction. *Energy* **2023**, *266*, 126419. [CrossRef]
25. Guo, J.; Liu, Y.; Zou, Q.; Ye, L.; Zhu, S.; Zhang, H. Study on optimization and combination strategy of multiple daily runoff prediction models coupled with physical mechanism and LSTM. *J. Hydrol.* **2023**, *624*, 129969. [CrossRef]
26. 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*, 129977. [CrossRef]
27. Liu, Y.; Ji, Y.; Liu, D.; Fu, Q.; Li, T.; Hou, R.; Li, Q.; Cui, S.; Li, M. A new method for runoff prediction error correction based on LS-SVM and a 4D copula joint distribution. *J. Hydrol.* **2021**, *598*, 126223. [CrossRef]

28. Chen, Z.; Lin, H.; Shen, G. TreeLSTM: A spatiotemporal machine learning model for rainfall-runoff estimation. *J. Hydrol. Reg. Stud.* **2023**, *48*, 101474. [CrossRef]

29. Dong, Y.; Su, H.; Zhu, J.; Bao, F. Towards Interpretable Deep Neural Networks by Leveraging Adversarial Examples. *arXiv* **2017**, arXiv:abs/1901.09035.

30. Wang, N.; Zhang, D.; Chang, H.; Li, H. Deep learning of subsurface flow via theory-guided neural network. *J. Hydrol.* **2020**, *584*, 124700. [CrossRef]

31. Jing, X.; Luo, J.; Zuo, G.; Yang, X. Interpreting runoff forecasting of long short-term memory network: An investigation using the integrated gradient method on runoff data from the Han River Basin. *J. Hydrol. Reg. Stud.* **2023**, *50*, 101549. [CrossRef]

32. Luo, X.; Zhang, D.; Zhu, X. Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge. *Energy* **2021**, *225*, 120240. [CrossRef]

33. Karpatne, A.; Watkins, W.; Read, J.S.; Kumar, V. Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling. *arXiv* **2017**, arXiv:abs/1710.11431.

34. He, T.; Wang, N.; Zhang, D. Theory-guided full convolutional neural network: An efficient surrogate model for inverse problems in subsurface contaminant transport. *Adv. Water Resour.* **2021**, *157*, 104051. [CrossRef]

35. Chen, Y.; Huang, D.; Zhang, D.; Zeng, J.; Wang, N.; Zhang, H.; Yan, J. Theory-guided hard constraint projection (HCP): A knowledge-based data-driven scientific machine learning method. *J. Comput. Phys.* **2021**, *445*, 110624. [CrossRef]

36. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]

37. Ni, Y.; Yu, Z.; Lv, X.; Qin, T.; Yan, D.; Zhang, Q.; Ma, L. Spatial difference analysis of the runoff evolution attribution in the Yellow River Basin. *J. Hydrol.* **2022**, *612*, 128149. [CrossRef]

38. Zhao, A.; Wang, D.; Xiang, K.; Zhang, A. Vegetation photosynthesis changes and response to water constraints in the Yangtze River and Yellow River Basin, China. *Ecol. Indic.* **2022**, *39*, 109331. [CrossRef]

39. Zhou, X.; Chen, W.; Liu, Q.; Shen, H.; Cai, S.; Lei, X. Future runoff forecast in Hanjiang River Basin based on Wetspa model and CMIP6 model. *Front. Environ. Sci.* **2022**, *10*, 980949. [CrossRef]

40. Kang, L.; Wen, Y.; Zhou, L.; Chen, H.; Ye, J. Drought driving mechanism and risk situation prediction based on machine learning models in the Yellow River Basin, China. *Geomat. Nat. Hazards Risk* **2023**, *14*, 2279493. [CrossRef]

41. Kong, D.; Miao, C.; Wu, J.; Duan, Q. Impact assessment of climate change and human activities on net runoff in the Yellow River Basin from 1951 to 2012. *Ecol. Eng.* **2016**, *91*, 566–573. [CrossRef]

42. Wang, Q.; Sun, Y.; Guan, Q.; Du, Q.; Zhang, Z.; Zhang, J.; Zhang, E. Exploring future trends of precipitation and runoff in arid regions under different scenarios based on a bias-corrected CMIP6 model. *J. Hydrol.* **2024**, *630*, 130666. [CrossRef]

43. Shi, C.; Zhou, Y.; Fan, X.; Shao, W. A study on the annual runoff change and its relationship with water and soil conservation practices and climate change in the middle Yellow River basin. *Catena* **2013**, *100*, 31–41. [CrossRef]

44. Belayneh, A.; Adamowski, J.; Khalil, B.; Ozga-Zielinski, B. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *J. Hydrol.* **2014**, *508*, 418–429. [CrossRef]

45. Foroumandi, E.; Nourani, V.; Dąbrowska, D.; Kantoush, S.A. Linking Spatial–Temporal Changes of Vegetation Cover with Hydroclimatological Variables in Terrestrial Environments with a Focus on the Lake Urmia Basin. *Land* **2022**, *11*, 115. [CrossRef]

46. Ghaderpour, E.; Ince, E.S.; Pagiatakis, S.D. Least-squares cross-wavelet analysis and its applications in geophysical time series. *J. Geod.* **2018**, *92*, 1223–1236. [CrossRef]

47. Bai, S.; Zico, K.J.; Vladlen, K. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:180301271.

48. Zhao, P.; Fan, Z.; Cao, Z.; Li, X. Intrusion Detection Model Using Temporal Convolutional Network Blend Into Attention Mechanism. *Int. J. Inf. Secur. Priv.* **2021**, *16*, 1–20. [CrossRef]

49. Zhou, D.; Wang, B. Battery health prognosis using improved temporal convolutional network modeling. *J. Energy Storage* **2022**, *51*, 104480. [CrossRef]

50. Wan, R.; Mei, S.; Wang, J.; Liu, M.; Yang, F. Multivariate Temporal Convolutional Network: A Deep Neural Networks Approach for Multivariate Time Series Forecasting. *Electronics* **2019**, *8*, 876. [CrossRef]

51. Zhou, L.; Kang, L. A Comparative Analysis of Multiple Machine Learning Methods for Flood Routing in the Yangtze River. *Water* **2023**, *15*, 1556. [CrossRef]

52. Chen, Y.; Zhang, D. Integration of knowledge and data in machine learning. *arXiv* **2022**, arXiv:2202.10337v2. [CrossRef]

53. Merabti, A.; Martins, D.S.; Meddi, M.; Pereira, L.S. Spatial and Time Variability of Drought Based on SPI and RDI with Various Time Scales. *Water Resour. Manag.* **2017**, *32*, 1087–1100. [CrossRef]

54. Wang, F.; Wang, Z.; Yang, H.; Zhao, Y. Study of the temporal and spatial patterns of drought in the Yellow River basin based on SPEI. *Sci. China Earth Sci.* **2018**, *61*, 1098–1111. [CrossRef]

55. Guo, Y.; Ming, B.; Huang, Q.; Wang, Y.; Zheng, X.; Zhang, W. Risk-averse day-ahead generation scheduling of hydro–wind–photovoltaic complementary systems considering the steady requirement of power delivery. *Appl. Energy* **2022**, *309*, 118467. [CrossRef]

56. Wang, Y.; Ding, Y.; Ye, B.; Liu, F.; Wang, J.; Wang, J. Contributions of climate and human activities to changes in runoff of the Yellow and Yangtze rivers from 1950 to 2008. *Sci. China Earth Sci.* **2012**, *56*, 1398–1412. [CrossRef]

57. Cui, J.; Jian, S. Spatiotemporal Variation of Runoff and Its Influencing Factors in the Yellow River Basin, China. *Water* **2023**, *15*, 2058. [CrossRef]

58. Wang, Y.; Wang, S.; Wang, C.; Zhao, W. Runoff sensitivity increases with land use/cover change contributing to runoff decline across the middle reaches of the Yellow River basin. *J. Hydrol.* **2021**, *600*, 126536. [CrossRef]

59. Zhao, P.; Lü, H.; Fu, G.; Zhu, Y.; Su, J.; Wang, J. Uncertainty of Hydrological Drought Characteristics with Copula Functions and Probability Distributions: A Case Study of Weihe River, China. *Water* **2017**, *9*, 334. [CrossRef]

60. He, Y.; Mu, X.; Jiang, X.; Song, J. Runoff Variation and Influencing Factors in the Kuye River Basin of the Middle Yellow River. *Front. Environ. Sci.* **2022**, *10*, 877535. [CrossRef]

61. Lin, Y.; Wang, D.; Meng, Y.; Sun, W.; Qiu, J.; Shangguan, W.; Cai, J.; Kim, Y.; Dai, Y. Bias learning improves data driven models for streamflow prediction. *J. Hydrol. Reg. Stud.* **2023**, *50*, 101557. [CrossRef]

62. Jin, J.; Zhang, Y.; Hao, Z.; Xia, R.; Yang, W.; Yin, H.; Zhang, X. Benchmarking data-driven rainfall-runoff modeling across 54 catchments in the Yellow River Basin: Overfitting, calibration length, dry frequency. *J. Hydrol. Reg. Stud.* **2022**, *42*, 101119. [CrossRef]

63. de la Fuente, A.; Meruane, V.; Meruane, C. Hydrological Early Warning System Based on a Deep Learning Runoff Model Coupled with a Meteorological Forecast. *Water* **2019**, *11*, 1808. [CrossRef]

64. Wang, H.; Qin, H.; Liu, G.; Huang, S.; Qu, Y.; Qi, X.; Zhang, Y. Hierarchical attention network for short-term runoff forecasting. *J. Hydrol.* **2024**, *638*, 131549. [CrossRef]

65. Huang, Q.; Wang, X. A Forecasting Model of Wind Power Based on IPSO–LSTM and Classified Fusion. *Energies* **2022**, *15*, 5531. [CrossRef]

66. Foroumandi, E.; Nourani, V.; Kantoush, S.A. Investigating the main reasons for the tragedy of large saline lakes: Drought, climate change, or anthropogenic activities? A call to action. *J. Arid Environ.* **2022**, *196*, 104652. [CrossRef]

67. Gao, Y.; Fu, S.; Cui, H.; Cao, Q.; Wang, Z.; Zhang, Z.; Wu, Q.; Qiao, J. Identifying the spatio-temporal pattern of drought characteristics and its constraint factors in the Yellow River Basin. *Ecol. Indic.* **2023**, *154*, 110753. [CrossRef]