TU Delft

Delft University of Technology

A Multi-Task Learning Based Runoff Forecasting Model for Multi-Scale Chaotic Hydrological Time Series

Zuo, Hui; Yan, Gaowei; Lu, Ruochen; Li, Rong; Xiao, Shuyi; Pang, Yusong

DOI
10.1007/s11269-023-03681-z

Publication date
2023

Document Version
Final published version

Published in
Water Resources Management

Citation (APA)
Zuo, H., Yan, G., Lu, R., Li, R., Xiao, S., & Pang, Y. (2023). A Multi-Task Learning Based Runoff Forecasting Model for Multi-Scale Chaotic Hydrological Time Series. *Water Resources Management*, *38 (2024)*(2), 481-503. https://doi.org/10.1007/s11269-023-03681-z

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Water Resources Management
https://doi.org/10.1007/s11269-023-03681-z

A Multi-Task Learning Based Runoff Forecasting Model for Multi-Scale Chaotic Hydrological Time Series

Hui Zuo[1] · Gaowei Yan[1] · Ruochen Lu[1] · Rong Li[1] · Shuyi Xiao[1] · Yusong Pang[2]

Received: 10 August 2023 / Accepted: 23 November 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract
Accurately predicting runoff is crucial for managing water resources, preventing and mitigating floods, scheduling hydropower plant operations, and protecting the environment. The hydrological dynamic composite system that forms runoff is complex and random, and seemingly random behavior may be caused by nonlinear variables in a simple deterministic system, which poses a challenge to runoff prediction. In this paper, we construct parallel and multi-timescale reservoirs from a chaotic theory perspective to simulate the stochasticity of chaotic systems. We propose a multi-task-based "Decomposition-Integration-Prediction" (Multi-SDIPC) model for runoff prediction. To validate our research results, we use the Catchment Attributes and Meteorology for Large-Sample Studies (CAMELS) dataset and compare our proposed model with 10 baseline models. The results show that our model has an average NSE metric of 0.83 and exhibits higher accuracy, better generalization, and greater stability than the other models in multi-step forecasting. Based on our findings, we recommend wider application of the Multi-SDIPC model in different regions of the world for medium or long-term runoff prediction.

Keywords Runoff prediction · Chaos theory · Reservoir computing · Multi-task learning · Convolutional neural network

✉ Gaowei Yan
yangaowei@tyut.edu.cn

Hui Zuo
zuohui092@163.com

Ruochen Lu
ruochen.lu@uq.net.au

Rong Li
lirong@tyut.edu.cn

Shuyi Xiao
xiaoshuyi@tyut.edu.cn

Yusong Pang
Y.Pang@tudelft.nl

[1] College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China

[2] Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft 2628CD, Holland

Published online: 30 November 2023

1 Introduction

The changing global climate and the impact of human activities have led to changes in rainfall patterns, making it more difficult to make decisions on measures such as flood warnings, flood prevention and mitigation, and water resource allocation. Therefore, it is crucial to establish an accurate and reliable runoff prediction model. Currently, runoff prediction models can be categorized into physically driven models and machine learning models. Physically driven models that use equations constrained by boundary conditions to simulate the physical and interacting processes of hydrological systems require significant a priori knowledge, leading to constraints on their application (Yaseen et al. 2019). Machine learning models based on statistics or deep learning can learn the complex relationship between variables only through a large amount of data (Gao et al. 2020; Adnan et al. 2021; Zhou et al. 2022; Malakoutian et al. 2022; Pérez-Alarcón et al. 2022). This type of model is often a black-box model, which needs to consider multiple aspects such as data quality, and feature engineering, and lacks physical interpretability in the face of the water imbalance problem. However, it is still widely used in the field of hydrology due to its high efficiency, low cost, and easy-to-implement advantages (Yin et al. 2022).

The formation of runoff is closely related to the spatial and temporal distribution of meteorological conditions, surface soil factors, and the operation of upstream water conservancy projects, forming a chaotic hydrological dynamics composite system with huge spatial and temporal variability, complex evolutionary laws, whose chaotic characteristics pose a great challenge to runoff prediction, and the traditional machine learning methods are unable to fulfill the prediction needs (Sivakumar 2000). Machine learning models that take chaos theory into account can effectively simulate chaotic patterns that may exist in real systems and have made some progress in solving hydrological problems on a global scale (Duan et al. 2020; Giri and Devercelli 2023). For example, Yu-tong et al. (2019) utilized chaotic wavelet neural networks for daily and monthly runoff flow prediction. Han et al. (2019) proposed an improved kernel recursive least squares algorithm for online prediction of multivariate chaotic time series. However, the above methods fail to produce satisfactory results in predicting chaotic hydrological time series and in distinguishing chaotic and random data, so reservoir computing (RC) was proposed and proved to be effective in predicting the trend of chaotic time series (Griffith et al. 2019; Vlachas et al. 2020). RC regarded a chaotic time series as a series of superpositions of random perturbations and deterministic trends, where the random perturbations obey a certain probability distribution, and the deterministic trends can be fitted with a linear regression model. It improves the prediction accuracy of chaotic time series by continuously updating and adjusting the parameters of the reservoir of random perturbations. Arcomano et al. (2020) applied RC to global atmospheric prediction models and demonstrated that parallel machine learning models based on RC can be applied to other geophysical hydrodynamic systems. Gupta et al. (2023) captured the nonlinearity of complex systems by adding radial basis functions (RBF) to the RC-based recurrent neural networks model to achieve model-free prediction of turbulent systems. Inspired by this, RC is introduced in this paper to predict chaotic runoff time series. However, most studies have focused on single-scale chaotic features, ignoring the existence of multi-scale chaotic features in hydrological time series. Only by studying as many features as possible on time scales can we truly understand the hydroclimatic and runoff processes.

To avoid ignoring the trend, season, and residual in the prediction of the model, scholars proposed a hybrid model combining the runoff series decomposition method

with a single model. Zhou and Dong (2012) used the X-12-ARIMA decomposition method to forecast the monthly and quarterly series of China's net crude oil imports for the past 16 years. However, this decomposition method is based on the assumption of the linear relationship, which cannot solve the complex chaotic nonlinear system problem well. Therefore, He et al. (2022) integrated seasonal-trend decomposition using loess (STL) and machine learning methods for predicting rainfall time series one step ahead based on historical rainfall and other meteorological data. The STL method has stronger robustness and is not disturbed by the outliers in the data, which has better application prospects in runoff time series prediction (McKitrick and Christy 2019), so this paper constructs a multi-timescale reservoir to study chaotic runoff time series. At the same time, RC in chaotic runoff prediction tends to ignore the existence of trend changes, periodic variations, and noise interference problems within its sequence. To address this problem, some scholars have proposed hybrid models that combine runoff series decomposition methods with a single model to fully extract the useful information hidden in the complex runoff series. Li et al. (2022) used variational modal decomposition (VMD) coupled with Gray Wolf Optimizer-based LSTM to implement monthly runoff prediction. However, this decomposition method is sensitive to noise, so He et al. (2022) integrated the STL and machine learning methods for predicting rainfall time series one step ahead based on historical rainfall and other meteorological data. The STL method has stronger robustness and is not disturbed by the outliers in the data, which has better application prospects in runoff time series prediction.

Prediction models based on time series decomposition can affect runoff prediction model performance due to the choice of prediction framework. For example, He et al. (2020) proposed a hybrid model VMD-GBRT based on VMD and gradient boosting regression (GBRT) and applied it to monthly runoff forecasting. This prediction framework is collectively called the "Decomposition-Prediction-Reconstruction" (DPR) framework, which can lead to degradation of the final runoff prediction performance due to the accumulation of prediction errors in each subseries. Subsequently, Xu et al. (2021) proposed a "Decomposition-Integration-Prediction" (DIP) framework, in which the decomposed subsequence of the original runoff sequence is input to multiple parallel neural networks, and then the output of each neural network is adaptively concatenated to obtain the prediction results. Although this prediction framework has an error correction function, the dependencies between high-dimensional complex chaotic features are difficult to extract effectively, which will lead to the degradation of the prediction performance of the model. Therefore, this paper proposes using convolutional neural networks with multiple convolutional kernels to obtain the complex correlations between high-dimensional chaotic features and ensure the diversity of extracted features (Ma et al. 2021).

Consider using knowledge from other hydrological time series to improve prediction and communicate high-dimensional chaotic information effectively. Multi-task learning is proposed to achieve shared information and knowledge transfer in the process of problem-solving to improve the overall performance of the target model. Zhang and Yang (2017) demonstrated that joint learning of multiple tasks leads to better performance empirically and theoretically than independent learning. This paper proposes a multi-task learning approach to improve runoff predictions. It uses a subtask reservoir to obtain high-dimensional chaotic information and a convolutional neural network to extract and transfer knowledge for enhanced performance.

To address the challenges brought by the complexity and randomness of hydrological dynamic composite systems to runoff prediction, and taking into account the multi-scale characteristics of hydrological time series, this paper proposes a multi-task based

Multi-SDIPC model, which simulates the randomness of chaotic systems by constructing parallel and multi-time scale reservoirs. The main contributions are summarized as follows:

1. Stochastic high-dimensional mapping of RC is used to simulate chaotic patterns to capture the nonlinear dynamic features in chaotic hydrological time series and convert them into predictable linear dynamic features, thereby predicting their future trends to improve the prediction model performance.
2. Extracting chaotic multi-timescale features by constructing sub-task multi-timescale reservoirs to enhance the original runoff series for useful knowledge transfer.
3. Extensive experiments with 10 baseline models on the CAMELS dataset.

2 Related Work

2.1 Seasonal-Trend Decomposition Procedures Based on LOESS(STL)

The STL is a robust method that does not require mathematical modeling and is widely used to decompose time series with different features (Theodosiou 2011; Mohsin et al. 2021). It consists of an internal and external iterative loop process. As shown in Fig. 1, the inner loop implements updating the seasonal and trend components to make them smooth; the outer loop implements updating the robustness weights of the residual components. It is assumed that the trend component and the cycle component after the k-1th iteration in the inner loop are: $T_v^{(k)}$, $S_v^{(k)}$, initially $T_v^{(k)} = 0$; and the following parameters: $n(p)$ is the number of samples in one cycle, and $n(s)$, $n(l)$, $n(t)$ are LOESS smoothing parameters in Step2, Step3, Step6 respectively. The specific steps of the inner loop are presented as follows:

Step 1: Detrending. Subtract the trend component of the previous iteration by $Y_v - T_v^{(k)}$.
Step 2: Cycle-subseries smoothing. Regress each subseries using LOESS of $q = n_{(s)}$, $d = 1$, with a delay of 1-time point before and after, and combine to obtain a time series $C_v^{(k+1)}$ of length $(N + 2 \times n_p)$.
Step 3: Low-pass filtering of smoothing cycle-subseries. The sequence of results $C_v^{(k+1)}$ of Step 2 is done as a sliding average of $n(p)$, $n(p)$, 3 in turn, and then the sequence $L_v^{(k+1)}$ is obtained by regression with LOESS of $q = n_{(l)}$, $d = 1$.
Step 4: Detrending of smoothed cycle-subseries. $S_v^{(k+1)} = C_v^{(k+1)} - L_v^{(k+1)}$.
Step 5: Deseasonalizing.$Y_v - S_v^{(k+1)}$.
Step 6: Trend Smoothing. Regression of the resulting series of Step 5 using $q = n_{(t)}$, $d = 1$ was performed to obtain the trend component $T_v^{(k+1)}$.

The iterations of the outer loop are used by the results of the iterations of the inner loop (trend T_v and seasonal S_v) to calculate the robustness weights, while the residuals can be obtained by $R_v = Y_v - T_v - S_v$. First definition,

$$h = 6 \times median(|R_v|) \tag{1}$$

Then, the robust weights for time point v are:

$$\rho_v = B(|R_v|/h) \tag{2}$$

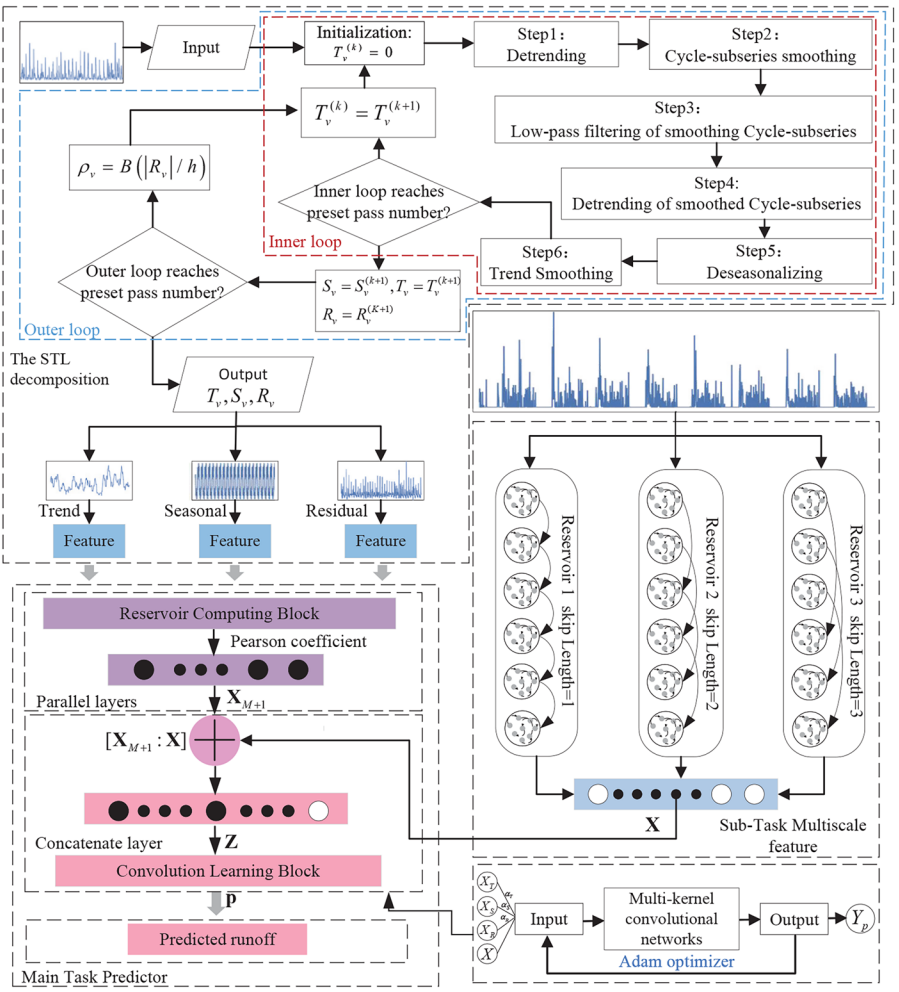


Fig. 1 Framework of Multi-SDIPC model

where B is the square weight function:

$$B(u)=\begin{cases}(1-u^2)^2, 0\le u<1\\ 0, u\ge 1\end{cases} \tag{3}$$

After that, the inner loop is further iterated using robustness weights ρ_v to reduce the effect of outliers on the regression. In this paper, the STL of the runoff time series is obtained:

$$Y=STL(Y)=T+S+R \tag{4}$$

where $T, S, R\in\mathbb{R}^D$ denotes trend, seasonal, and residual, respectively.

2.2 Reservoir Computing(RC)

The RC is an extension framework of the Recurrent Neural Network, which views chaotic time series as a superposition of a series of stochastic perturbations and deterministic trends, and its prediction accuracy of chaotic time series is improved by continuously updating and adjusting the parameters of the reservoir of random perturbations (Jaeger 2001). The RC consists of an input layer, a reservoir layer, and an output layer. The input weights and reservoir weights are randomly initialized and do not require training; and the reservoir weights are large sparse matrices with non-zero elements indicating the activated neurons in the reservoir, which have short-term memory. The output weights are then adjusted by solving a linear regression for training. The architecture of RC is expressed by the following equation:

$$\mathbf{x}(t) = f\left(\mathbf{W}^{\text{res}}\mathbf{x}(t-1) + \mathbf{W}^{in}\mathbf{u}(t)\right) \tag{5}$$

$$\mathbf{y}(t) = f^{\text{out}}\left(\mathbf{W}^{\text{out}}\mathbf{x}(t)\right) \tag{6}$$

where $\mathbf{u}(t) \in \mathbb{R}^{\text{D}}, \mathbf{x}(t) \in \mathbb{R}^{N}$ and $y(t) \in \mathbb{R}^{L}$ denote the input, reservoir, and output of the RC, respectively. $\mathbf{u}(t)$ denotes the input value at moment t of the time series, $\mathbf{W}^{in} \in \mathbb{R}^{N\times D}$ (N is the number of reservoir nodes, D is the input dimension) denotes the connection weight between the input layer and the reservoir, $\mathbf{x}(t)$ denotes the reservoir at moment t of the time series, $\mathbf{W}^{res} \in \mathbb{R}^{N\times N}$ denotes the connection weight of the reservoir itself, and $\mathbf{W}^{\text{out}} \in \mathbb{R}^{L\times N}$ (L denotes the output dimension) denotes the connection weight between the reservoir and the output layer. $f(\cdot)$ and $f^{out}(t)$ denote the activation functions on the reservoir and output layers.

3 Main Research

3.1 The Multi-SDIPC Framework for Runoff Forecasting

The Multi-SDIPC framework is shown in Fig. 1 and designed as follows:

1. The STL decomposition: Decomposition of chaotic runoff sequences using STL decomposition.
2. Parallel layer: A multi-input parallel reservoir layer is constructed to obtain the nonlinear dynamic features in each sub-series through high-dimensional mapping to obtain the correlation between multiple explanatory components and response variables.
3. Sub-task multiscale features: Constructing sub-task multi-timescale reservoirs to extract multi-timescale chaotic features for explanatory variables such as rainfall, surface water input, average daily temperature, and potential evapotranspiration to achieve more accurate predictions to realize useful knowledge transfer.
4. Concatenate layer: Constructing a concatenated convolutional layer to extract complex feature dependencies in high-dimensional concatenated explanatory variables using a multi-kernel convolutional neural network, and finally outputting runoff prediction results through a maximum pooling layer and a fully connected layer.

3.2 Sub-Task Multiscale Feature

There are multi-timescale structures in time series, Ma et al. (2021) designed an RC-based method to capture the multi-timescale structures in time data. By combining the use of multi-task learning ideas, this paper extracts implicit multi-timescale feature information from hydrological time series features to achieve information sharing between tasks and rational utilization, so that the overall prediction can achieve better results. In this paper, a reservoir is used to extract multi-timescale explanatory variables, as shown in the Sub-Task Multi-scale Feature in Fig. 1. A multi-timescale reservoir consisting of M independent reservoir connections is used, where each reservoir consists of N nodes, and each reservoir captures long-term dependent features by jumping connected. The dependency of different reservoirs is capable of extracting different timescale features.

Given a D-dimensional hydrological time series input:

$$\mathbf{U} = (u(1), \ldots, u(t), \ldots, u(T))^T \tag{7}$$

where $\mathbf{U} \in \mathbb{R}^{D\times T}$,$u(t) \in \mathbb{R}^D$, T denotes the length of the time series; then the state update equation for the i-th reservoir of the reservoir pool with d_i period at time t is expressed as:

$$x_i(t) = \begin{cases} f\left(\mathbf{W}_i^{res} x(0) + \mathbf{W}_i^{in} u(t)\right), if\ t - d_i \le 0 \\ f\left(\mathbf{W}_i^{res} x_i\left(t - d_i\right) + \mathbf{W}_i^{in} u(t)\right), if\ t - d_i > 0 \end{cases} \tag{8}$$

where $x(0)$ denotes the initial state, $\mathbf{W}_i^{in} \in \mathbb{R}^{N\times D}(i = 1, \ldots, M)$ denotes the input weight matrix corresponding to the reservoir with a period; d_i denotes the period in the reservoir, reservoirs with different periods can obtain time dependence at different time scales, the larger the period, the wider the range of time dependence; $\mathbf{W}_i^{\text{res}} \in \mathbb{R}^{N\times N}(i = 1, \ldots, M)$ denotes the matrix of state feedback weights corresponding to the reservoir with period d_i; $f(\cdot)$ denotes the activation function in the reservoir (usually $\tanh(\cdot)$). Collecting all the reservoir states of the i-th reservoir in time order constitutes the matrix $\mathbf{X}_i$:

$$\begin{aligned} \mathbf{X}_i &= \mathcal{F}(U) = \mathcal{F}\left((u(1), \ldots, u(t), \ldots, u(T))^T\right) \\ &= \left(x_i(1), \ldots, x_i(t), \ldots x_i(T)\right)^T \\ &= \begin{pmatrix} x_i^1(1) & \cdots & x_i^n(1) & \cdots & x_i^N(1) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_i^1(t) & \cdots & x_i^n(t) & \cdots & x_i^N(t) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_i^1(T) & \cdots & x_i^n(T) & \cdots & x_i^N(T) \end{pmatrix}_{T\times N} \end{aligned} \tag{9}$$

Then, all reservoir states are expressed as:

$$\mathbf{X} = \left\{\mathbf{X}_i\right\}_{i=1}^{M} \tag{10}$$

3.3 Main-Task Predictor

This section describes the main task of runoff prediction, which consists of a parallel input layer, a concatenate layer, and a high-dimensional convolutional learning block, the framework of which is shown in Fig. 2.

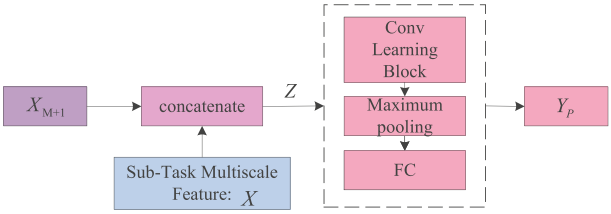


Fig. 2 Main-task predictor

3.3.1 Parallel Layer

The runoff time series Y is decomposed by STL to obtain three explanatory components: trend T, seasonal S, and residual R. To improve the accuracy of the prediction model, it is necessary to extract the chaotic characteristics of the corresponding explanatory components, so three parallel reservoir networks are established in the parallel input layer. The decomposed three explanatory components are simultaneously passed through the parallel reservoir to obtain the outputs $\mathbf{x}_T$, $\mathbf{x}_S$, $\mathbf{x}_R$. Since the influence of each explanatory component on the response variable is not identical, Pearson coefficients are introduced to give weights to each component so that the model learns more useful knowledge and finally obtains the weighted explanatory components.

The three explanatory components $\mathbf{x}_T$, $\mathbf{x}_S$ and $\mathbf{x}_R$ are respectively calculated with the runoff response variable Y to calculate Pearson coefficients and their absolute values are used as adjustment weights for the three components. The absolute value α_j is calculated as shown below:

$$\alpha_j = \left| \frac{\mathrm{Cov}\left(\mathbf{x}_j, Y\right)}{\sqrt{\mathrm{Var}\left(\mathbf{x}_j\right)} \cdot \sqrt{\mathrm{Var}(Y)}} \right| \quad j = T, S, R \tag{11}$$

where $\mathrm{Cov}\left(\mathbf{x}_j, Y\right)$ denotes the covariance operator between the different explanatory vectors and the response vector; $Var(\mathbf{x}_j)$ denotes the variance operator of the different explanatory vectors, and $Var(Y)$ denotes the variance operator of the response vector.

The final weighted explanatory component is obtained as follows:

$$\mathbf{X}_{M+1} = \alpha_T \cdot \mathbf{x}_T + \alpha_S \cdot \mathbf{x}_S + \alpha_R \cdot \mathbf{x}_R \tag{12}$$

3.3.2 Concatenate Layer

The previous section describes the extraction of chaotic multi-timescale explanatory variables $\mathbf{X}$ using a multi-timescale reservoir. For knowledge transfer, $\mathbf{X}$ is concatenated in the main task prediction. The chaotic multi-timescale explanatory variables $\mathbf{X}$ generated during the subtask and the weighted explanatory component $\mathbf{X}_{M+1}$ are concatenated to obtain the high-dimensional concatenated explanatory variables $\mathbf{Z}$, i.e.,$\mathbf{Z} = [\mathbf{X} : \mathbf{X}_{M+1}]$. The concatenated explanatory variables are specified in the following equation:

$$\begin{aligned} \mathbf{z}_i^{t:t+l-1} &= \mathbf{x}_i(t) \oplus \mathbf{x}_i(t+1) \oplus \cdots \oplus \mathbf{x}_i(t+l-1) \\ i &= 1, \ldots, M+1 \end{aligned} \tag{13}$$

$$\mathbf{Z} = \left\{ \mathbf{z}_i^{t:t+l-1} \right\}_{t=1}^{T-l+1} \tag{14}$$

where $\oplus$ denotes the linkage operator, i denotes the presence of i concatenated explanatory variables, $\mathbf{z}_i^{t:t+l-1} \in \mathbb{R}^{l \times N}$ denotes the sequence fragment from time t to time $t+l-1$

in the i-th concatenated explanatory variable, $\mathbf{Z}$ denotes the sequence of all concatenated explanatory components from time 1 to time $t+l-1$, and T is the length of the time series.

3.3.3 Conv Learning Block

To obtain multi-scale temporal dynamic dependencies in high-dimensional aggregated explanatory variables, a convolutional learning module with multiple convolutional kernels is constructed in this paper (Cui et al. 2016; Ma et al. 2021). The convolutional learning block consists of a convolutional layer, a maximum pooling layer, and a fully connected layer (as shown in Fig. 2). The convolutional layer uses multiple convolutional kernels of different sizes to extract deeper multiscale features for each concatenated interpretation component; the maximum pooling layer preserves the invariance of multiscale features and removes redundant information. Figure 3 shows the multi-kernel convolution operation.

After obtaining the concatenated explanatory variables $\mathbf{Z}$ in the concatenate layer, it enters the multi-kernel convolution module to obtain the dependencies among the explanatory variables. The output of the convolution layer with explanatory variables passing through convolution kernels of different sizes is:

$$c_i^{l,k}(t) = f\left(w_i^{l,k} \times \mathbf{z}_i^{t:t+1-1} + b_i^{l,k}\right) \tag{15}$$

where $w_i^{l,k} \in \mathbb{R}^{l\times N}(i = 1, \ldots, M+1, l = l_1, \ldots, l_J, k = 1, \ldots, K)$, J and K denote the scale type of the convolution kernel and the number of convolution kernels of the corresponding scale, respectively, $w_i^{l,k}$ denotes the use of the k-th convolution kernel of size l at the ith concatenated explanatory variable, $f(\cdot)$ and $b_i^{l,k}$ denote the activation function and the corresponding bias, $c_i^{l,k}(t) \in \mathbb{R}$ denotes the result obtained by convolving the i-th concatenated explanatory variable at time t using the convolution kernel $w_i^{l,k}$. Then, the result of the full convolution of the i-th concatenated explanatory variable is denoted as:

$$\mathbf{c}_i^{l,k} = \left(c_i^{l,k}(1), \ldots, c_i^{l,k}(t), \ldots, c_i^{l,k}(T-l+1)\right)^T \tag{16}$$

At the end of the convolution process, the corresponding result is obtained by entering the maximum pooling layer expressed as:

$$p_i^{l,k} = \arg\max_{t\in[1,T+l-1]} c_i^{l,k}(t) \tag{17}$$

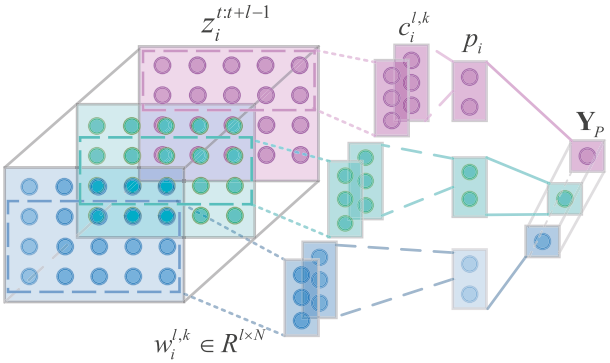


Fig. 3 Multi-kernel convolutional operations

$$\mathbf{p}_i = (p_i^{l_1,1}, \ldots, p_i^{l_1,K}, \ldots, p_i^{l_j,1}, \ldots, p_i^{l_i,K})^T \tag{18}$$

$$\mathbf{p} = \mathbf{p}_1 \oplus \cdots \oplus \mathbf{p}_i \oplus \cdots \oplus \mathbf{p}_M \oplus \mathbf{p}_{M+1} \tag{19}$$

After the maximum pooling phase, a fully connected layer is used to fuse the features on all time scales to obtain the final prediction results expressed as:

$$Y_p = f\left(\mathbf{W}^{fus}\mathbf{p} + \mathbf{b}^{fus}\right) \tag{20}$$

where $\mathbf{W}^{fus}$ and $\mathbf{b}^{fus}$ denote the weight matrix and bias corresponding to the fully connected layer, and $f(\cdot)$ denote the activation function.

3.4 Procedure of the Multi-SDIPC Model

In summary, the pseudo-code of the Multi-SDIPC algorithm proposed in this paper is as follows:

input : Chaotic runoff time series Y and hydrological time series such as rainfall, surface water input, potential evapotranspiration and average daily air temperature $\mathbf{U}$.

output: Runoff Y_p at the time to be predicted.

1 Time series decomposition. The STL of the chaotic runoff time series Y by Eq. (4) yields the three explanatory components of trend T, seasonal S, and residual R.

2 Parallel layers: Input of three explanatory components to a multi-input parallel reservoir, $\mathbf{x}_T$, $\mathbf{x}_S$, $\mathbf{x}_R$ obtained by Eq. (5).

3 Obtain the weight coefficients. Calculate the Pearson coefficients by Eq. (11) to obtain the weights α_T, α_S and α_R of the three explanatory components; finally, use Eq. (12) to obtain the concatenated weighted explanatory component $\mathbf{X}_{M+1}$.

4 Construct sub-task multi-timescale reservoirs. The explanatory variables such as precipitation, rainfall, potential evapotranspiration and mean daily air temperature are selected as inputs and denoted as $\mathbf{U}$; the reservoir state $x_i(t)$ of the i-th reservoir with d_i period at time t is obtained by Eq. (8), and the matrix $\mathbf{X}_i$ of all reservoir states of the ith reservoir is obtained by Eq. (9); The final multi-timescale explanatory variable $\mathbf{X} = \{\mathbf{X}_i\}_{i=1}^{M}$ is calculated by Eq. (10).

5 Concatenate layer. The high-dimensional concatenated explanatory variables $\mathbf{Z}$, i.e. $\mathbf{Z} = [\mathbf{X}_{M+1} : \mathbf{X}]$, are obtained by Eqs. (13) and (14).

6 Constructing the convolutional learning block. Firstly, the high-dimensional concatenated explanatory variables $\mathbf{Z}$ are obtained by convolutional operations of different sizes of convolutional kernels through Eq. (15) to obtain $c_i^{l,k}(t)$; secondly, the result $\mathbf{c}_i^{l,k}$ after convolution of the i-th concatenated explanatory variable is obtained through Eq. (16); $p_i^{l,k}$ is then obtained from the maximum pooling layer by Eq. (17), and all pooling results are concatenated and represented as $\mathbf{p}$ by Eqs. (18) and (19).

7 The final runoff prediction result Y_p is obtained by the fully connected layer of Eq. (20).

Algorithm 1 Multi-SDIPC

4 Case Study

4.1 CAMELS Dataset

This paper uses the CAMELS dataset (Addor et al. 2017), which provides hydrometeorological forcing data for 671 small and medium-sized basins in the contiguous United States, divided into 18 hydrologic units spanning the entire continental United States and covering a wide range of hydroclimatic conditions. The CAMELS dataset contains basin summary on a daily time scale from October 1, 1980, to December 31, 2014, meteorological forcing data, and observed runoff values. The daily meteorological data consists of Daymet data, Maurer data, and NLDAS data (Maurer et al. 2002; Cosgrove et al. 2003; Thornton et al. 2022). This dataset's diverse catchment characteristics make it ideal for large sample and hydrological studies. Daymet data enables researchers to model ecosystem processes in regions lacking meteorological data (Thornton et al. 2022). Daymet data has a high spatial resolution (1 km grid, while the other two datasets have a spatial resolution of 12 km grid), so the Daymet dataset was chosen for this study, in which the first 70% is the training set and the second 30% is the testing set.

This study selects four hydrologic, including 241 catchments. The specific reasons are as follows: first, these four hydrologic units, including the New England region in the northeast (hydrologic unit 01), the Arkansas-White-Red region in the center (hydrologic unit 11), the South Atlantic-Gulf region (hydrologic unit 03), and the Pacific Northwest region (hydrologic unit 17), span a wide range of locations across the United States, with relatively broad geographic distribution, widely varying watershed attributes, and different topographic and soil conditions (Yin et al. 2022). Secondly, to conserve available resources, these 241 catchment areas allow for better control of computational costs than all 671 catchment areas.

4.2 Methodology Assessment Metrics

To evaluate the predictive runoff model performance, this paper uses three evaluation metrics Nash-Sutcliffe efficiency factor (NSE), mean absolute error (MAE), percentage bias (PBIAS), top 20 % peak flow bias (FHV) to measure the performance effectiveness of different models, which are described as follows:

$$NSE = 1 - \frac{\sum_{i}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i}^{n} \left(y_i - \bar{y}\right)^2} \tag{21}$$

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

$$PBIAS = \frac{\sum_{i=1}^{n} \hat{y}_i - \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i} \times 100\% \tag{23}$$

$$FHV = \frac{\sum_{h=1}^{H} \left(QS_h - QO_h\right)}{\sum_{h=1}^{H} QO_h} \times 100\% \tag{24}$$

where y_i denotes the ith runoff observation, $\hat{y}$ denotes the average of all runoff observations, $\hat{y}_i$ denotes the ith runoff prediction, n denotes the total number of samples in the runoff series, QO_h denotes runoff observation, QS_h denotes runoff prediction, h = 1,2,.... H is the index of the flow with exceedance probability below 0.02.

4.3 Data Preprocessing

Before training the model, the input data were first standardized, and then the explanatory variables with correlation coefficients greater than 0.1 with the response variables were selected as inputs to the model, and their correlation coefficients heatmaps are shown in Fig. 4.

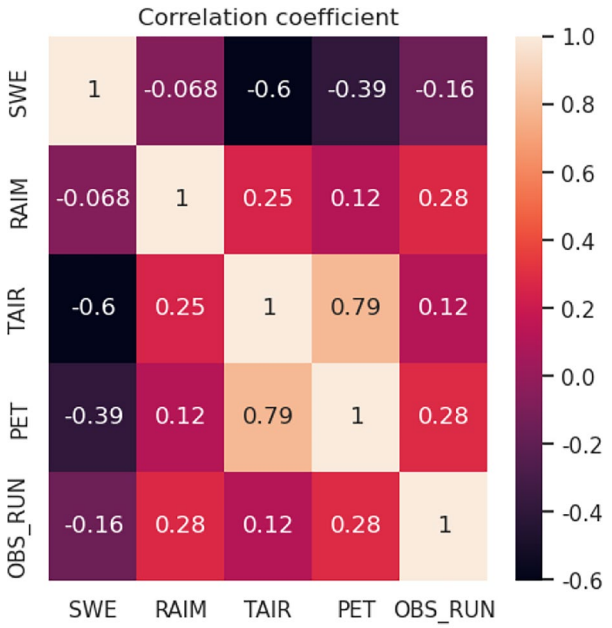
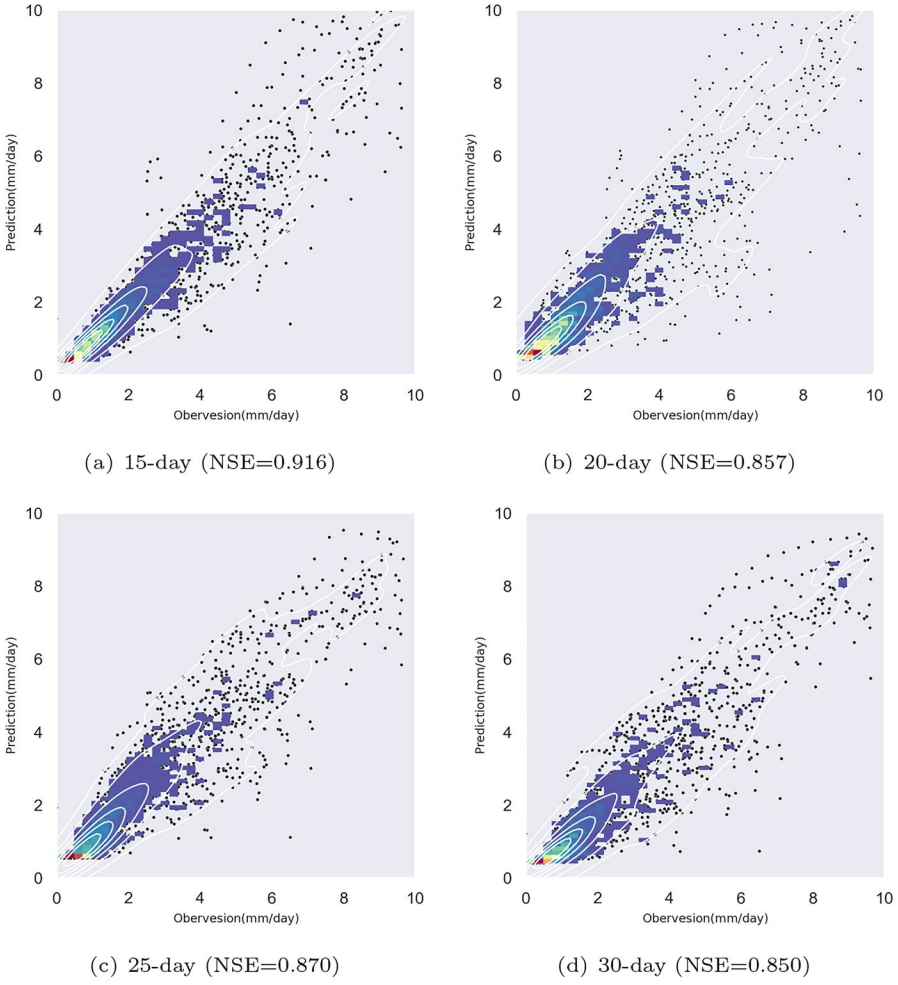


Fig. 4 Heatmap of relevant variables

4.4 Input Time Step Selection

The choice of input time step is crucial for the predictive performance of a model. In this paper, we design experiments on the prediction accuracy corresponding to the Multi-SDIPC model at different time steps (15-day, 20-day, 25-day, 30-day). From Fig. 5 comparing the scatter density plots of different time steps, it can be seen that: the density distribution of the time step of 15-day is smoother and more uniform, the data distribution is more centralized, and the prediction effect is the best, so this paper chooses 15-day as the input time step.



(a) 15-day (NSE=0.916)

(b) 20-day (NSE=0.857)

(c) 25-day (NSE=0.870)

(d) 30-day (NSE=0.850)

Fig. 5 Predictive performance for different time-steps

4.5 Experimental Results and Discussion

This paper focuses on short-term runoff prediction by selecting 15 past runoff sequence lengths as inputs to predict runoff values at the next moment. To evaluate the prediction performance of the Multi-SDIPC model proposed in this paper, eight other models were selected for comparison: LSTM (Tian et al. 2023), CLSTM(consists of a convolutional neural network and LSTM), ESN, ConvMESN (Change the ConvMESN classification model to the regression model for comparison) (Ma et al. 2021), TLSTM (Kratzert et al. 2018), LSTM-STS (Xiang et al. 2020), LSTM-SS (Yin et al. 2022), VMD-GRU-SFS (Xu et al. 2021), Multi-SDPR, and Multi-SDIP. To better analyze the experimental results of the models, we first list four RQs to guide the experimental process as follows:

RQ1: Does the proposed Multi-SDIPC model obtain the best performance among all the compared methods?
RQ2: Does the hybrid model have better prediction performance than the single model?
RQ3: Does the use of multi-task learning improve the model prediction performance?
RQ4: Does the DIP framework improve the error accumulation problem?

To verify the prediction performance of the model in different hydrological regions, the Multi-SDIPC model was experimented with other baseline models on 241 catchments, as shown in Table 1. The bolded numbers in the table represent the optimal results, and the rest of the tables are the same. Figure 6 represents the NSE, MAE, PBIAS, and FHV metrics of the proposed model in this paper compared to other baseline models, it can be seen that the model outperforms in all metrics comparisons, which indicates that the model has excellent predictive ability and good overall generalization ability in the face of large regional differences and different hydrological conditions, and thus can answer RQ1. The models are divided into single and hybrid models according to whether they include time series decomposition methods, and the first seven models (LSTM, CLSTM, ESN, ConvMESN, TLSTM, LSTM-STS, LSTM-SS) in Table 1 are regarded as single models, and the last four models (VMD-GRU-SFS, Multi-SDPR, Multi-SDIP, Multi-SDIPC) are regarded as hybrid models. From the experimental results, it can be reflected that the single models of LSTM, ESN, ConvMESN, CLSTM, and TLSTM showed poor generalization (NSE <0.7) when predicting the 241 catchments, the LSTM-SS significantly improves the single model modeling accuracy due to the model's ability to reduce prediction noise and errors, similar to physical models. The four hybrid models, the VMD-GRU-SFS, Multi-SDPR, Multi-SDIP, and Multi-SDIPC, have better overall prediction performance than a single model. The decomposition-based prediction model is better than a single model in extracting useful information and identifying trends accurately from complex hydrological series, resulting in higher prediction accuracy, robustness, and generalization. It can effectively answer RQ2 to prove that the hybrid model has a greater improvement in the model prediction performance. To answer RQ4, this paper compares the DPR and DIP frameworks for runoff prediction. Results show that DPR has a large deviation due to error accumulation, while DIP effectively reduces the cumulative impact of errors. This is consistent with other studies (Xu et al. 2021); The model extracts dynamic dependencies of high-dimensional variables using multiple convolutional kernels, which is supported by other studies (Ma et al. 2021).

The models are divided into regional models and individual models according to the size of the prediction data set, corresponding to Tables 1 and 2 respectively. As the prediction

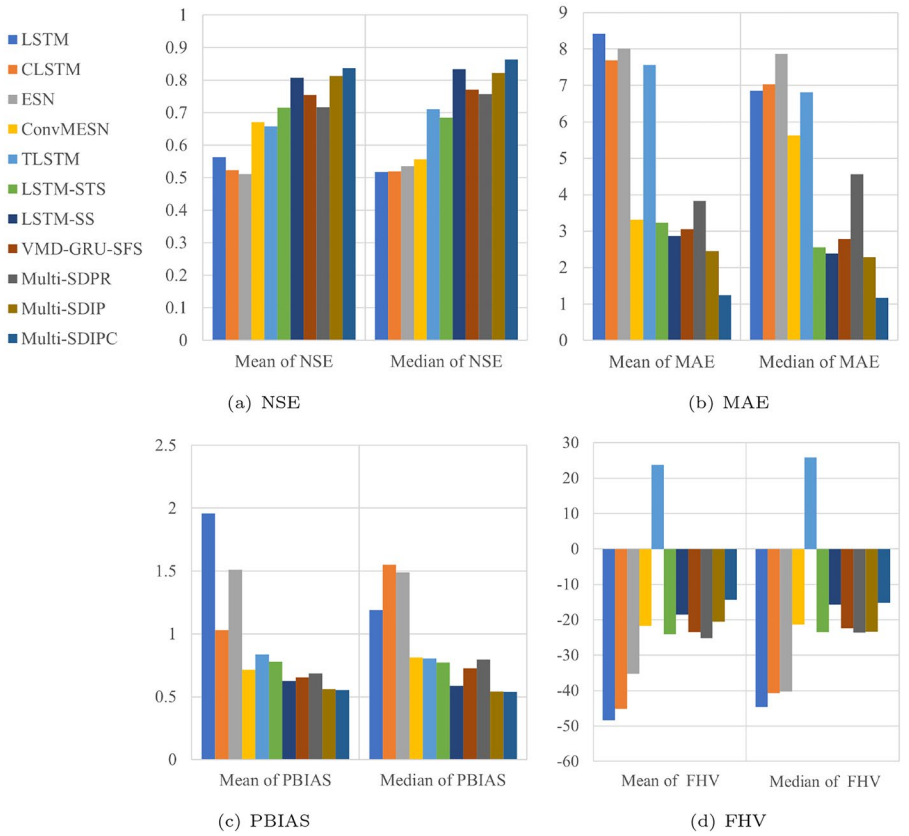


Fig. 6 Comparison of NSE, MAE, PBIAS and FHV metrics for regional models

range expands, the prediction accuracy of regional models is significantly lower than that of individual models. Moreover, the prediction accuracy of a single model for multiple stations is lower than that of a single station, which indicates that the single model may lack the generalization ability of the model due to ignoring the chaotic characteristics of the hydrological time series itself and the temporal and spatial dependence of the characteristics; Hybrid models have better accuracy in predicting a single site or multiple regions. This suggests that the hybrid model with better prediction performance and some generalization performance is more suitable for the prediction of runoff in a wide range of areas, which strengthens the answer to RQ2, this is consistent with other studies (He et al. 2020).

The experimental results for the single-individual model and the hybrid-individual model at site 13011500 are plotted and displayed in Figs. 7 and 8. Figure 7 shows the baseline model, which is unable to accurately predict future runoff trends. The scatterplot of observed versus predicted runoff is more dispersed than that of the Multi-SDIPC model. On the other hand, the fit chart of the Multi-SDIPC model in Fig. 8 better captures the future trend direction of runoff sequences. Its scatter clustering map is closer to the ideal line than the other two models. The Multi-SDIPC model outperforms both the Multi-SDPR and the Multi-SDIP models, which may not effectively capture high

Table 1 Comparison of predicted outcome metrics of Multi-SDIPC model and 10 baseline models over 241 catchments

	Model	Mean of NSE	Median of NSE	Mean of MAE	Median of MAE	Mean of PBIAS	Median of PBIAS	Mean of FHV	Median of FHV
Single-region-model	LSTM	0.563	0.477	8.42%	6.85%	1.96	1.19	-46.4%	-44.6%
	CLSTM	0.523	0.519	7.69%	7.03%	1.03	1.55	-45.1%	-40.7%
	ESN	0.511	0.536	8.01%	7.86%	1.51	1.49	-35.3%	-40.3%
	ConvMESN	0.671	0.556	3.31%	5.63%	0.715	0.814	-21.7%	-21.3%
	TLSTM	0.658	0.711	7.56%	6.81%	0.835	0.804	23.8%	25.9%
	LSTM-STS	0.715	0.684	3.23%	2.55%	0.779	0.773	-24.1%	-23.5%
	LSTM-SS	0.807	0.834	2.87%	2.39%	0.624	0.586	-18.6%	-15.7%
Hybrid-region model	VMD-GRU-SFS	0.754	0.770	3.05%	2.78%	0.655	0.726	-23.5%	-22.4%
	Multi-SDPR	0.717	0.757	3.82%	4.56%	0.686	0.795	-25.2%	-23.6%
	Multi-SDIP	0.813	0.822	2.45%	2.28%	0.562	0.541	-20.5%	-23.4%
	Multi-SDIPC	**0.837**	**0.863**	**1.24%**	**1.17%**	**0.553**	**0.538**	**-14.3%**	**-14.2%**

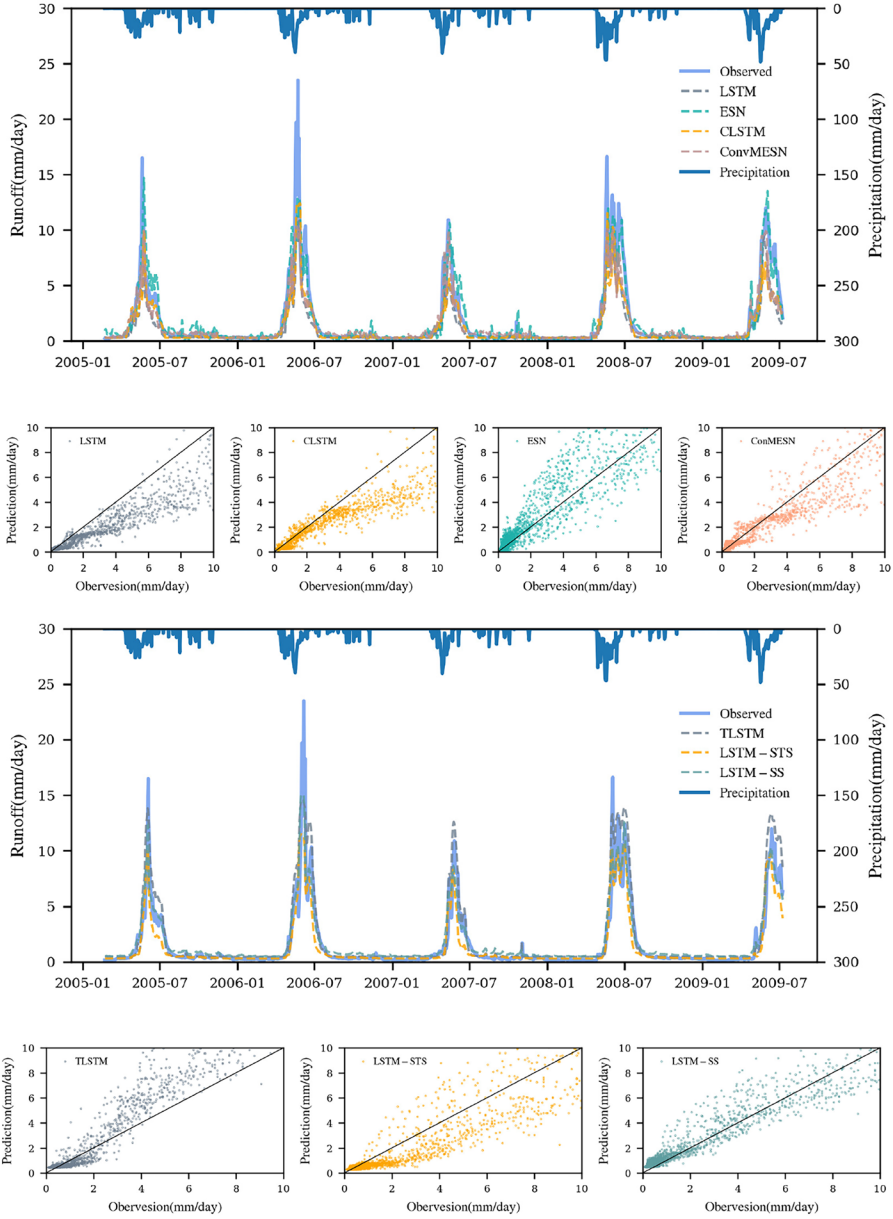


Fig. 7 Fitted plots and scatter plots of experimental results of single-individual model

flow in the runoff process. The Multi-SDPR model accumulates prediction errors over time, resulting in scattered cluster plots dispersed around the ideal fit line. On the other hand, the Multi-SDIPC model has a more stable overall prediction and a certain tracking ability for future trends of runoff, but it has a certain prediction error for extreme

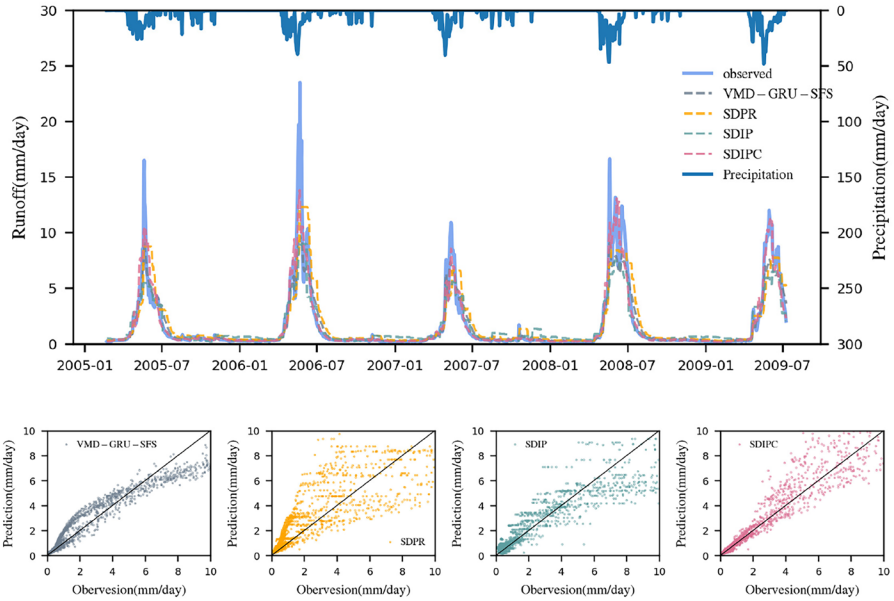


Fig. 8 Fitted plots and scatter plots of experimental results of hybrid-individual model

events. Figure 9 shows that the Multi-SDIPC model performs better and exhibits more stable performance when predicting runoff for a single catchment.

To investigate the effect of multi-task learning on the performance of the prediction model, the following ablation experiments are designed for the Multi-SDIPC model to answer RQ3. To make the experimental conditions include the hydrological properties of different hydrological units and thus reflect the generalization characteristics of the model, one catchment area was selected from each of the four hydrological units of the data set: site 01013500, site 02312200, site 07071500 and site 13083000 for the

Table 2 Comparison of the Multi-SDIPC model with 10 baseline models for predicted outcome metrics at site 13011500

	Model	NSE	MAE	PBIAS	FHV
Single-individual model	LSTM	0.673	0.617	5.14%	-23.3%
	CLSTM	0.743	0.873	3.55%	-22.6%
	ESN	0.686	0.876	4.93%	-16.5%
	ConvMESN	0.766	0.714	5.50%	-18.5%
	TLSTM	0.709	0.537	4.66%	15.8%
	LSTM-STS	0.751	0.684	3.67%	-20.6%
	LSTM-SS	0.802	0.499	2.50%	-16.3%
Hybrid-individual model	VMD-GRU-SFS	0.789	0.555	3.17%	-12.4%
	Multi-SDPR	0.737	0.681	3.02%	-25.1%
	Multi-SDIP	0.806	0.490	2.99%	-18.2%
	Multi-SDIPC	**0.882**	**0.358**	**0.15%**	**-11.3%**

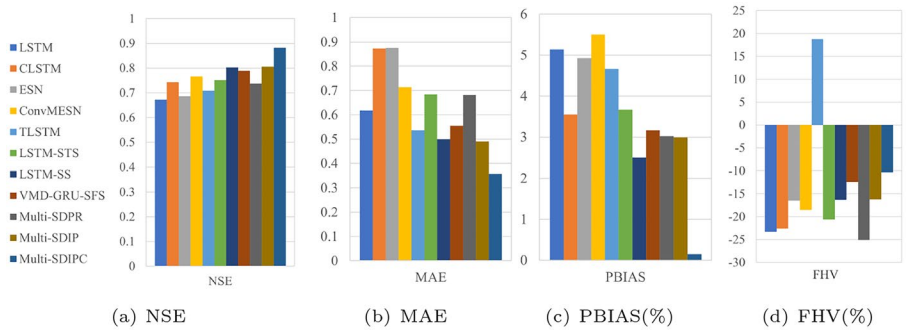


Fig. 9 Comparison of NSE, MAE, PBIAS and FHV metrics for individual models

ablation experiments, the Fig. 10 represents the STL decomposition of the runoff data for each of the four sites.

- w/o Main: The Multi-SDIPC model without STL runoff series decomposition, using multi-timescale auxiliary explanatory variables for prediction.
- w/o Sub: The Multi-SDIPC model has no sub-task multi-timescale extractor and uses the explanatory variable components of the STL runoff series decomposition for prediction.

The results of the ablation experiments presented in Table 3 highlight the superior performance of the Multi-SDIPC model in all metrics. This suggests that multi-task learning can effectively transfer knowledge and significantly improve the model's prediction performance. On the other hand, the w/o Main model displayed poor performance across all metrics, indicating the importance of time series decomposition in extracting complex information from runoff series. The w/o Sub model showed moderate performance across all indicators, but the metrics were lower than those of the Multi-SDIPC model. This suggests that sub-task learning is useful in acquiring chaotic information from complex runoff sequences, which can improve the model's predictive ability.

As can be seen from Table 1, the Multi-SDIPC model proposed in this paper exhibits better performance than other baseline models under the 1-day ahead runoff prediction condition. It is advisable to conduct the runoff prediction experiment 7 days in advance as an extension experiment to test the performance of the model for long-term future runoff prediction.

The results of the model for predicting runoff 7 days in advance are shown in Table 4. As time and prediction steps increase, indicators fluctuate less and the Multi SDIPC model remains stable with good generalization ability.

Table 3 Multi-task ablation experiments

Model	Mean of NSE	Mean of MAE	Mean of PBIAS	Mean of FHV
w/o Main	0.816	0.314	5.69%	-13.1%
w/o Sub	0.819	0.256	3.34%	-16.3%
Multi-SDIPC	**0.910**	**0.314**	**1.28%**	**-10.6%**

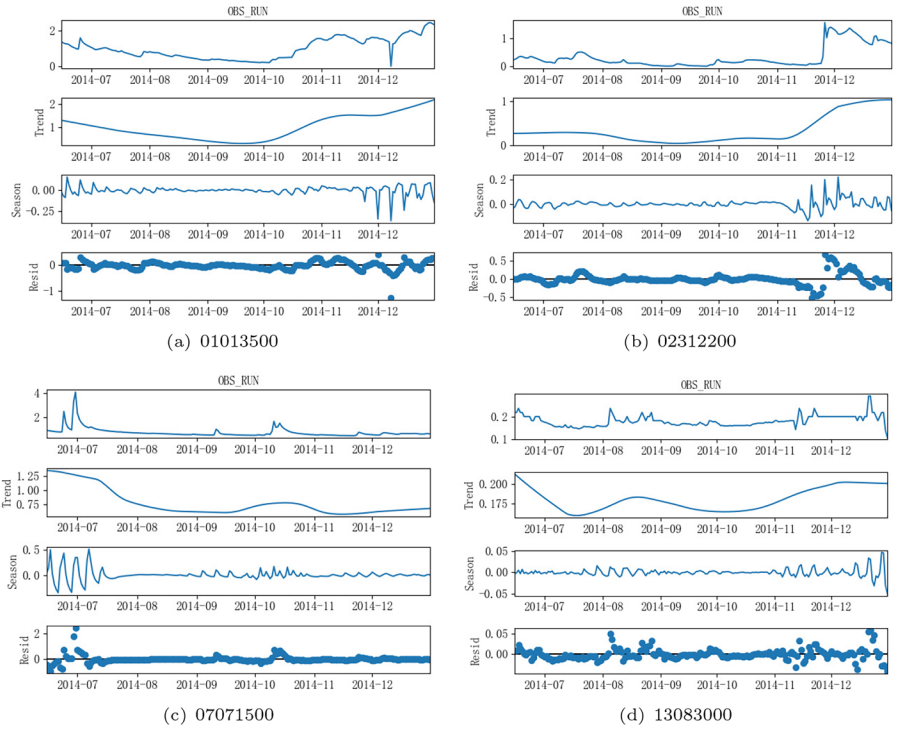


Fig. 10 The STL decomposes runoff time series

Table 4 Multi-SDIPC model predicts results 7 days in advance at site 13313000

Length / Metrics	1-day-ahead	2-day-ahead	3-day-ahead	4-day-ahead	5-day-ahead	6-day-ahead	7-day-ahead
NSE	0.923	0.901	0.887	0.876	0.853	0.833	0.804
MAE	0.303	0.346	0.348	0.452	0.431	0.451	0.492
PBIAS	0.53%	0.92%	1.68%	1.10%	1.51%	2.38%	2.87%
FHV	-9.03%	-12.8%	-14.8%	-17.6%	-23.4%	-21.1%	-18.5%

5 Conclusion

In this paper, the Multi-SDIPC model is proposed to address the chaotic and multi-scale characteristics of hydrological dynamic composite systems. The RC is utilized to simulate the potential chaotic patterns in the hydrological system. Additionally, a sub-task multi-timescale reservoir is constructed to extract the multi-scale chaotic features of the auxiliary explanatory variables. The study's key findings are as follows:

1. Compared to other models, the Multi-SDIPC model achieves better accuracy and generalization by effectively capturing both chaotic and multi-scale features of hydrological time series.

2. Compared to the single model, the hybrid model has better predictive performance and stability, demonstrating that the hybrid model can improve runoff predictions.
3. Through a series of multi-task ablation experiments, it has been demonstrated that multi-task learning contributes to improving the accuracy of model predictions. Additionally, experiments performed on 7-day runoff prediction have revealed that the Multi-SDIPC model exhibits high overall prediction accuracy with minimal fluctuations. This makes it a suitable choice for medium to long-term predictions.

As with many studies, this paper has its limitations. The Multi-SDIPC model proposed in this study did not provide a significant improvement in predicting extreme runoff events. Therefore, further research is needed to explore ways to enhance the accuracy of predicting extreme runoff events.

Author Contributions Hui Zuo: conceptualization, methodology, software, visualization, formal analysis, writing-original draft preparation. Gaowei Yan: conceptualization, methodology, writing-review and editing, supervision, funding acquisition. Ruochen Lu: supervision, software, writing-review & editing. Rong Li: conceptualization, methodology, writing-review & editing. Shuyi Xiao: conceptualization, methodology writing-review & editing. Yusong Pang: conceptualization, supervision, Writing-review & editing.

Funding This work was supported by the National Natural Science Foundation of China (61973226, 62003233), the Shanxi Provincial Department of Water Resources 2023 Water Technology Research and Promotion Subsidy Projects(2023GM17), the Shanxi Province Major Special Program of Science and Technology "Unveiling and Commanding" Project(202201090301013), the Natural Science Foundation of Shanxi Province(202203021222101).

Data Availability The data that support the findings of this study are openly available in Hydrology and Earth System Sciences at https://doi.org/10.5194/hess-21-5293-2017, 2017.

Declarations

Ethics Approval This paper does not contain any studies with human participants or animals performed by any of the authors.

Consent to Participate Consent.

Consent for Publication Consent.

Conflict of Interest The authors declare no competing interests.

References

Addor N, Newman AJ, Mizukami N, Clark MP (2017) The camels data set: catchment attributes and meteorology for large-sample studies. Hydrol Earth Syst Sci 21(10):5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Adnan RM, Mostafa RR, Kisi O, Yaseen ZM, Shahid S, Zounemat-Kermani M (2021) Improving streamflow prediction using a new hybrid elm model combined with hybrid particle swarm optimization and grey wolf optimization. Knowl-Based Syst 230:107379. https://doi.org/10.1016/j.knosys.2021.107379

Arcomano T, Szunyogh I, Pathak J, Wikner A, Hunt BR, Ott E (2020) A machine learning-based global atmospheric forecast model. Geophys Res Lett 47(9):e2020GL087776. https://doi.org/10.1029/2020GL087776

Cosgrove BA, Lohmann D, Mitchell KE, Houser PR, Wood EF, Schaake JC, Robock A, Marshall C, Sheffield J, Duan Q, Luo L, Higgins RW, Pinker RT, Tarpley JD, Meng J (2003) Real-time and retrospective forcing in the north american land data assimilation system (nldas) project. J Geophys Res Atmos 108(D22). https://doi.org/10.1029/2002JD003118

Cui Z, Chen W, Chen Y (2016) Multi-scale convolutional neural networks for time series classification. CoRR abs/1603.06995

Gao S, Huang Y, Zhang S, Han J, Wang G, Zhang M, Lin Q (2020) Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. J Hydrol 589:125188. https://doi.org/10.1016/j.jhydrol.2020.125188

Giri F, Devercelli M (2023) Chaos arising from the hydrological behaviour of a floodplain river during the last century. River Res Appl 39(2):241–254. https://doi.org/10.1002/rra.4080

Griffith A, Pomerance A, Gauthier DJ (2019) Forecasting chaotic systems with very low connectivity reservoir computers. Chaos: An Interdisciplinary Journal of Nonlinear Science 29(12). https://doi.org/10.1063/1.5120710

Gupta V, Li LK, Chen S, Wan M (2023) Model-free forecasting of partially observable spatiotemporally chaotic systems. Neural Netw 160:297–305. https://doi.org/10.1016/j.neunet.2023.01.013

Han M, Zhang S, Xu M, Qiu T, Wang N (2019) Multivariate chaotic time series online prediction based on improved kernel recursive least squares algorithm. IEEE Trans Cybern 49(4):1160–1172. https://doi.org/10.1109/TCYB.2018.2789686

He R, Zhang L, Chew AWZ (2022) Modeling and predicting rainfall time series using seasonal-trend decomposition and machine learning. Knowl-Based Syst 251:109125. https://doi.org/10.1016/j.knosys.2022.109125

He X, Luo J, Li P, Zuo G, Xie J (2020) A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting. Water Resources Management: An International Journal, Published for the European Water Resources Association (EWRA) 34(2):865–884. https://doi.org/10.1007/s11269-020-02483-x

Jaeger H (2001) The "echo state'' approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German National Research Center for Information Technology GMD Technical Report 148(34):13

Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M (2018) Rainfall-runoff modelling using long short-term memory (lstm) networks. Hydrology and Earth System Scviences 22(11):6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Li BJ, Sun GL, Liu Y, Wang WC, Huang XD (2022) Monthly runoff forecasting using variational mode decomposition coupled with gray wolf optimizer-based long short-term memory neural networks. Water Resour Manage 36(6):2095–2115. https://doi.org/10.1007/s11269-022-03133-0

Ma Q, Chen E, Lin Z, Yan J, Yu Z, Ng WWY (2021) Convolutional multitimescale echo state network. IEEE Trans Cybern 51(3):1613–1625. https://doi.org/10.1109/TCYB.2019.2919648

Malakoutian MMA, Samaei SY, Khaksar M, Malakoutian Y (2022) A prediction of future flows of ephemeral rivers by using stochastic modeling (ar autoregressive modeling). Sustainable Operations and Computers 3:330–335. https://doi.org/10.1016/j.susoc.2022.05.003

Maurer EP, Wood AW, Adam JC, Lettenmaier DP, Nijssen B (2002) A long-term hydrologically based dataset of land surface fluxes and states for the conterminous united states. J Clim 15(22):3237–3251. https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2

McKitrick R, Christy J (2019) Assessing changes in us regional precipitation on multiple time scales. J Hydrol 578:124074. https://doi.org/10.1016/j.jhydrol.2019.124074

Mohsin AKM, Hongzhen L, Iqbal MM, Salim ZR, Hossain A, Kafy AA (2021) Forecasting e-waste recovery scale driven by seasonal data characteristics: A decomposition-ensemble approach. Waste Manag Res 40:870–881. https://doi.org/10.1177/0734242X211061443

Pérez-Alarcón A, Garcia-Cortes D, Fernández-Alvarez JC, Martínez-González Y (2022) Improving monthly rainfall forecast in a watershed by combining neural networks and autoregressive models. Environ Process 9(3):53. https://doi.org/10.1007/s40710-022-00602-x

Sivakumar B (2000) Chaos theory in hydrology: important issues and interpretations. J Hydrol 227(1):1–20. https://doi.org/10.1016/S0022-1694(99)00186-9

Theodosiou M (2011) Forecasting monthly and quarterly time series using stl decomposition. Int J Forecast 27(4):1178–1195. https://doi.org/10.1016/j.ijforecast.2010.11.002

Thornton M, Shrestha R, Wei Y, Thornton P, Kao SC, Wilson B (2022) Daymet: Daily surface weather data on a 1-km grid for North America, version 4 r1

Tian Q, Gao H, Tian Y, Jiang Y, Li Z, Guo L (2023) Runoff prediction in the xijiang river basin based on long short-term memory with variant models and its interpretable analysis. Water 15(18). https://doi.org/10.3390/w15183184

Vlachas P, Pathak J, Hunt B, Sapsis T, Girvan M, Ott E, Koumoutsakos P (2020) Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. Neural Netw 126:191–217. https://doi.org/10.1016/j.neunet.2020.02.016

Xiang Z, Yan J, Demir I (2020) A rainfall-runoff model with lstm-based sequence-to-sequence learning. Water Resour Res 56(1): e2019WR025326. https://doi.org/10.1029/2019WR025326

Xu Z, Zhou J, Mo L, Jia B, Yang Y, Fang W, Qin Z (2021) A novel runoff forecasting model based on the decomposition-integration-prediction framework. Water 13(23). https://doi.org/10.3390/w13233390

Yaseen ZM, Sulaiman SO, Deo RC, Chau KW (2019) An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. J Hydrol 569:387–408. https://doi.org/10.1016/j.jhydrol.2018.11.069

Yin H, Wang F, Zhang X, Zhang Y, Chen J, Xia R, Jin J (2022) Rainfall-runoff modeling using long short-term memory based step-sequence framework. J Hydrol 610:127901. https://doi.org/10.1016/j.jhydrol.2022.127901

Yu D, Guobin Xu YW, Yang D (2020) The middle huaihe river stability analysis and optimization of hydrological chaos forecasting model. Geomat Nat Haz Risk 11(1):1805–1826. https://doi.org/10.1080/19475705.2020.1815870

Yu-tong Z, Xiao-min W, Ting L (2019) Characteristic analysis and prediction of runoff based on chaotic wavelet neural network. In 2019 Chinese Control And Decision Conference (CCDC), pp. 1765–1769

Zhang Y, Yang Q (2017) 09. An overview of multi-task learning. Natl Sci Rev 5(1):30–43. https://doi.org/10.1093/nsr/nwx105

Zhou T, Wen X, Feng Q, Yin Z, Yang L (2022) Research on runoff prediction of shule river based on bma multiple model combination. J Glaciol Geocryol 44(5):1606. https://doi.org/10.7522/j.issn.1000-0240.2022.0141

Zhou Z, Dong X (2012) Analysis about the seasonality of china's crude oil import based on x–12-arima. Energy 42(1):281–288. https://doi.org/10.1016/j.energy.2012.03.058

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.