DAR-type model based on "long memory-threshold" structure: a competitor for daily streamflow prediction under changing environment

Huimin Wang[1], Songbai Song[2,3], Gengxi Zhang[1*], Thian Yew Gan[4], Zhuoyue Peng[1]

[1]College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou 225009, China
[2]College of Water Resources and Architectural Engineering, Northwest A & F University, Yangling 712100, China
[3]Key Laboratory for Agricultural Soil and Water Engineering in Arid Area of Ministry of Education, Northwest A & F University, Yangling 712100, China
[4]Department of Civil and Environmental Engineering, University of Alberta, Edmonton T6G2R3, Canada

Correspondence to: Gengxi Zhang (gengxizhang@yzu.edu.cn)

Abstract. The non-stationarity and non-linearity of streamflow have increased with changes of the environment, challenging accurate streamflow prediction. Furthermore, the overlook of long-term memory features could lead to biases in model parameter estimation and testing of time series properties. The classical linear Autoregressive-Generalized Autoregressive Conditional Heteroskedasticity (AR-GARCH) model has a narrow parameter range, and the moment conditional requirements for parameter estimation are relatively strict, limiting its applicability and prediction accuracy in modeling and predicting daily streamflow. Under the premise of long-term memory, a dual-threshold double autoregressive (DTDAR) model is proposed to capture the non-linear patterns in streamflow series. Using 15 hydrological stations in the Yellow River basin in China as an example, DTDAR models are compared with AR-GARCH and long short-term memory (LSTM) models to assess their applicability and predictive ability. The results indicate that the DTDAR models have a stronger predictive ability for daily streamflow than the AR-GARCH-type and LSTM models. The nonlinear changes of the daily streamflow time series are reflected in multiple linear structures by adding the threshold, improving the accuracy of the single linear structure method. The NSE values of the FDTDAR and TAR-GARCH models are higher than those of the DAR and AR-GARCH models by 0.013-0.556 and 0.031-0.582, respectively. Compared to the normal distribution, the Student's t distribution for residuals is a better choice for predicting daily streamflow time series in the study area. This study enriches the stochastic hydrological models and improves the accuracy of streamflow prediction.

1 Introduction

The behavioural characteristics of hydrological systems have become increasingly complex in the face of climate change and human activities (Lyu et al., 2023; Ma et al., 2024; Matic et al., 2022; Sivakumar, 2009). Time series analysis method has gained wide attention in hydrological simulation and prediction due to its outstanding ability to describe the structure, function, and development trend of the system, and has become one of the hot issues in stochastic hydrology research (Can et al., 2012;

Chen et al., 2021; Yang et al., 2022). As such, several stochastic models have been proposed, among which the regression model has a simple form and is easy to implement.

Traditional regression models, including the autoregressive (AR) model, the moving average (MA) model, and their combined form (ARMA model), are well-suited for stationary time series but have significant limitations. For non-stationary series, the differencing ARMA (ARIMA) model is an effective method, and its key difference from the ARMA model is the addition of a difference step to smooth the sequence (Wang et al., 2019). However, both the ARIMA and ARMA models struggle to handle seasonality in time series. The seasonal ARIMA (SARIMA) model is the preferred option for monthly streamflow time series with annual cycles resulting from seasonal variations (Adnan et al., 2019; Modarres, 2007). However, the illegible seasonal features in the daily streamflow series bring a challenge to the applicability of the SARIMA model. To overcome this issue, seasonal standardization is necessary to preprocess daily streamflow time series, removing the influence of seasonality for subsequent research (Guo et al., 2021a).

In the last decades, the conditional heteroscedastic behaviour (ARCH effect) of the mean model residual series obtained increasing attention due to changing environment (Guo et al., 2021b; Nazeri-Tahroudi et al., 2022; Wang et al., 2023a). The generalized Autoregressive Conditional Heteroscedasticity (GARCH) model has been often employed to improve the modelling accuracy of the AR-type model by eliminating the ARCH effect in the residuals (Fathian et al., 2019; Fathian and Vaheddoost, 2021; Pandey et al., 2019; Zha et al., 2020). However, the AR-GARCH model, a typical regression model for streamflow forecasting, has a more cumbersome combined form than a single model, hindering its application. Compared with the AR-GARCH, the double autoregressive (DAR) model proposed by Ling (2007) is more concise in form, can describe the first- and second-order moments behaviour of time series simultaneously, and has been applied maturely in the economic field (Hansen, 2021; Jiang et al., 2020; Li et al., 2019; Liu et al., 2018). In addition, the AR model requires the autoregressive and moving average parameters to fall within the range of -1~1, and the sum is lower than 1, while the GARCH model restricts non-negative values and the sum (except for the constant) is below 1. In contrast, the definition range of the DAR model parameter is broader, with first-order moment selection throughout the real number field and non-negative second-order moment constraints. However, the DAR model has not yet been applied in the field of hydrology, and its prediction performance has not been verified.

Dimitriadis and Koutsoyiannis (2015) demonstrated that long-term persistence—characterized by the temporal dependence between historical and contemporary streamflow—is fundamental to daily streamflow prediction. And it has been reported that time-varying volatility detected in daily streamflow series could be attributed to long-term memory (Dimitriadis et al., 2021; Graves et al., 2017; Grimaldi, 2004), which may lead to a spurious ARCH effect in residuals of the pure AR model, rendering the constructed GARCH model insufficient and affecting prediction performance. Both the combined AR-GARCH model and the novel DAR model are powerless to reproduce the long-term persistence of the daily streamflow time series. Long-term memory was always neglected in the vast majority of existing streamflow simulation and prediction studies. The research on long memory began with Hurst (1951) and Mandelbrot and Wallis (1969), and Hurst proposed the Hurst index (H, $H>0.5$ indicates that the time series has long memory) to identify it. Although this characteristic was initially identified within

the field of hydrology, at that time, there lacked a research methodology widely recognized by most hydrologists that could effectively reproduce the long-term memory of daily streamflow. Granger (1978) proposed the concept of "fractionally differencing", which laid the foundation for the development of later long-memory models (such as the ARFIMA model). Hosking (1981) followed up in the field of hydrology. Montanari et al (1996; 1997) still believed that the fractional difference model is the only method to describe the daily streamflow time series with long-term persistence.

Hydrological time series are highly nonlinear, as proven in numerous studies (Delforge et al., 2022; Feng et al., 2022; Guo et al., 2021b; Miller, 2022; Yuan et al., 2021), which poses significant challenges to traditional mean models (e.g., AR, ARMA). The threshold autoregressive model based on the AR model can capture and predict the nonlinearity in the mean behaviour of hydrological time series (Tong, 1983). Several recent studies (Fathian et al., 2019; Gharehbaghi et al., 2022; Huang et al., 2021; Kolte et al., 2023) have introduced a new combination model coupled with the nonlinear mean model, namely artificial intelligence-GARCH, for prediction work. Guo et al. (2021b) combined the Self-Exciting TAR model with the GARCH model to predict groundwater depth and achieved satisfactory prediction accuracy. However, the applicability of the TAR-GARCH model to hydrological time series is limited by narrow parameter constraints (Guo et al., 2021b; Li et al., 2016).

This study aims to improve the prediction accuracy of daily streamflow time series by constructing a novel model with high applicability that can simultaneously capture seasonality, non-stationarity, long-term memory, and nonlinearity of daily streamflow. We propose the Fractional-differenced Dual-Threshold Double Autoregressive (FDTDAR) framework that systematically integrates seasonal normalization, fractional differencing for long-memory modeling, and a dual-threshold structure to capture regime-specific nonlinearities. Furthermore, we evaluate alternative residual distributions—comparing the standard Gaussian distribution against the heavy-tailed Student's t-distribution—to improve residual characterization. The framework is applied to daily streamflow data from 15 gauging stations located in the Yellow River Basin. For robust evaluation, the historical data are divided into a 70% calibration period and a 30% out-of-sample testing period to assess the one-day-ahead forecasting performance. The FDTDAR framework is rigorously evaluated against both the classical linear AR-GARCH model and the long short-term memory (LSTM) network. Evaluation metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of determination (R2), Nash-Sutcliffe efficiency coefficient (NSE), and Absolute Maximum Error (AME), providing a multi-faceted assessment of model robustness.

2 Study area and data

2.1 Study area

The research area of this study (Figure 1) is the Yellow River Basin in China, which has a main stream length of approximately 5464 km and a catchment area of about 795,000 km2, ranking second in China and fifth in the world. The Yellow River originates from the Bayan Har Mountains in Qinghai Province, Western China, and flows through 9 provinces before emptying into the Bohai Sea. This study utilizes daily streamflow records from 15 hydrological stations in the Yellow River Basin. These stations are selected based on rigorous criteria to ensure statistical robustness: (i) a minimum continuous record length of 10

years; (ii) a data completeness requirement of 100%, with zero missing daily values throughout the selected study period to ensure continuous time-series modeling; and (iii) spatial representativeness covering the upper, middle, and lower reaches of the basin. These stations (Figure 1) include those located on the mainstream (Tangnaihai (TNH), Lanzhou (LZ), Xiaheyan (XHY), Shizuishan (SZS), Toudaoguai (TDG), Fugu (FG), Longmen (LM), Tongguan (TG), Sanmenxia (SMX), Aishan (AS), and Lijin (LJ)) and the largest tributary Weihe River Basin (Zhangjiashan (ZJS), Xianyang (XY), Lintong (LT), and Zhuangtou (ZT)).

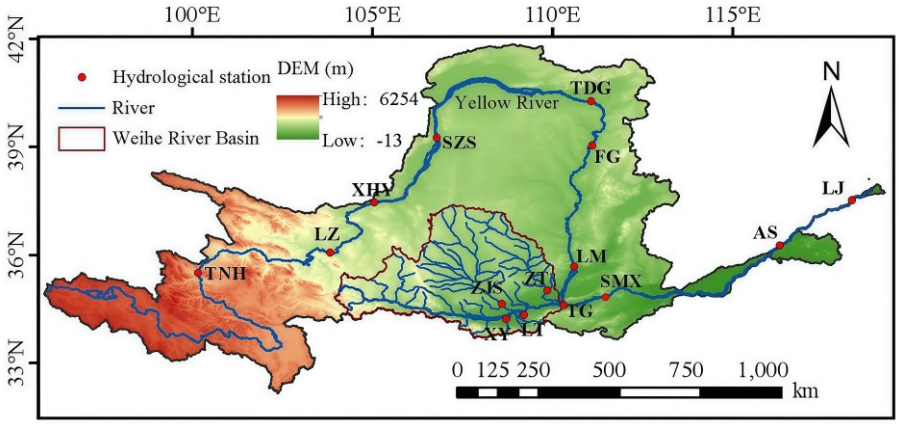

Figure 1: The Yellow River basin and locations of 15 hydrological stations.

2.2 Daily streamflow time series data

This study selects the measured daily streamflow time series from 15 hydrological stations in the Yellow River Basin (Table 1). Figure S1 depicts the temporal variation of the daily streamflow time series for the 15 hydrological stations. The daily streamflow record for each station is partitioned chronologically: the initial 70% of the data serves as the calibration period for identifying statistical characteristics and estimating model parameters, while the remaining 30% is reserved as an independent testing set. This setup allows for assessing the model fit during calibration and conducting a robust, out-of-sample evaluation of prediction accuracy across the various models.

Table 1: Basic information of daily streamflow series for 15 stations.

Stations	Abbreviation	Water system	Latitude/N	Longitude/E	Control area/km^2	Period
Tangnaihai	TNH	Yellow River	35.5	100.15	121972	2007.01.01-2022.12.31
Lanzhou	LZ	Yellow River	36.06	103.82	222600	2009.01.01-2022.12.31

Xiaheyan	XHY	Yellow River	37.45	105.05	254142	2007.01.01-2021.12.31
Shizuishan	SZS	Yellow River	39.25	106.78	309146	2007.01.01-2022.12.31
Toudaoguai	TDG	Yellow River	40.27	111.06	367898	2009.01.01-2022.12.31
Fugu	FG	Yellow River	39.03	111.08	404000	2007.01.01-2022.12.31
Longmen	LM	Yellow River	35.67	110.58	497552	2009.01.01-2022.12.31
Tongguan	TG	Yellow River	34.6	110.3	682141	2007.01.01-2022.12.31
Sanmenxia	SMX	Yellow River	34.82	111.37	688421	2007.01.01-2022.12.31
Aishan	AS	Yellow River	36.25	116.3	749136	2007.01.01-2022.12.31
Lijin	LJ	Yellow River	37.52	118.3	751869	2007.01.01-2022.12.31
Zhangjiashan	ZJS	Jinghe	34.64	108.59	43216	2007.01.01-2022.12.31
Xianyang	XY	Weihe River	34.32	108.7	46827	1985.01.01-2022.12.31
Lintong	LT	Weihe River	34.43	109.2	97299	1985.01.01-2007.12.31

Zhuangtou	ZT	Beiluohe	35	109.84	25154	2007.01.01-2022.12.31

3 Methodology

3.1 De-seasonal standardization

For a given original daily streamflow time series, the mathematical expression for deseasonalization is as follows:

$$\begin{cases} \mu^m = \frac{1}{N}\sum_{n=1}^{N} X_t^{(n,m)} \\ \sigma^m = \left[\frac{1}{N}\sum_{n=1}^{N}\left(X_t^{(n,m)} - \mu^m\right)^2\right]^{0.5} \end{cases} \quad (1)$$

$$Y_t = \frac{X_t - \mu^m}{\sigma^m} \quad (2)$$

where $X_t^{(n,m)}(n = 1, \cdots, N; m = 1, \cdots, M)$ (M=366 in a leap year and 365 in a normal year) is the original daily streamflow time series of the m-th day in the n-th year, μ^m and σ^m are the mean and standard deviation of the variable for each calendar day, respectively; Y_t $(t = 1, \cdots, T)$ represents the daily streamflow time series after removing seasonality, and the training period length is T.

3.2 Statistical diagnostic methods for streamflow series

Before initiating the modeling procedure, it is essential to understand the inherent statistical properties of daily streamflow time series. The primary objective of this section is to introduce a comprehensive diagnostic analysis to evaluate four key characteristics of time series: long-term persistence, non-stationarity, ARCH effects, and nonlinearity (Figure 2). These features are critical in determining the suitability and effectiveness of various modeling approaches. Moreover, understanding their interconnections provides important theoretical support for the logical consistency and robustness of the subsequent modeling framework (Figure 2).

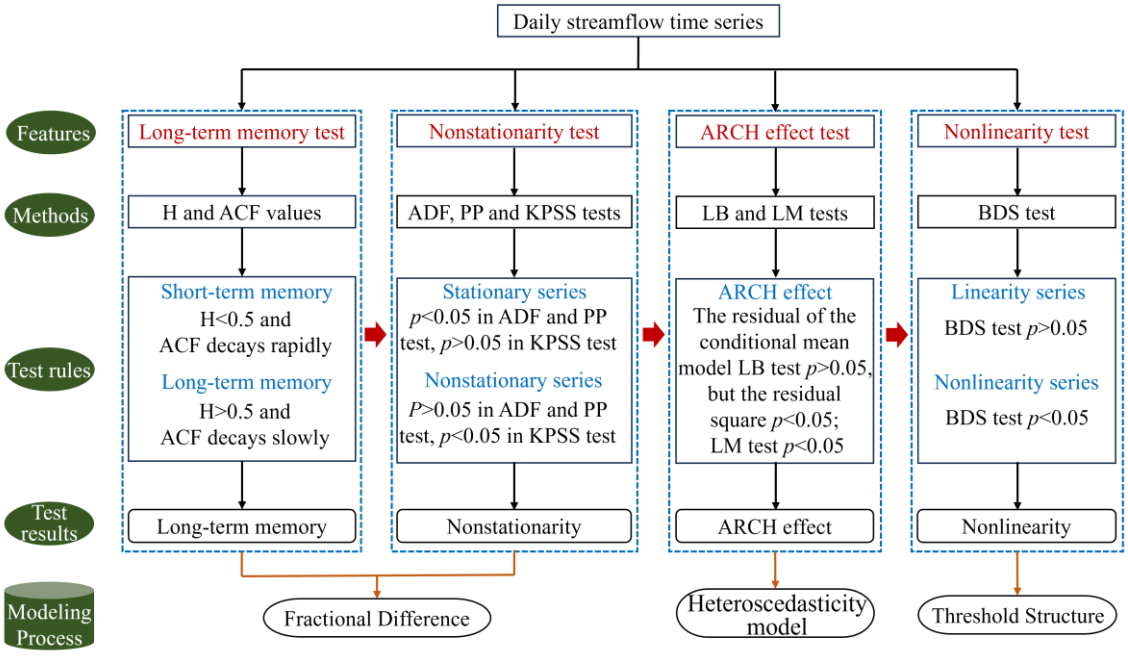

Figure 2: Schematic flowchart of the analysis pipeline, illustrating the sequence of statistical diagnostic tests and the motivated selection of the FDTDAR framework components.

The daily streamflow time series exhibits seasonal characteristics due to the cyclical influence of the four seasons. Therefore, deseasonalization is a necessary preprocessing step before modeling. In this section, the presence of long-term memory in the daily streamflow series is identified by the Hurst exponent (H) calculated using the rescaled range (R/S) analysis method and the trend of the autocorrelation function (ACF). The influence of seasonality on long-term memory is also assessed (Lo, 1991). On this basis, the stationarity of the series is evaluated using three unit root tests: the Augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests (Dickey and Fuller, 1981; Kwiatkowski et al., 1992; Peter and Perron, 1988). The null hypotheses of the ADF and PP tests assume that the time series has a unit root, while the KPSS test considers the null hypothesis as the time series being a stationary process. Due to the tendency of unit root tests to favor the null hypothesis, to rigorously determine the non-stationarity of the daily streamflow time series, if any of the three methods indicates the presence of a unit root process, then the daily streamflow time series at that station is considered non-stationary. Stationarity is a fundamental prerequisite for model construction. The method of achieving stationarity differs depending on the presence of long memory: non-long-memory series are differenced using integer orders, while long-memory series require fractional differencing. Subsequently, the Lagrange Multiplier (LM) tests is employed to detect the presence of ARCH effects, which is the most critical link in choosing to build a heteroscedastic model in this study (Ljung and Box, 1978). The ARCH effect reflects the limitations of linear regression models and highlights the necessity of characterizing the variation in second-order moments. Some studies have suggested a close relationship between ARCH effects and long-term memory in hydrological time series, neglecting long-term memory may lead to spurious detection of

ARCH effects (Wang et al., 2023b; Wang et al., 2023c). Finally, the Brock, Dechert, and Scheinkman (BDS) test is applied to examine whether the series exhibited nonlinear dynamics (Broock et al., 1996). The null hypothesis (H0) of the BDS test is that the time series is independently and identically distributed (i.i.d.). If the null hypothesis is rejected ($p < 0.05$), it indicates that the series is not i.i.d. and is a nonlinear stochastic process. Since most existing heteroscedastic models are constructed based on linear structures, if the series demonstrates significant nonlinearity, it may be necessary to improve the model expression accordingly.

3.3 Fractional differenced dual-threshold double autoregressive (FDTDAR) model

In this section, we developed the Fractional-differenced Dual-Threshold Double Autoregressive (FDTDAR) model. This framework integrates fractional integration to account for long-memory effects within a non-linear threshold-based double autoregressive structure. The daily streamflow time series always exhibits long-term memory, non-stationarity, conditional heteroscedasticity, and nonlinear characteristics. Therefore, these properties should be comprehensively considered in the model construction process. The model proposed in this study, referred to as the FDTDAR model, is an extension of the ordinary double autoregressive (DAR) model, a type of heteroscedasticity model. Specifically, the long memory characteristics of daily streamflow are described using the fractional difference method. On the premise of satisfying stationarity, the threshold idea is considered in the DAR model, that is, in the process of characterizing the first-order moment and the second-order moment, the threshold is introduced to segment, so as to express the nonlinear changes of the daily streamflow series. This processing method is inspired by the threshold autoregression (TAR) model. The FDTDAR model is therefore capable of accommodating the coexistence of multiple complex features in the daily streamflow series. The mathematical formulation of the model is presented as follows:

The fractional difference is used to describe the long-term memory of the daily streamflow series (Y_t) after removing the seasonal components. The specific mathematical form is expressed as:

$$y_t = (1 - L)^d Y_t \tag{3}$$

where y_t is the daily streamflow series after difference, d is the difference order, and $LY_t = Y_{t-1}$ is the lag operator. In order to facilitate the calculation, the binomial expansion of $(1 - L)^d$ is performed as follows:

$$(1 - L)^d = 1 - dL + \frac{d(d-1)L^2}{2!} - \frac{d(d-1)(d-2)L^3}{3!} + \cdots\cdots \tag{4}$$

Eq. (4) shows that infinite data information before the observation point needs to be used in the difference process, which is almost impossible in actual calculation, and a reasonable approximation is necessary. Therefore, Eq. (4) is substituted into Eq. (3) and sorted to obtain the difference formula used in this study. The specific structure is as follows:

$$y_t = \sum_{k=0}^{t} \omega_k \cdot Y_{t-k} \tag{5}$$

$$\omega_k = -\omega_{k-1} \frac{d-k+1}{k} \tag{6}$$

where the weight ω_k ($\omega_0 = 1$) means that the amount of information required for each data point is different, and k represents the time lag.

The fractional differenced DAR (FDAR) combines the first- and second-order moment of time series, and its general structure is as follows (Ling, 2007):

$$y_t = \varphi + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t \sqrt{\alpha + \sum_{i=1}^{p} b_i y_{t-i}^2} \tag{7}$$

where y_t is the daily streamflow time series processed by fractional difference (Eq. (5)). φ and α are constants, $a_1 \cdots a_p$ and $b_1 \cdots b_p$ are the coefficients of the FDAR model with the order p, ε_t represents the residual series with mean 0 and variance 1. Eq. (7) shows that the mathematical form used to describe the first-order and second-order moment characteristics of the time series in FDAR model are linear structures, that is $A = \varphi + \sum_{i=1}^{p} a_i y_{t-i}$ and $B = \omega + \sum_{i=1}^{p} b_i y_{t-i}^2$. Thus, both parts are assigned thresholds r_1 and r_2 to build the FDTDAR model, and the corresponding lag orders are d_1 and d_2, respectively. The Θ is the set of model parameters. When $d_1 \neq d_2$, the mathematical form of the FDTDAR model is expressed as:

$$y_t = \begin{cases} \varphi_{11} + \sum_{i=1}^{p_{11}} a_{1i}^1 y_{t-i} + \varepsilon_t \sqrt{\alpha_{11} + \sum_{i=1}^{q_{11}} b_{1i}^1 y_{t-i}^2} & if\ y_{t-d_1} \le r_1 \ \ y_{t-d_2}^2 \le r_2 \\ \varphi_{21} + \sum_{i=1}^{p_{21}} a_{1i}^2 y_{t-i} + \varepsilon_t \sqrt{\alpha_{21} + \sum_{i=1}^{q_{12}} b_{1i}^2 y_{t-i}^2} & if\ y_{t-d_1} > r_1 \ \ y_{t-d_2}^2 \le r_2 \\ \varphi_{12} + \sum_{i=1}^{p_{12}} a_{2i}^1 y_{t-i} + \varepsilon_t \sqrt{\alpha_{12} + \sum_{i=1}^{q_{21}} b_{2i}^1 y_{t-i}^2} & if\ y_{t-d_1} \le r_1 \ \ y_{t-d_2}^2 > r_2 \\ \varphi_{22} + \sum_{i=1}^{p_{22}} a_{2i}^2 y_{t-i} + \varepsilon_t \sqrt{\alpha_{22} + \sum_{i=1}^{q_{22}} b_{2i}^2 y_{t-i}^2} & if\ y_{t-d_1} > r_1 \ \ y_{t-d_2}^2 > r_2 \end{cases} \tag{8}$$

where φ_{11}, φ_{21}, φ_{12}, φ_{22}, α_{11}, α_{21}, α_{12} and α_{22} are constants, represented by the set C; the order of the model is p_{11}, p_{21}, p_{12}, p_{22}, q_{11}, q_{12}, q_{21}, and q_{22}, and the set is O; $a_{11}^1 \cdots a_{1p_{11}}^1$, $a_{11}^2 \cdots a_{1p_{21}}^2$, $a_{21}^1 \cdots a_{2p_{12}}^1$, $a_{21}^2 \cdots a_{2p_{22}}^2$ and $b_{11}^1 \cdots b_{1q_{11}}^1$, $b_{11}^2 \cdots b_{1q_{12}}^2$, $b_{21}^1 \cdots b_{2q_{21}}^1$, $b_{21}^2 \cdots b_{2q_{22}}^2$ represent the first-order and second-order moment coefficients, respectively, which expressed by the set E. When $d_1 = d_2 = d_0$, the threshold of FDTDAR (Θ) model is $r_1 = r_2 = r_0$, and its general form is expressed as:

$$y_t = \begin{cases} \varphi_{10} + \sum_{i=1}^{p_1} a_{1i} y_{t-i} + \varepsilon_t \sqrt{\alpha_{10} + \sum_{i=1}^{q_1} b_{1i} y_{t-i}^2} & if\ y_{t-d_0} \le r_0 \\ \varphi_{20} + \sum_{i=1}^{p_2} a_{2i} y_{t-i} + \varepsilon_t \sqrt{\alpha_{20} + \sum_{i=1}^{q_2} b_{2i} y_{t-i}^2} & if\ y_{t-d_0} > r_0 \end{cases} \tag{9}$$

where φ_{10}, φ_{20}, α_{10} and α_{20} are constants, $a_{11} \cdots a_{1p_1}$, $a_{21} \cdots a_{2p_2}$, $b_{11} \cdots b_{1q_1}$, and $b_{21} \cdots b_{2q_2}$ represent the coefficients of the model with orders p_1, p_2, q_1 and q_2.

In fact, Li et al. (2016) proposed a similar form of Eq. (9) ($d_1 = d_2$) and further studied the quasi-maximum likelihood estimation of the model. The significant difference between this model and the TAR-GARCH model is that the former makes the dynamic behavior of the conditional variance visible by specifying it in the observation function. Daren and Huay-min (2004) investigated the structure of such models in general settings, such as the strict stationarity and V-uniform ergodicity. The FDTDAR model we proposed (Eq. (8)) is further extended based on the work of previous researchers. For a further discussion on the structure of this model, please refer to Li et al. (2016) and Daren and Huay-min (2004).

For clarity, the term DAR refers to the standard linear double autoregressive model, while DT-DAR denotes the model with threshold-switching nonlinearities. The term FDTDAR is used exclusively for our proposed final model, which further integrates fractional differencing to handle long-range dependencies.

3.4 Parameter estimation of FDTDAR model

The parameters of the FDTDAR model are estimated using the quasi-maximum likelihood estimation (QMLE) method, considering both the Gaussian distribution (FDTDAR-n) and the Student's t distribution (FDTDAR-t) for the residuals.

(1) Gaussian distribution

Assume that the time series y_t is a sample from the FDTDAR model (taking Eq. (8) as an example), given an initial value $\{y_{t-p}, \cdots, y_0\}$, where $p = max\{p_{11}, p_{21}, p_{12}, p_{22}, q_{11}, q_{12}, q_{21}, q_{22}\}$. The conditional log-likelihood function is defined as:

$$Ln(\theta) = \sum_{t=1}^{T} l_t(\theta) = \sum_{t=1}^{T} \left[-\frac{1}{2} log h_t(\theta) - \frac{[y_t - u_t(\theta)]^2}{2h_t(\theta)} \right] \quad (10)$$

where θ is the set of model parameters, $\theta = (C', E', r_1, r_2)' \equiv (C', A', B', r_1, r_2)'$. And $C' = (\varphi_{11}, \varphi_{21}, \varphi_{12}, \varphi_{22}, \alpha_{11}, \alpha_{21}, \alpha_{12}, \alpha_{22})'$, $A = (A_1, A_2, A_3, A_4)$, $A_1 = \left(a_{11}^1 \cdots a_{1p_{11}}^1\right)$, $A_2 = \left(a_{11}^2 \cdots a_{1p_{21}}^2\right)$, $A_3 = \left(a_{21}^1 \cdots a_{2p_{12}}^1\right)$, $A_4 = \left(a_{21}^2 \cdots a_{2p_{22}}^2\right)$; $B_1 = \left(b_{11}^1 \cdots b_{1q_{11}}^1\right)$, $B_2 = \left(b_{11}^2 \cdots b_{1q_{12}}^2\right)$, $B_3 = \left(b_{21}^1 \cdots b_{2q_{21}}^1\right)$, $B_4 = \left(b_{21}^2 \cdots b_{2q_{22}}^2\right)$. $u_t(\theta)$ and $h_t(\theta)$ represent the conditional mean and conditional variance of the FDTDAR model, respectively. Their specific mathematical expressions are:

$$\begin{aligned} u_t(\theta) = {} & \left(\varphi_{11} + A_1' Z_{1,t-1}\right) I\left(y_{t-d_1} \le r_1, y_{t-d_2}^2 \le r_2\right) \\ & + \left(\varphi_{21} + A_2' Z_{2,t-1}\right) I\left(y_{t-d_1} > r_1, y_{t-d_2}^2 \le r_2\right) \\ & + \left(\varphi_{12} + A_3' Z_{3,t-1}\right) I\left(y_{t-d_1} \le r_1, y_{t-d_2}^2 > r_2\right) \\ & + \left(\varphi_{22} + A_4' Z_{4,t-1}\right) I\left(y_{t-d_1} > r_1, y_{t-d_2}^2 > r_2\right) \end{aligned} \quad (11)$$

$$\begin{aligned} h_t(\theta) = {} & \left(\alpha_{11} + B_1' S_{1,t-1}\right) I\left(y_{t-d_1} \le r_1, y_{t-d_2}^2 \le r_2\right) \\ & + \left(\alpha_{21} + B_2' S_{2,t-1}\right) I\left(y_{t-d_1} > r_1, y_{t-d_2}^2 \le r_2\right) \\ & + \left(\alpha_{12} + B_3' S_{3,t-1}\right) I\left(y_{t-d_1} \le r_1, y_{t-d_2}^2 > r_2\right) \\ & + \left(\alpha_{22} + B_4' S_{4,t-1}\right) I\left(y_{t-d_1} > r_1, y_{t-d_2}^2 > r_2\right) \end{aligned} \quad (12)$$

where $Z_{1,t-1} = \left(1, y_{t-1}, \cdots, y_{t-p_{11}}\right)'$, $Z_{2,t-1} = \left(1, y_{t-1}, \cdots, y_{t-p_{21}}\right)'$, $Z_{3,t-1} = \left(1, y_{t-1}, \cdots, y_{t-p_{12}}\right)'$, $Z_{4,t-1} = \left(1, y_{t-1}, \cdots, y_{t-p_{22}}\right)'$; $S_{1,t-1} = \left(1, y_{t-1}^2, \cdots, y_{t-q_{11}}^2\right)'$, $S_{2,t-1} = (1, y_{t-1}^2, \cdots, y_{t-q_{21}}^2)'$, $S_{3,t-1} = \left(1, y_{t-1}^2, \cdots, y_{t-q_{12}}^2\right)'$, $S_{4,t-1} = \left(1, y_{t-1}^2, \cdots, y_{t-q_{22}}^2\right)'$. $I(\cdot)$ is an indicator function.

Given the initial value of the observation, the quasi-maximum likelihood estimate (QMLE) of the parameter θ can be defined by the following formula:

$$\hat{\theta} = argmaxLn(\theta) \tag{13}$$

where $\hat{\theta}$ is the quasi-maximum likelihood estimator of parameter vector θ.

(2) Student's t distribution

The normal distribution is the most commonly used conditional distribution form in the DAR model. However, for the significant peak-tailed characteristics in hydrological time series, the heavy-tailed distribution of linear model residuals may be insufficient, and even the residual series obtained from volatility are required to be heavy-tailed. Therefore, the normal distribution assumption of the classic DAR model may be difficult to meet the needs of hydrological time series modeling. It is worthwhile to introduce skewed conditional distribution into the DAR model to enhance its ability to characterize the heavy-tail characteristics of time series.

The expansion of the conditional distribution in the classic ARCH model starts from the Student's t distribution. Regarding the ARCH model, Engle (1982, 1987, 1990, 1993) mentioned after extensive research that the conditional distribution of the volatility model can adopt a non-Gaussian form. Therefore, Bollerslev (1986) first used the student's t distribution to describe the distribution form of the residuals. This study used the student's t distribution to expand the conditional distribution of the DAR model.

Therefore, the student's t distribution is considered in the FDTDAR model, that is $\varepsilon_t \xrightarrow{i.i.d} t(0,1;k)$, where k is the degree of freedom of the student's t distribution. The most common probability density function of the student's t distribution is expressed as:

$$f(y_t,k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{y_t^2}{k}\right)^{\frac{-(k+1)}{2}} \tag{14}$$

where $\Gamma(\cdot)$ is the Gamma function. M_t is used to represent the scale parameter of the FDTDAR model, then the probability density function of the time series y_t is expressed as:

$$f(y_t|M_t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)}\frac{1}{\sqrt{M_t}}\left[1+\frac{(y_t-u_t)^2}{M_t k}\right]^{\frac{-(k+1)}{2}} \tag{15}$$

When the degrees of freedom k>2, the mean of the residual series ε_t is 0, and the variance h_t is expressed as:

$$E(y_t^2) = \frac{M_t k}{k-2} \tag{16}$$

Therefore, for series y_t with variance h_t and freedom k, the scale parameter M_t can be written as:

$$M_t = \frac{h_t(k-2)}{k} \tag{17}$$

The probability density function then becomes:

$$f(y_t|h_t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{(k-2)\pi}\Gamma\left(\frac{k}{2}\right)}\frac{1}{\sqrt{h_t}}\left[1+\frac{(y_t-u_t)^2}{h_t(k-2)}\right]^{\frac{-(k+1)}{2}} \tag{18}$$

where the freedom degree k is to be estimated, the conditional log-likelihood function is:

$$Ln(\theta) = \sum_{t=1}^{T} l_t(\theta) = \sum_{t=1}^{T} ln[f(y_t|h_t)]$$

$$= \sum_{t=1}^{T} ln\left[\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{(k-2)\pi}\Gamma\left(\frac{k}{2}\right)}\frac{1}{\sqrt{h_t}}\left[1+\frac{(y_t-u_t)^2}{h_t(k-2)}\right]^{\frac{-(k+1)}{2}}\right] \tag{19}$$

$$= -\frac{T}{2}ln\left[\frac{\pi(k-2)\Gamma\left(\frac{k}{2}\right)^2}{\Gamma\left(\frac{k+1}{2}\right)^2}\right] - \frac{1}{2}\sum_{t=1}^{T} lnh_t - \frac{k+1}{2}\sum_{t=1}^{T} ln\left[1+\frac{(y_t-u_t)^2}{h_t(k-2)}\right]$$

where u_t and h_t are calculated by Eq. (11) and Eq. (12), respectively. If there exists a vector $\hat{\theta}$ such that $Ln(\theta)$ has the maximum value, then in the setting of maximum likelihood estimation, vector $\hat{\theta}$ is considered as the maximum likelihood estimator for the parameter vector θ, and it is determined by Eq. (13).

In this study, the quasi-maximum likelihood estimation under both residual distributions is implemented using the *nlminb* function in R version 4.4.1, which is an efficient numerical optimization tool. It returns the negative value of the likelihood function; therefore, the actual likelihood value used in the estimation is the negative of the function's output. For both residual distribution scenarios, the model orders (p_{11}, p_{21}, p_{12}, p_{22}, q_{11}, q_{12}, q_{21}, and q_{22}; p_1, p_2, q_1 and q_2) are exhaustively searched in the range of 0 to 10. The initial values of the parameters are determined based on the autocorrelation functions at different lag orders up to the corresponding model order. Additionally, the *nlminb* function allows for setting constraints on parameter values, which are specified based on the theoretical requirements of the model. Specifically, the parameters associated with the first-order moments are allowed to vary over the entire real line, while the parameters related to the second-order moments are constrained to be greater than 0. The degrees of freedom of the Student's t distribution are estimated jointly with the other model parameters, and a constraint of greater than 2 is imposed to ensure numerical stability.

3.5 Order determination of the FDTDAR model

Obviously, as the order increases, the FDTDAR model's ability to describe time series becomes stronger. However, the accompanying large parameter set adds complexity to the model structure and reduces computational speed. Therefore, it is necessary to use a metric tool to determine the optimal order for the model. The evaluation of model quality generally depends on two aspects: the likelihood function value of parameter estimation and the number of unknown parameters in the model. A larger likelihood function value and a greater number of parameters indicate superior model fitting. However, the risk of "overfitting" arises when the simulation performance is excessively superior, leading to a decrease in accuracy during the prediction phase.

The process of model order determination is essentially an optimization task aimed at balancing the two aspects mentioned above. In practical computations, we aim for a larger likelihood function value while minimizing the number of model parameters. The Bayesian Information Criterion (BIC) exhibits outstanding effectiveness in the context of model order determination. The calculation formula of BIC based on DTDAR model is:

$$\boldsymbol{BIC = kln(n) - 2ln(\hat{L})} \quad (20)$$

where n is the sample size, k is the number of parameters, and $\hat{\boldsymbol{L}}$ is the maximized value of the likelihood function of the model.

During the order determination process, we restrict the value of p within the range of 1 to 25 to select the minimum BIC value. And $p_{ij}\ (i,j=1,2)$ and $q_{ij}\ (i,j=1,2)$ take values within $[1,p]$, $d_1 \in [1, min(p_{ij})]$, $d_2 \in [1, min(q_{ij})]$, $r_1 \in [min(y_t), max(y_t)]$, $r_2 \in [min(y_t^2), max(y_t^2)]$.

3.6 FTAR-GARCH model

To evaluate the performance of the proposed FDTDAR model, we compare it with the FTAR-GARCH model, a widely used long-memory heteroscedastic model that also incorporates threshold effects.

The TAR model (Tong, 1983) is based on the AR model by adding a threshold to achieve the nonlinear description of the time series mean behavior. The mathematical structure of a two-stage TAR model is written as:

$$y_t = \begin{cases} \omega_{10} + \sum_{i=1}^{p^1} \beta_i^1 y_{t-i} + e_t^1 & if\ y_{t-c} \le \tau \\ \omega_{20} + \sum_{i=1}^{p^2} \beta_i^2 y_{t-i} + e_t^2 & if\ y_{t-c} > \tau \end{cases} \quad (21)$$

where ω_{10} and ω_{20} are constants, β_i^1 and β_i^2 represent the coefficients of model with orders p^1 and p^2, the e_t^1 and e_t^2 are residuals with mean 0 and variance σ^2, and c and τ express specified time delay and threshold, respectively. The conditional heteroscedasticity information in the residual series of the TAR model is captured by the GARCH (1,1) model (Bollerslev, 1986), and its specific form is as follows:

$$\sigma_t^2 = \omega + \nu e_{t-1}^2 + \theta \sigma_{t-1}^2 \quad (22)$$

$$e_t = \sigma_t \eta_t \quad \eta_t \sim N(0,1) \quad (23)$$

where ω is a constant, ν and θ are coefficients of the GARCH model, σ_t^2 is the conditioned time-varying variance of the residual series.

3.7 Long short-term memory network

To evaluate the proposed FDTDAR framework against modern nonlinear data-driven approaches, a long short-term memory (LSTM) network is implemented as a benchmark. LSTM is a specialized type of recurrent neural network (RNN) designed to overcome the vanishing gradient problem, making it particularly effective for capturing long-term dependencies in daily streamflow time series (Hochreiter and Schmidhuber, 1997; Kratzert et al., 2018). The core of the LSTM architecture is the memory cell state (ct), which acts as an "information highway" regulated by three specific gate mechanisms: the forget gate (ft), the input gate (it), and the output gate (ot). These gates determine what information to discard from the previous state, what new information to add, and what to pass as the hidden state (ht) to the next time step. In this study, the LSTM model is configured with one hidden layer of 128 units determined through a trial-and-error process. To ensure a fair comparison with the statistical models, the same 70%/30% chronological split for calibration and testing was applied. The model provides a

purely data-driven non-linear baseline to contrast with the statistically interpretable regime-switching logic of the FDTDAR framework.

3.8 Comparative evaluation methods

Six metrics are used in this study to compare and evaluate the prediction performance of the models, namely Mean Absolute Error (MAE), Mean Relative Error (MRE), Root Mean Squared Error (RMSE), Coefficient of determination (R^2), Nash-Sutcliffe efficiency coefficient (NSE), and Absolute Maximum Error (AME) (Dawson et al., 2007; Liu et al., 2020; Moriasi et al., 2007). The smaller the MAE and RMSE, the closer the R^2 and NSE are to 1, indicating that the model prediction performance is better.

In addition, the DM test is used to compare the accuracy of two models (model 1 and model 2), and its null hypothesis is that the prediction accuracy of model 1 is higher than that of model 2.

The residual diagnostics of the model are conducted using the autocorrelation function (ACF), by examining whether the residual autocorrelations at various lag times fall within the 95% confidence interval around 0. If ACF values lie within this interval, the residuals can be considered approximately white noise, indicating that the model has effectively captured the variation structure of the time series.

An overview of the process of building the model in this study is shown in Figure 3.

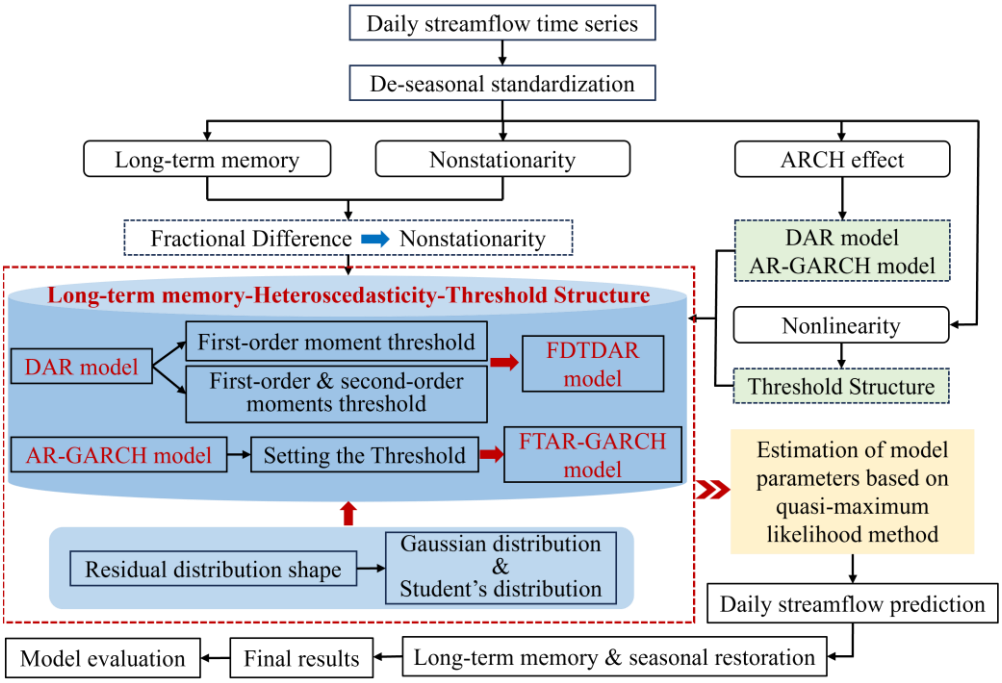

Figure 3: Methodological workflow of the Fractional-differenced dual-threshold double autoregressive (FDTDAR) framework for non-linear and long-memory streamflow prediction.

4 Results

4.1 Statistical diagnostics and model selection rationale

The results of various diagnostic tests on the daily streamflow time series are presented in Figure 4, Figure S2, and Tables S1–S3. The ACF (Figure 4) and the Hurst exponent results (Table S1) indicate that the daily streamflow series exhibits long-term memory characteristics, and that seasonality weakens the strength of this memory. The results of the 3 unit root tests (Table S2) reveal that the series is non-stationary. Considering its long-memory properties, fractional differencing was adopted to achieve stationarity, which satisfied the model assumptions. The LM test yields consistent results (Tables S4). The nonlinearity test results (Table S3) suggest that the daily streamflow time series exhibits significant nonlinear behavior, implying that the model structure in this study should be modified to accommodate nonlinear dynamics.

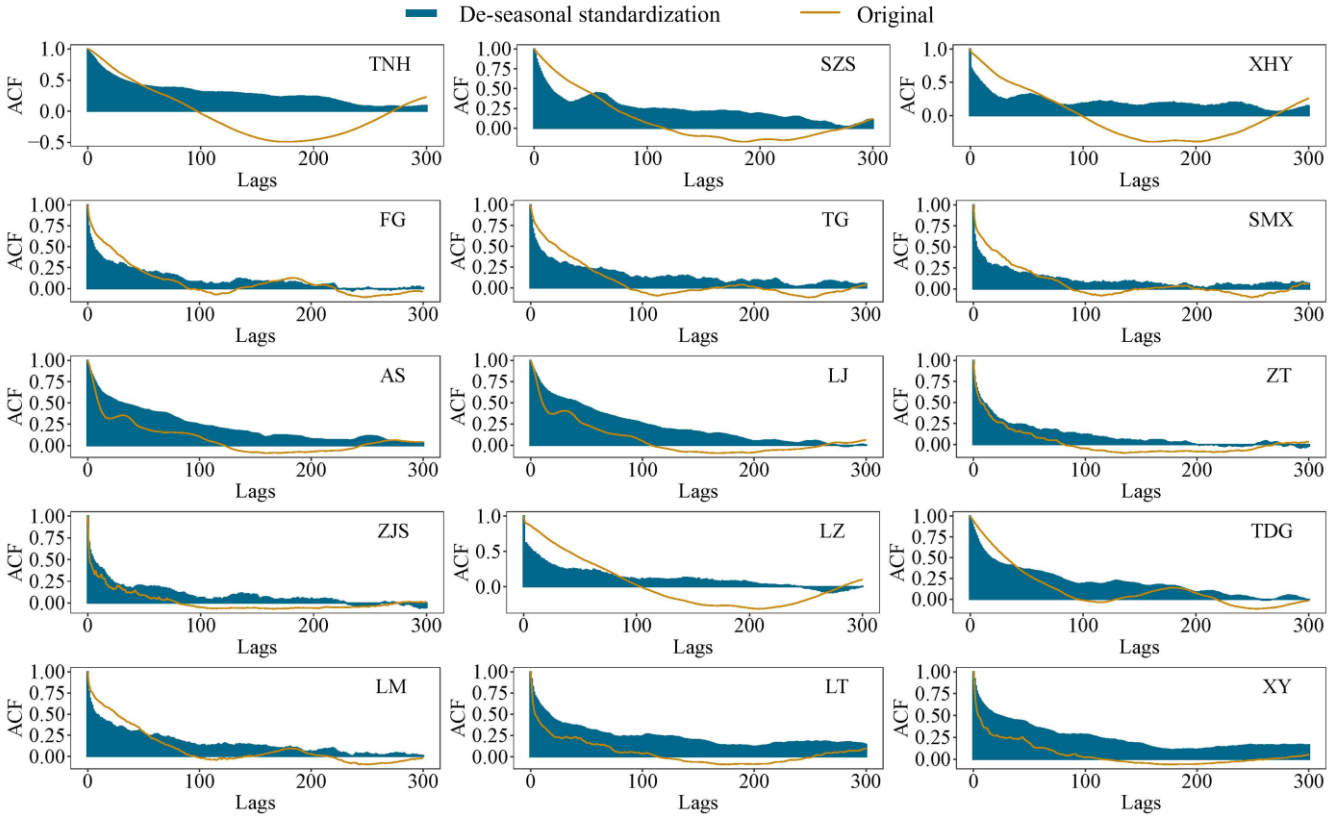

Figure 4: Autocorrelation coefficient (ACF) for original and de-seasonal standarded daily streamflow time series in 15 stations.

The diagnostic results presented above provide critical insights into the statistical characteristics of the daily streamflow time series, including long-term memory, non-stationarity, heteroscedasticity, and nonlinearity. These findings form a solid foundation for selecting appropriate modeling strategies in the subsequent sections. Specifically, the detection of long-term memory and non-stationarity justifies the use of fractional differencing methods to ensure model validity. The presence of

ARCH effects highlights the need for models capable of capturing time-varying volatility, while the identified nonlinear dynamics suggest that linear structural models may be insufficient. DAR-type models have emerged as strong competitors to the widely used heteroscedastic GARCH models, due to their simple mathematical formulation, ease of parameter estimation, and ability to accurately characterize real-world data. Based on the results of this section, the next section introduces an enhanced modeling framework (FDTDAR model) based on the DAR family, which integrates long-memory components, heteroscedasticity, and nonlinear dynamics to better capture the complex behavior of daily streamflow.

Figure 5 illustrates the presence of the FDTDAR model at various hydrological stations. In the FDTDAR models with equal lag times (Eq. (9)), the value of $\boldsymbol{d_0}$ is consistently 1, while in another form (Eq. (8)), the lag times ($\boldsymbol{d_1}$ and $\boldsymbol{d_2}$) have values of 1 to 3. Furthermore, the order of FDTDAR-n and FDTDAR-t models generally ranges from 1 to 5 days, except for ZT and LJ stations, where the FDTDAR-t model order reaches 8 and 10.

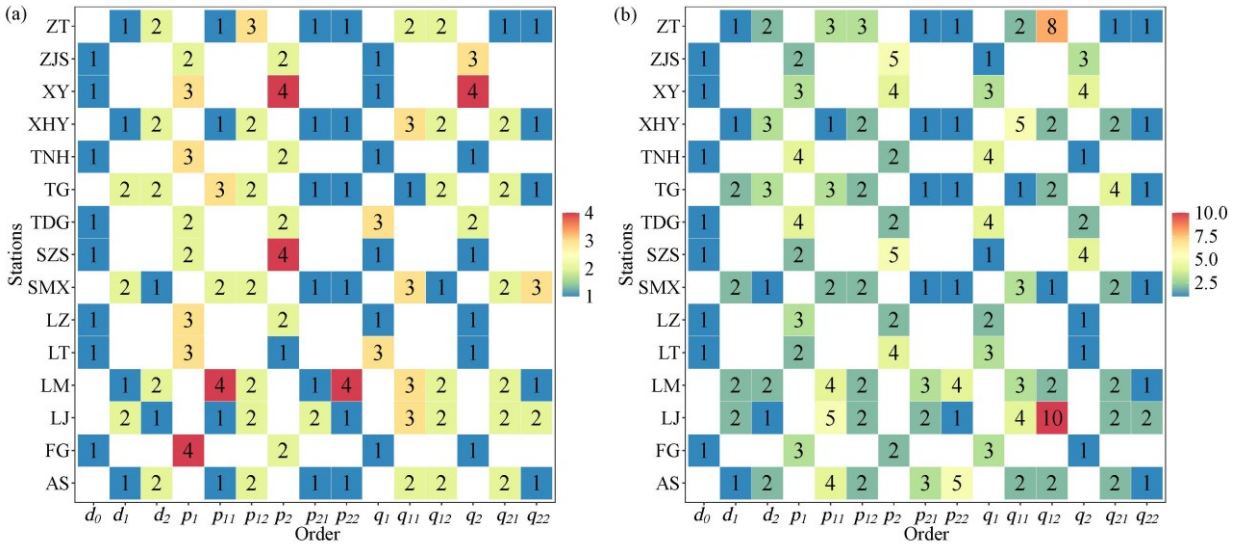

Figure 5: The order of the FDTDAR model for 15 hydrological stations, where (a) and (b) represent the normal distribution and the Student's t distribution, respectively.

Due to intensified climate change and enhanced human activities, the nonlinearity of daily streamflow time series may continue to strengthen. FDTDAR models with the same lag time essentially use a threshold at the mean level to simultaneously constrain the behavior of the first and second moments, which may be challenging in addressing the challenges brought about by nonlinear changes at the second-moment level in daily streamflow time series. Except for ZT and LJ stations, the complexity of the FDTDAR models under the two residual distribution forms is generally consistent.

The thresholds for the FDTDAR model with different residual distributions at 15 hydrological stations are presented in Table 2. The inputs for each model are daily streamflow time series that have undergone seasonal standardization and fractional differencing. The thresholds for the FDTDAR models based on two residual distributions are identical for the XHY, ZT, LZ, LM, LT and XY hydrological stations.

Table 2: Thresholds of the FDTDAR model in 15 hydrological stations, where n and t represent normal and Student's t distributions, respectively. r1, r2 and r0 are the thresholds for classifying streamflow into different flow regimes, where r1 and r2 are derived from Eq. (8) and r0 is derived from Eq. (9).

Stations	Distributions	r_1	r_2	r_0	Stations	Distributions	r_1	r_2	r_0
TNH	n			-80.35	ZT	n	-4.33	0.90	
	t			9.07		t	-4.33	0.90	
SZS	n			-63.93	ZJS	n			-4.67
	t			-4.74		t			8.67
XHY	n	-86.81	26.62		LZ	n			-139.90
	t	-86.81	26.62			t			-139.90
FG	n			-176.50	TDG	n			9.87
	t			270.43		t			-15.19
TG	n	-153.72	30.60		LM	n	-150.08	33.73	
	t	14.44	290.93			t	-150.08	33.75	
SMX	n	316.73	394.62		LT	n			-39.98
	t	-15.73	394.62			t			-39.98
AS	n	-16.43	69.82		XY	n			-21.80
	t	7.00	56.85			t			-21.80
LJ	n	-31.26	38.37						
	t	65.93	90.52						

4.2 Comparison between FDTDAR and FTAR-GARCH models

Figure 6 compares the predictive performance of various long memory threshold models for 15 hydrological stations. For mainstream hydrological stations, higher levels of average error (MAE and MRE) and extreme error (RMSE and AME), and lower R2 and NSE values, indicate poor predictive accuracy for the four models at the SMX station. For the vast majority of stations, the R^2 and NSE values of both FDTDAR and FTAR-GARCH models exceed 0.8, but the FDTDAR model yields smaller MAE and RMSE values. In addition, the shape of the residual distribution exerts a certain influence on predictive performance, with models based on the Student's t distribution exhibiting smaller error metrics and larger R^2 and NSE values.

Overall, compared with the FTAR-GARCH model, the FDTDAR model demonstrates superior predictive capability, and the Student's t distribution is more suitable than the normal distribution for characterizing the variability of daily streamflow.

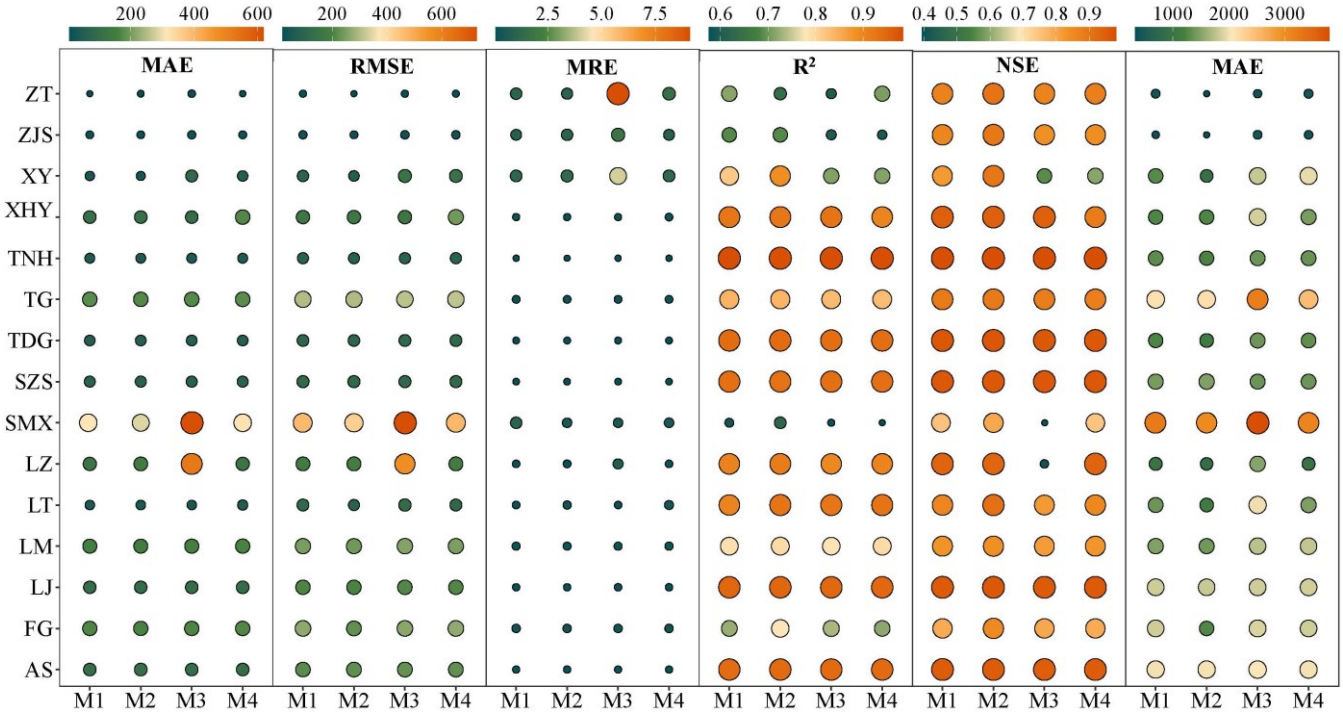

Figure 6: Prediction evaluation indicators of long memory threshold models, where M1, M2, M3 and M4 are FDTDAR-n, FDTDAR-t, FTAR-GARCH-n and FTAR-GARCH-t models, respectively.

Furthermore, Figure 6 illustrates that, in the case of four tributary hydrological stations (ZT, ZJS, LT, and XY), the FDTDAR type models exhibit relatively lower values of MAE, RMSE, MRE, and AME, while simultaneously having higher R^2 and NSE values compared to FTAR-GARCH models. The predictive results of FDTDAR and FTAR-GARCH models vary under different distribution assumptions, with the t-distribution yielding superior predictive performance. In short, FDTDAR models demonstrated outstanding predictive capabilities in the daily streamflow time series at tributary hydrological stations, and the t-distribution significantly improved the accuracy of model predictions.

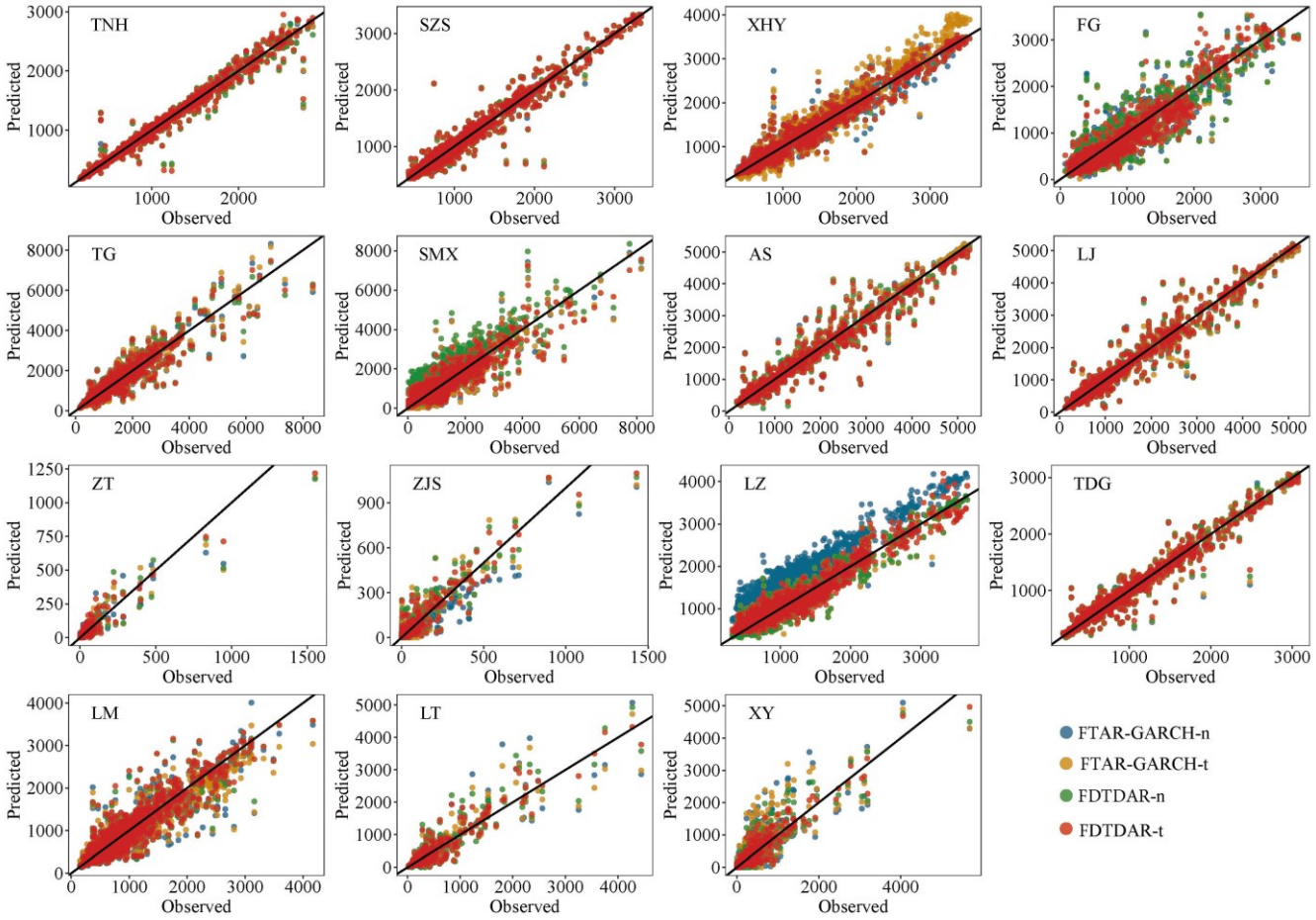

Figure 7: Scatter plots of observed versus predicted daily streamflow for the 15 stations during the testing period. Colors represent different model variants (FTAR-GARCH-n, FTAR-GARCH-t, FDTDAR-n, and FDTDAR-t).

Predictive performance of the FDTDAR-n, FDTDAR-t, FTAR-GARCH-n, and FTAR-GARCH-t models at 15 hydrological stations reveals that, for the majority of stations, the degree of clustering of predicted points around the 1:1 line is stronger for the FDTDAR models compared to the FTAR-GARCH models (Figure 7). Under the assumption of the Student's t distribution, both FDTDAR and FTAR-GARCH models exhibit better matching with observed daily streamflow compared to their counterparts assuming normal distribution (FDTDAR-n and FTAR-GARCH-n models). This suggests that the FDTDAR models have stronger predictive capabilities for daily streamflow time series compared to the FTAR-GARCH models, and the student's t distribution improves the predictive performance and peak description ability of the models under the classical normal distribution assumption.

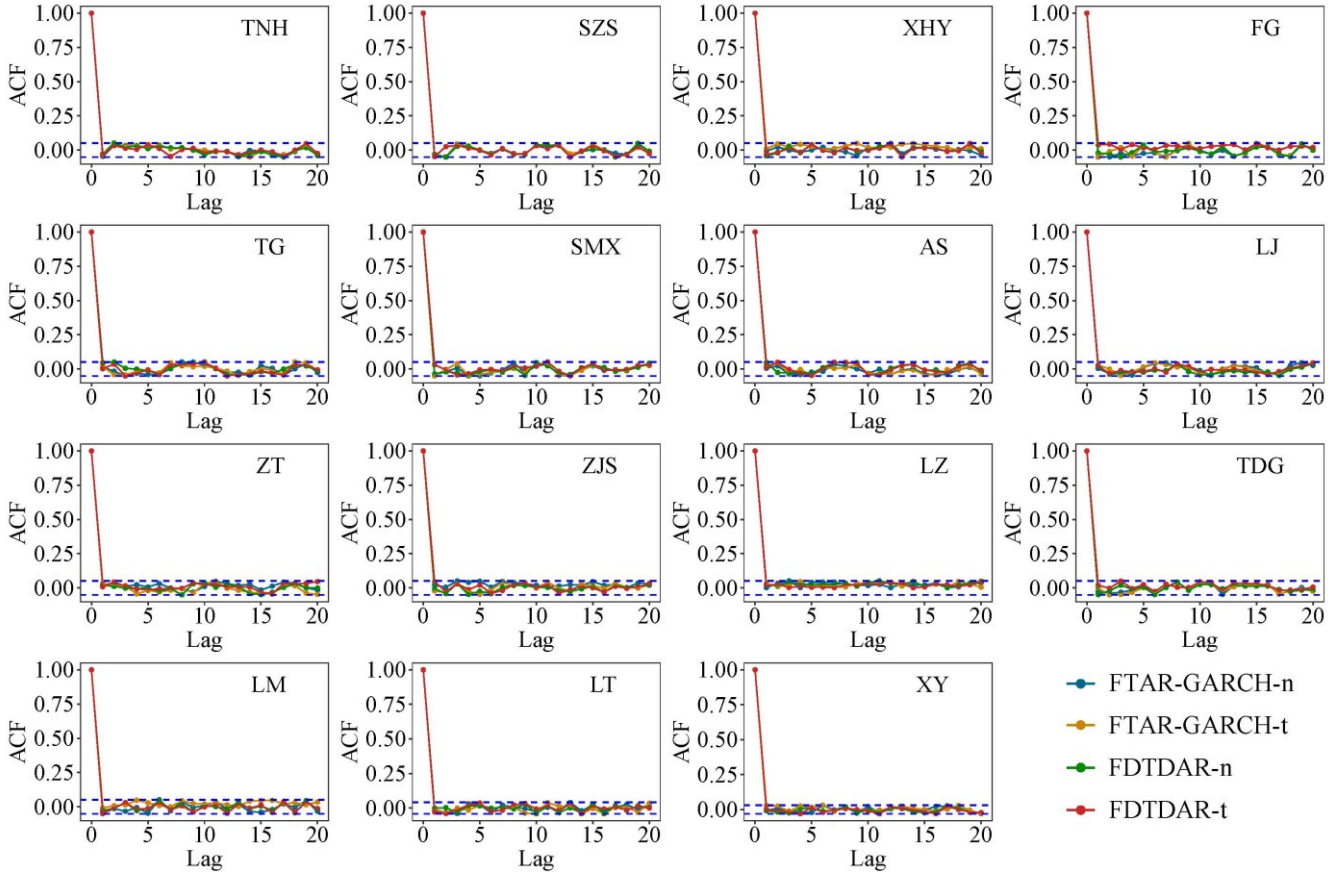

Figure 8: Autocorrelation function (ACF) plots of the standardized residuals for the 15 gauging stations during the testing period. The subplots compare the residual series derived from the FTAR-GARCH-n, FTAR-GARCH-t, FDTDAR-n, and FDTDAR-t models. The horizontal blue dashed lines represent the 95% confidence intervals. The decay of ACF values to within these significance bounds across all stations indicates that the model residuals are largely uncorrelated, approximating a white noise process.

Figure 8 shows that the ACF of the residual series from the long memory threshold FDTDAR and FTAR-GARCH models, based on two different residual distributions, are within the confidence intervals, indicating the absence of autocorrelation in the residual series. This implies that both model types effectively capture practical information in the daily streamflow time series, and the model predictions are reliable.

4.3 Effectiveness evaluation of long memory threshold structure

This section compares the prediction performance of DAR-type and AR-GARCH-type models under three different structures (integer differencing, long memory, and long memory threshold) and two residual distribution assumptions. In the point

prediction comparison, R^2 and NSE values are chosen to assess the prediction accuracy of the model. The DM test is employed to identify the optimal prediction model structure and type for each station.

Figure 9 reveals that, at most stations, under the "long memory threshold" structure that considers nonlinear changes, both AR-GARCH-type and DAR-type models exhibit higher R^2 and NSE values than the two linear structures. Furthermore, the FAR-GARCH and FDAR models, which account for long-term memory, significantly improve the NSE values of the classical AR-GARCH and DAR models based on integer differencing. This indicates that the long-term memory and strong nonlinearity in the daily streamflow time series significantly influence the predictive performance of the models, and the long memory threshold structure is effective in improving the accuracy of daily streamflow predictions.

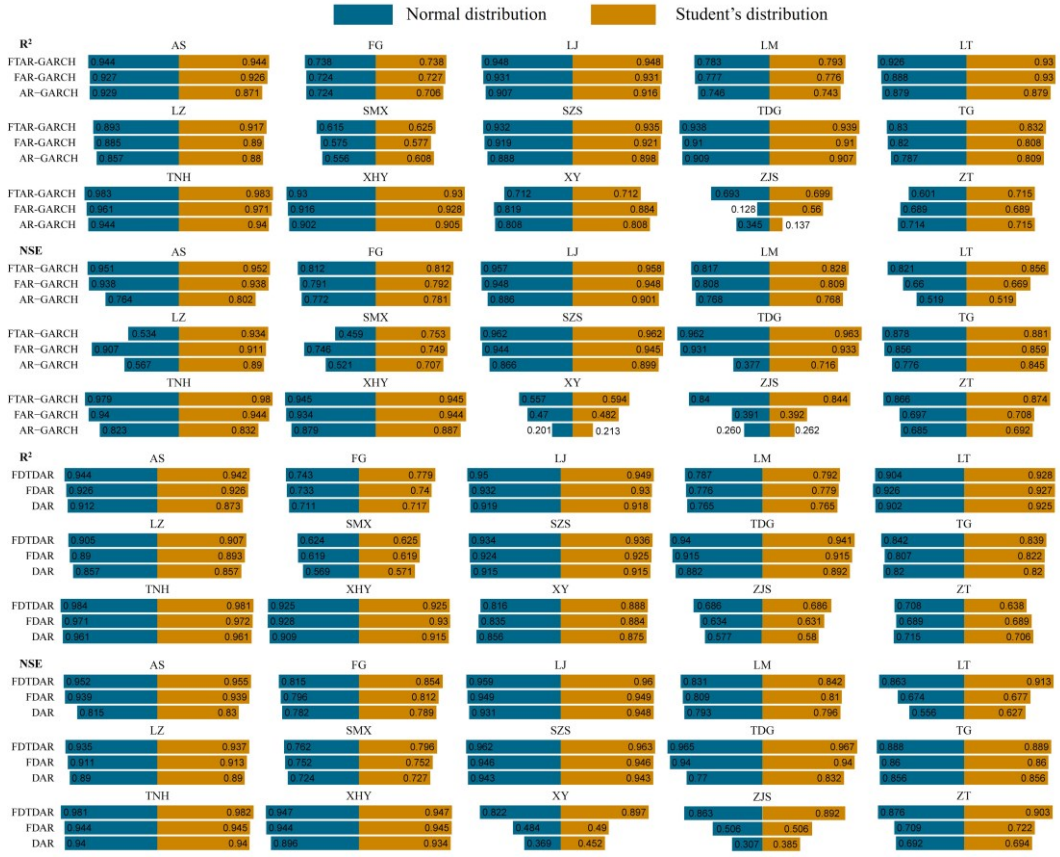

Figure 9: Comparison of prediction accuracy for DAR-type and AR-GARCH-type models based on three structures with two residual distributions at 15 stations. (Integer difference structure: AR-GARCH and DAR; Long memory structure: FAR-GARCH and FDACR; Long memory threshold structure: FTAR-GARCH and FDTDAR)

The comprehensive evaluation results of the predictive performance of DAR-type and AR-GARCH-type models under different structures are presented in Table 3. In the group of Linear vs. Nonlinear, the linear models are derived from the advantageous models in the difference group, while the two superior models in the residual distribution comparison groups come from the optimal models of each distribution. It can be observed that, except for the ZJS station, the long memory

structure based on fractional differencing effectively enhances the predictive ability of the two type models under traditional integer differencing. The *p*-values less than 0.05 indicate that the long-memory threshold structure significantly contributes to improving the prediction accuracy of the two type models (DAR and AR-GARCH) with linear changing characteristics (excluding the ZJS station). Additionally, the results of different residual comparison group indicate that XHY, FG, TG, SMX, AS, and TDG stations are more suitable for the t-distribution in DAR-type model, while in AR-GARCH-type, SZS, XHY, TG, AS, ZT, ZJS and LT stations exhibit this preference.

Table 3. *p*-value of DM test for DAR-type and AR-GARCH-type models under different structures (M1 vs. M2 shows that the original hypothesis of the DM test is that the prediction performance of M1 is better than M2).

Stations	DAR-type model					AR-GARCH-type model				
	Normal distribution		Student's t distribution			Normal distribution		Student's t distribution		
	Difference Integer vs. Fraction	Linear vs. Nonlinear	Difference Integer vs. Fraction	Linear vs. Nonlinear	n vs. t	Difference Integer vs. Fraction	Linear vs. Nonlinear	Difference Integer vs. Fraction	Linear vs. Nonlinear	n vs. t
TNH	0	0	0	0	0.95	0	0	0	0	0.01
SZS	0	0	0	0	0.98	0	0	0	0	0.96
XHY	0	0	0	0	0.00	0	0	0	0	0.95
FG	0	0	0	0	0.00	0	0	0	0	0.01
TG	0	0	0	0	0.00	0	0	0	0	0.98
SMX	0	0	0	0	0.00	0	0	0	0	0.01
AS	0	0	0	0	0.02	0	0.01	0	0	0.80
LJ	0	0	0	0	0.98	0	0	0	0	0.00
ZT	0	0	0	0	0.91	0	0	0	0	0.28
ZJS	0.05	0.07	0.06	0.06	0.16	0.1	0.07	0.06	0.08	0.98
LZ	0	0	0	0	0.51	0	0	0	0	0.00
TDG	0	0	0	0	0.00	0	0	0	0	0.00
LM	0	0	0	0	0.95	0	0	0	0	0.00
LT	0.01	0	0	0	0.25	0	0.02	0	0.01	0.26
XY	0	0.02	0	0.01	0.20	0	0.01	0	0.02	0.02

Further comparison was conducted on the preferred models of the two types at each station, with the null hypothesis (DM hypothesis) set as DAR-type model being superior to AR-GARCH-type. The optimal models for each station were ultimately determined, with XHY, FG, TG, SMX, AS, and TDG stations favouring the FDTDAR-t model, the DAR-n model performing well at the ZJS station, and the remaining stations adopting the FDTDAR-n model.

5 Discussions

5.1 Properties of daily streamflow time series

The daily streamflow time series is influenced by natural and human factors, presenting seasonality, long-term persistence, non-stationarity, and time-varying volatility (ARCH effect), which play a crucial role in simulation and prediction. However, the long-term persistence has been overlooked in most studies. The integer difference method is often used to address non-stationary issues to smooth the series (Myronidis et al., 2018; Yan et al., 2022), but it also eliminates the long-term memory information of daily streamflow, resulting in incomplete input information for the model and affecting the simulation and prediction accuracy. Some scholars (Mehdizadeh et al., 2019; O'Connell et al., 2016; Yang and Bowling, 2014) have also studied long-term persistence, demonstrating its importance in predicting the daily streamflow process. However, those researchers have not conducted the second-order moment components, making it impossible to prove whether the ARCH effect is related to long-term persistence. In recent years, an increasing number of researchers (Dimitriadis et al., 2021; Graves et al., 2017; Grimaldi, 2004) have reported that the detected ARCH effect in the classic AR model may be suspicious due to the lack of consideration of long-term memory, and the results of the further constructed time-varying fluctuation model are thought-provoking. This implies that the results obtained in existing studies on daily streamflow prediction by time series models may be biased due to an incomplete understanding of the characteristics of the research targets. Given the existing problems in the research, in addition to the seasonality and non-stationarity of the daily streamflow time series, this study also considers long-term persistence and time-varying volatility, which improved the prediction accuracy while ensuring the correctness of the ARCH effect.

In addition, time irreversibility is also an important property of streamflow time series, although it is negligible in other hydrological cycle components (e.g., precipitation, evapotranspiration) (Vavoulogiannis et al., 2021). Stationary processes in statistics are typically reversible over time, but the nonlinearity of underlying dynamic procedures is reflected in time irreversibility. River dynamics argues that the time irreversibility of streamflow sequences mainly depends on various triggering factors in the formation process and exhibits nonlinear changes. Specifically, the increase in streamflow is largely determined by short-term meteorological drivers, which are inherently the dynamic characteristics of precipitation, while groundwater dynamics respond to the streamflow reduction process (Serinaldi and Kilsby, 2016). Table S3 provides clear evidence of the nonlinear nature of daily streamflow, supporting the existence of time irreversibility. Threshold regression models such as TAR-GARCH and DTDAR have been successful in catching the nonlinear characteristics of daily streamflow time series, and numerous studies have confirmed their effectiveness in solving nonlinear problems. Therefore, the modeling approach used in this study, incorporating long memory, thresholds, and time-varying fluctuations, is more suitable for describing daily streamflow time series.

5.2 DAR-type models vs. AR-GARCH-type models

From the general situation of the model, both the DAR and the AR-GARCH models can grasp the first- and second-order moment information of the daily streamflow time series, but the DAR model has a simpler structure. One of the most significant differences between these models is that the second-order moment description of the GARCH model uses past conditional fluctuations, as well as past daily streamflow information, to estimate the current conditional variance. In contrast, the DAR model relies solely on past streamflow information. Previous period streamflow values and conditional fluctuations reflect the fluctuation aggregation and long-term behavior characteristics of the time series, respectively. Nevertheless, most of the literature (Guo et al., 2021b; Modarres and Ouarda, 2013b; Pandey et al., 2019) indicated that the classic GARCH model is insufficient in describing fluctuation aggregation effect, which significantly exists in hydrological time series (Figure S1). Moreover, limited data availability in practice means that short-term volatility clustering is often prioritized over long-term volatility. To avoid long-term persistent interference, the daily streamflow time series is processed with an effective fractional difference approach before being used as the model input. Taken together, these constraints result in better prediction performance and accuracy for the DAR-type models than for the AR-GARCH-type models.

5.3 Threshold regression models

Geosciences acknowledge that data are the only reliable source of information for model construction, derivation, and prediction. The concepts of stationarity and non-stationarity are considered modeling options rather than properties of data because stochastic models are inherently mathematical constructs (Dimitriadis and Koutsoyiannis, 2018; Koutsoyiannis and Montanari, 2015). However, model properties such as randomness, determinism, stationarity, and non-stationarity should always align with the data. Furthermore, stationarity is ergodic in practice, i.e., the selected sample data is representative of the entire stochastic process (Koutsoyiannis and Montanari, 2015). Meanwhile, the short-term behavior of hydrological predictions can allow model structure to be inferred from sample data. These provide evidence for the rationality of using data to build models and predict in this study. In regression analysis, the stability of coefficient estimates is usually studied, but the external force disturbance that causes structural breaks in the streamflow time series cannot be avoided, which hinders the applicability of ordinary linear regression models.

Affected by multiple factors such as global climate change and intense human activities, the nonlinearity of the basin hydrological system has been enhanced. As early as the 1990s, Xia et al. (1997) had conducted in-depth and systematic research on nonlinear hydrological systems, and explored the nonlinear process of streamflow generation and transformation, as well as the nonlinear mechanism of streamflow formation and transformation from response units to watershed scales. He further proposed a new nonlinear model for streamflow simulation and prediction. In practice, the hydrological gain of streamflow magnitude generated by precipitation is closely related to highly nonlinear controlling factors (such as precipitation, underlying surface conditions, soil moisture, etc.), resulting in the system gain of streamflow generated by precipitation being non-steady,

which is a nonlinear relationship. Therefore, linear mechanism process analysis is difficult to meet the needs of basin streamflow simulation and prediction.

In view of the nonlinear changes in water resource systems (Wu et al. 2023), a series of multivariate analysis methods based on nonlinear relationships have been developed, where numerous conditioning factors are input into models to simulate streamflow (Song et al. 2022). These methods improve the simulation and prediction accuracy by comprehensively considering the influence of climate factors, soil properties, terrain conditions and vegetation types on streamflow (Tiwari et al. 2022). Regarding the land surface water balance, precipitation reaches the ground after being intercepted by vegetation leaves, and infiltrates through the soil or becomes surface streamflow. The water intercepted by the leaves and a portion of the water in the soil returns to the atmosphere through evaporation, and the soil moisture absorbed by vegetation for growth is transpired through the leaves, ultimately forming the cycle of atmospheric-land surface water. At present, coupled process and data-driven multivariate models are developing rapidly, which can take into account the mechanisms of streamflow and further enhance the simulation and prediction accuracy of single-type models (Zhong et al. 2023).

The univariate method constructed in this study is based on various performance characteristics of streamflow time series, reducing the uncertainty of input factors by generalizing the rainfall-runoff process in the basin, making practical applications more convenient. The threshold concept is used to explain the nonlinear variations in the series, which can more accurately describe the linear variation process of daily streamflow time series within different threshold intervals, and has strong robustness and applicability. Moreover, the threshold model is effective in capturing the asymmetric effects of rising and falling changes in daily streamflow time series, offering higher flexibility in grasping variation information compared to linear structure models, with high potential for simulation and prediction.

5.4 Limitations of FDTDAR models

The DAR-type models offer a simpler and more interpretable approach to time series modeling. They are particularly advantageous in scenarios where domain interpretability, limited sample sizes, and computational efficiency are critical. Their transparent structure facilitates parameter estimation, theoretical analysis, and diagnostic evaluation, making them especially suitable for hydrological applications that typically involve noisy data, sparse observations, and operational constraints, thereby enhancing their robustness and reliability in real-world application (Li et al., 2019; Ling, 2007).

However, the proposed FDTDAR model also has inherent limitations. First, although it can effectively capture regime-switching behavior and certain nonlinear dynamics, it may fall short in modeling the complex and high-dimensional variations commonly present in large-scale hydrological or climate datasets. Second, the modeling capacity of the FDTDAR framework is constrained by the number of thresholds and lag terms that can be practically specified, which limits its structural flexibility compared with more adaptive, data-driven models such as LSTM. Third, the prediction performance of FDTDAR models often depends heavily on the correct setting of thresholds and lag structures. This adjustment process can be challenging in practice and often requires expert domain knowledge or a large number of traversal trials, which may reduce the generalization ability of the model across different regions or datasets.

In contrast, machine learning models, particularly deep learning approaches, can learn complex patterns, nonlinearities, and long-term dependencies directly from the data without strong prior assumptions about model structure (van Cranenburgh et al., 2022). Figure 10 shows that the prediction accuracy of the FDTDAR-t model is comparable to, or even slightly higher than, that of the LSTM model. These models often demonstrate outstanding predictive performance in many benchmark tasks. However, their "black-box" nature raises concerns about interpretability and transparency, which are critical in scientific and decision-making contexts such as hydrology (Beven, 2020; Hosseini et al., 2025). In addition, the high computational cost, demand for large training datasets, and sensitivity to hyperparameter tuning may restrict their applicability in resource-limited environments or real-time forecasting systems.

Given these trade-offs, this study focuses on DAR-type models as a theoretically grounded and operationally tractable alternative to traditional AR-GARCH models. While the proposed FDTDAR framework strikes a balance between model simplicity and nonlinear expressiveness, we acknowledge that future research should include comprehensive comparisons with state-of-the-art machine learning models. Such efforts would help further assess the strengths and limitations of the proposed approach and explore the potential of hybrid frameworks that combine the interpretability of statistical models with the flexibility of machine learning techniques.

In terms of the model residuals, although the Student's t distribution used in this study effectively captures heavy tails, it does not account for potential asymmetry in the distribution of hydrological residuals. In real-world streamflow processes, especially during flood or drought events, residuals may exhibit skewness in addition to heavy tails. Therefore, more flexible distributions, such as the skewed or generalized t-distributions, may offer a better fit for such cases. While these distributions were not explored in the present study, they represent a promising direction for future research aimed at enhancing the model's adaptability to extreme or asymmetric behavior.

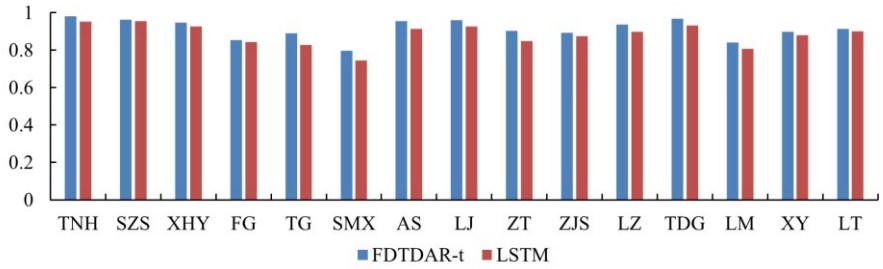

Figure 10: Comparison of prediction accuracy (NSE) during the validation period between LSTM and FDTDAR-t models

6 Conclusion

The nonlinear changes of the daily streamflow time series driven by the external environment cannot be ignored, and the strong non-stationarity and volatility, as well as its long-term memory, together lead to an increase in the difficulty of streamflow prediction, which brings great challenges to the traditional time series models. To address this issue, this study introduced a

novel DAR model and further proposed the FDTDAR model, which considered these critical features and added a threshold for improving the accuracy of daily streamflow time series prediction. The main conclusions of this study can be summarized as follows:

(1) The DAR model is a better alternative to the AR-GARCH model for predicting daily streamflow time series. Both the linear structure DAR and the threshold FDTDAR models outperform the AR-GARCH, TAR-GARCH, and LSTM models in terms of predictive ability.

(2) The nonlinear changes of the daily streamflow time series are reflected in multiple linear structures by adding the threshold, improving the accuracy of the single linear structure method. The NSE values of the FDTDAR and TAR-GARCH models are higher than those of the DAR and AR-GARCH models by 0.013-0.556 and 0.031-0.582, respectively.

(3) The hypothesis type based on the t-distribution for residuals stands out in the prediction of daily streamflow time series for both FDTDAR and FTAR-GARCH models and is a better choice than the traditional normal distribution.

Code and data availability. The seasonally normalized and fractionally differencing daily streamflow data in this study can be downloaded from https://doi.org/10.6084/m9.figshare.26795140.v1. The codes for calculating results and plotting figures is available upon request.

Author contributions. HW: conceptualization, methodology, software, formal analysis, data curation, writing – original draft, writing – review and editing, and visualization; SS: conceptualization, methodology, investigation, data curation, writing – review and editing, supervision, project administration, and funding acquisition; GZ: conceptualization, methodology, investigation, data curation, writing – review and editing, supervision, project administration, and funding acquisition; TG: Formal analysis and writing – review and editing; ZP: Formal analysis and writing – review and editing.

Competing interests. The author has declared that there are no competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Financial support. This study was supported financially by the National Natural Science Foundation of China (Grant Nos. 52509036 and 52079110).

References

Adnan, R.M., Meshram, S.G., Mostafa, R.R., Towfiqul IsIam, A.R.M., Abba, S.I., Andorful, F., and Chen, Z.H.: Application of soft computing models in streamflow forecasting. Proceedings of the Institution of Civil Engineers-water Management, 172(3): 123-134. https://doi.org/10.1680/jwama.16.00075, 2019.

Beven, K., 2020. Deep learning, hydrological processes and the uniqueness of place. Hydrological Processes, 34(16).

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3): 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

Broock, W.A., Scheinkman, J.A., Dechert, W.D., LeBaron, B., 1996. A test for independence based on the correlation dimension. Econometric Reviews, 15(3): 197-235. https://doi.org/10.1080/07474939608800353

Can, İ., Tosunoğlu, F., and Kahya, E.: Daily streamflow modelling using autoregressive moving average and artificial neural networks models: case study of Çoruh basin, Turkey. Hydrological Processes, 26(4): 567-576. https://doi.org/10.1111/j.1747-6593.2012.00337.x, 2012.

Chen, S., Xu, J.J., Li, Q.Q., Wang, Y.Q., Yuan, Z., and Wang, D.: Nonstationary stochastic simulation method for the risk assessment of water allocation. Environmental Science-water Research & Technology, 7(1): 212-221. https://doi.org/10.1039/d0ew00695e, 2021.

Daren, B.H.C., Huay-min, H.P., 2004. Stability and the Lyapounov Exponent of Threshold AR-ARCH Models. The Annals of Applied Probability, 14(4): 1920-1949.

Dawson, C.W., Abrahart, R.J., See, L.M., 2007. HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environmental Modelling & Software, 22(7): 1034-1052. https://doi.org/10.1016/j.envsoft.2006.06.008

Delforge, D., de Viron, O., Vanclooster, M., Van Camp, M., Watlet, A., 2022. Detecting hydrological connectivity using causal inference from time series: synthetic and real karstic case studies. Hydrol. Earth Syst. Sci., 26(8): 2181-2199. https://doi.org/10.5194/hess-26-2181-2022

Dickey, D.A., Fuller, W.A., 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. Econometrica, 49(4): 1057-1072. https://doi.org/10.2307/1912517

Dimitriadis, P., Koutsoyiannis, D., 2018. Stochastic synthesis approximating any process dependence and distribution. Stochastic Environmental Research and Risk Assessment, 32(6): 1493-1515. https://doi.org/10.1007/s00477-018-1540-2

Dimitriadis, P., Koutsoyiannis, D., 2015. Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes. Stochastic Environmental Research and Risk Assessment, 29(6): 1649-1669. https://doi.org/10.1007/s00477-015-1023-7

Dimitriadis, P., Koutsoyiannis, D., Iliopoulou, T., Papanicolaou, P., 2021. A Global-Scale Investigation of Stochastic Similarities in Marginal Distribution and Dependence Structure of Key Hydrological-Cycle Processes. Hydrology, 8(2): 59.

Engle R., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4): 987-1007.

Engle R.F., 1990. Stock Volatility and the Crash of '87: Discussion. The Review of Financial Studies, 3(1): 103-106.

Engle, R.F., Lilien, D.M., Robins, R.P., 1987. Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model. Econometrica, 55(2): 391-407.

Engle, R F, Ng, V K., 1993. Measuring and Testing the Impact of News on Volatility. The Journal of Finance, 48(5): 1749-1778.

Fathian, F., Fakheri-Fard, A., Modarres, R., van Gelder, P.H.A.J.M., 2018. Regional scale rainfall–runoff modeling using VARX–MGARCH approach. Stochastic Environmental Research and Risk Assessment, 32(4): 999-1016. https://doi.org/10.1007/s00477-017-1428-6

Fathian, F., Mehdizadeh, S., Kozekalani Sales, A., Safari, M.J.S., 2019. Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. Journal of Hydrology, 575: 1200-1213. https://doi.org/10.1016/j.jhydrol.2019.06.025

Fathian, F., Vaheddoost, B., 2021. Modeling the volatility changes in Lake Urmia water level time series. Theoretical and Applied Climatology, 143(1): 61-72. https://doi.org/10.1007/s00704-020-03417-8

Feng, Z.K., Niu, W.J., Wan, X.Y., Xu, B., Zhu, F.L., Chen, J., 2022. Hydrological time series forecasting via signal decomposition and twin support vector machine using cooperation search algorithm for parameter identification. Journal of Hydrology, 612: 128213. https://doi.org/10.1016/j.jhydrol.2022.128213

Gharehbaghi, V.R., Kalbkhani, H., Farsangi, E.N., Yang, T.Y., Mirjalili, S., 2022. A data-driven approach for linear and nonlinear damage detection using variational mode decomposition and GARCH model. Engineering with Computers. https://doi.org/10.1007/s00366-021-01568-4

Granger, C.W.J., 1978. New Classes of Time Series Models. Journal of the Royal Statistical Society. Series D (The Statistician), 27(3/4): 237-253. https://doi.org/10.2307/2988186

Graves, T., Gramacy, R., Watkins, N., Franzke, C., 2017. A Brief History of Long Memory: Hurst, Mandelbrot and the Road to ARFIMA, 1951–1980. Entropy, 19(9). https://doi.org/10.3390/e19090437

Grimaldi, S., 2004. Linear Parametric Models Applied to Daily Hydrological Series. Journal of Hydrologic Engineering, 9(5): 383-391. https://doi.org/10.1061/(ASCE)1084-0699(2004)9:5(383)

Guo, S., Xiong, L., Zha, X., Zeng, L., Cheng, L., 2021a. Impacts of the Three Gorges Dam on the streamflow fluctuations in the downstream region. Journal of Hydrology, 598: 126480. https://doi.org/10.1016/j.jhydrol.2021.126480

Guo, T.L., Song, S.B., Ma, W.J., 2021b. Point and Interval Forecasting of Groundwater Depth Using Nonlinear Models. Water Resouces Research, 57(12). https://doi.org/10.1029/2021WR030209

Hansen, A.L., 2021. Modeling persistent interest rates with double-autoregressive processes. Journal of Banking & Finance, 133. https://doi.org/10.1016/j.jbankfin.2021.106302

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Computation*, 9(8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hosking, J.R.M., 1981. Fractional Differencing. Biometrika, 68(1): 165-176. https://doi.org/10.2307/2335817

Hosseini, F., Prieto, C., Álvarez, C., 2025. An explainable AI approach for interpreting regionally optimized deep neural networks in hydrological prediction. Journal of Hydrology, 661: 133689. https://doi.org/10.1016/j.jhydrol.2025.133689

Huang, Y.M., Dai, X.Y., Wang, Q.W., Zhou, D.Q., 2021. A hybrid model for carbon price forecasting using GARCH and long short-term memory network. Applied Energy, 285. https://doi.org/10.1016/j.apenergy.2021.116485

Hurst, H.E., 1951. Long-Term Storage Capacity of Reservoirs. Transactions of the American Society of Civil Engineers, 116(1): 770-799. https://doi.org/10.1061/TACEAT.0006518

Jiang, F.Y., Li, D., Zhu, K., 2020. Non-standard inference for augmented double autoregressive models with null volatility coefficients. Journal of Econometrics, 215(1): 165-183. https://doi.org/10.1016/j.jeconom.2019.08.009

Kolte, A., Roy, J.K., Vasa, L., 2023. The impact of unpredictable resource prices and equity volatility in advanced and emerging economies: An econometric and machine learning approach. Resources Policy, 80. https://doi.org/10.1016/j.resourpol.2022.103216

Koutsoyiannis, D., Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. Hydrological Sciences Journal, 60(7-8): 1174-1183. https://doi.org/10.1080/02626667.2014.959959

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 Sciences, 22(11): 6005-6022. https://doi.org/10.5194/hess-22-6005-2018

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics, 54(1): 159-178. https://doi.org/10.1016/0304-4076(92)90104-Y

Li, D., Guo, S., Zhu, K., 2019. Double AR model without intercept: An alternative to modeling nonstationarity and heteroscedasticity. Econometric Reviews, 38(3): 319-331. https://doi.org/10.1080/07474938.2017.1310080

Li, D., Ling, S., Zhang, R., 2016. On a Threshold Double Autoregressive Model. Journal of Business & Economic Statistics, 34(1): 68-80.

Ling, S., 2007. A Double AR(p) Model: Structure and Estimation. Statistica Sinica, 17: 161-175.

Liu, C., Cui, N., Gong, D., Hu, X., Feng, Y., 2020. Evaluation of seasonal evapotranspiration of winter wheat in humid region of East China using large-weighted lysimeter and three models. Journal of Hydrology, 590: 125388. https://doi.org/10.1016/j.jhydrol.2020.125388

Liu, F., Li, D., Kang, X.M., 2018. Sample path properties of an explosive double autoregressive model. Econometric Reviews, 37(5): 484-490. https://doi.org/10.1080/07474938.2015.1092841

Ljung, G.M., Box, G.E.P., 1978. On a Measure of Lack of Fit in Time Series Models. Biometrika, 65(2): 297-303. https://doi.org/10.2307/2335207

Lo, A. W., 1991. Long-term memory in stock market prices. Econometrica, 59(5): 1279-1313. https://doi.org/10.2307/2938368

Lyu, Y., Chen, H., Cheng, Z., He, Y., Zheng, X., 2023. Identifying the impacts of land use landscape pattern and climate changes on streamflow from past to future. Journal of Environmental Management, 345: 118910. https://doi.org/10.1016/j.jenvman.2023.118910

Ma, X., Li, Z., Ren, Z., Shen, Z., Xu, G., Xie, M., 2024. Predicting future impacts of climate and land use change on streamflow in the middle reaches of China's Yellow River. Journal of Environmental Management, 370: 123000. https://doi.org/10.1016/j.jenvman.2024.123000

Mandelbrot, B.B., Wallis, J.R., 1969. Computer Experiments With Fractional Gaussian Noises: Part 1, Averages and Variances. Water Resources Research, 5(1): 228-241. https://doi.org/10.1029/WR005i001p00228

Matic, P., Bego, O., Males, M., 2022. Complex Hydrological System Inflow Prediction using Artificial Neural Network. Tehnicki Vjesnik-Technical Gazette, 29(1): 172-177. https://doi.org/10.17559/TV-20200721133924

Mehdizadeh, S., Fathian, F., Adamowski, J.F., 2019. Hybrid artificial intelligence-time series models for monthly streamflow modeling. Applied Soft Computing, 80: 873-887. https://doi.org/10.1016/j.asoc.2019.03.046

Miller, R.L., 2022. Nonstationary streamflow effects on backwater flood management of the Atchafalaya Basin, USA. Journal of Environmental Management, 309: 114726. https://doi.org/10.1016/j.jenvman.2022.114726

Modarres, R., 2007. Streamflow drought time series forecasting. Stochastic Environmental Research and Risk Assessment, 21(3): 223-233. https://doi.org/10.1007/s00477-006-0058-1

Modarres, R., Ouarda, T., 2013a. Modeling rainfall–runoff relationship using multivariate GARCH model. Journal of Hydrology, 499: 1-18. https://doi.org/10.1016/j.jhydrol.2013.06.044

Modarres, R., Ouarda, T.B.M.J., 2013b. Modelling heteroscedasticty of streamflow times series. Hydrological Sciences Journal, 58(1): 54-64. https://doi.org/10.1080/02626667.2012.743662

Montanari, A., Rosso, R., Taqqu, M.S., 1996. Some long-run properties of rainfall records in Italy. Journal of Geophysical Research: Atmospheres, 101(D23): 29431-29438. https://doi.org/10.1029/96JD02512

Montanari, A., Rosso, R., Taqqu, M.S., 1997. Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. Water Resources Research, 33(5): 1035-1044. https://doi.org/10.1029/97WR00043

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Veith, T., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE, 50. https://doi.org/10.13031/2013.23153

Myronidis, D., Ioannou, K., Fotakis, D., Dörflinger, G., 2018. Streamflow and Hydrological Drought Trend Analysis and Forecasting in Cyprus. Water Resources Management, 32(5): 1759-1776. https://doi.org/10.1007/s11269-018-1902-z

Nazeri-Tahroudi, M., Ramezani, Y., De Michele, C., Mirabbasi, R., 2022. Bivariate Simulation of Potential Evapotranspiration Using Copula-GARCH Model. Water Resources Management, 36(3): 1007-1024. https://doi.org/10.1007/s11269-022-03065-9

O'Connell, P.E., Koutsoyiannis, D., Lins, H.F., Markonis, Y., Montanari, A., Cohn, T., 2016. The scientific legacy of Harold Edwin Hurst (1880-1978). Hydrological Sciences Journal-Journal DES Sciences Hydrologiques, 61(9): 1571-1590. https://doi.org/10.1080/02626667.2015.1125998

Pandey, P.K., Tripura, H., Pandey, V., 2019. Improving Prediction Accuracy of Rainfall Time Series By Hybrid SARIMA–GARCH Modeling. Natural Resources Research, 28(3): 1125-1138. https://doi.org/10.1007/s11053-018-9442-z

Peter, C.B.P., Perron, P., 1988. Testing for a Unit Root in Time Series Regression. Biometrika, 75(2): 335-346. https://doi.org/10.2307/2336182

Serinaldi, F., Kilsby, C.G., 2016. Irreversibility and complex network behavior of stream flow fluctuations. Physica A: Statistical Mechanics and its Applications, 450: 585-600. https://doi.org/10.1016/j.physa.2016.01.043

Sivakumar, B., 2009. Nonlinear dynamics and chaos in hydrologic systems: latest developments and a look forward. Stochastic Environmental Research and Risk Assessment, 23(7): 1027-1036. https://doi.org/10.1007/s00477-008-0265-z

Song, Z.H., Xia, J., Wang, G.S., She, D.X., Hu, C., Hong, S., 2022. Regionalization of hydrological model parameters using gradient boosting machine. Hydrology and Earth System Sciences, 26(2): 505-524. https://doi.org/10.5194/hess-26-505-2022

Tiwari, A.D., Mukhopadhyay, P., Mishra, V., 2022. Influence of bias correction of meteorological and streamflow forecast on hydrological prediction in India. Journal of Hydrometeorology, 23(7): 1171-1192. https://doi.org/10.1175/JHM-D-20-0235.1

Tong, H., 1983. Threshold models in non-linear time series analysis. Springer. https://doi.org/10.1007/978-1-4684-7888-4

van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., Walker, J., 2022. Choice modelling in the age of machine learning - Discussion paper. Journal of Choice Modelling, 42: 100340. https://doi.org/10.1016/j.jocm.2021.100340

Vavoulogiannis, S., Iliopoulou, T., Dimitriadis, P., Koutsoyiannis, D., 2021. Multiscale Temporal Irreversibility of Streamflow and Its Stochastic Modelling. Hydrology, 8(2): 63

Wang, H., Song, S., Zhang, G., Ayantobo, O.O., Guo, T., 2023a. Stochastic Volatility Modeling of Daily Streamflow Time Series. Water Resources Research, 59(1): e2021WR031662. https://doi.org/10.1029/2021WR031662

Wang, H., Song, S., Zhang, G., Ayantoboc, O.O., 2023b. Predicting daily streamflow with a novel multi-regime switching ARIMA-MS-GARCH model. Journal of Hydrology: Regional Studies, 47: 101374. https://doi.org/10.1016/j.ejrh.2023.101374

Wang, H., Song, S., Zhang, G., 2023c. Daily runoff simulation based on seasonal long-term memory DAR model. Journal of Hydraulic Engineering, 54(6): 686-695. https://doi.org/10.13243/j.cnki.slxb.20220982. (in Chinese with English Abstract)

Wang, L.L., Li, X., Ma, C.F., Bai, Y.L., 2019. Improving the prediction accuracy of monthly streamflow using a data-driven model based on a double-processing strategy. Journal of Hydrology, 573: 733-745. https://doi.org/10.1016/j.jhydrol.2019.03.101

Wu, Z.H., Liu, D.D., Mei, Y.D., Guo, S.L., Xiong, L.H., Liu, P., Chen, J., Yin, J.B., Zeng, Y.J., 2023. A nonlinear model for evaluating dynamic resilience of water supply hydropower generation-environment conservation nexus system. Water Resources Research, 59. https://doi.org/10.1029/2023WR034922

Xia, J., O'Connor, K.M., Kachroo, R.K., Liang, G.C., 1997. A non-linear perturbation model considering catchment wetness and its application in river flow forecasting. Journal of Hydrology, 200(1): 164-178. https://doi.org/10.1016/S0022-1694(97)00013-9

Yan, B.W., Mu, R., Guo, J., Liu, Y., Tang, J.L., Wang, H., 2022. Flood risk analysis of reservoirs based on full-series ARIMA model under climate change. Journal of Hydrology, 610. https://doi.org/10.1016/j.jhydrol.2022.127979

Yang, G.X., Bowling, L.C., 2014. Detection of changes in hydrologic system memory associated with urbanization in the Great Lakes region. Water Resources Research, 50(5): 3750-3763. https://doi.org/10.1002/2014WR015339

Yang, L.H., Zhong, P.A., Zhu, F.L., Ma, Y.F., Wang, H., Li, J.Y., Xu, C.J., 2022. A comparison of the reproducibility of regional precipitation properties simulated respectively by weather generators and stochastic simulation methods. Stochastic Environmental Research and Risk Assessment, 36(2): 495-509. https://doi.org/10.1007/s00477-021-02053-6

Yuan, R.F., Cai, S.Y., Liao, W.H., Lei, X.H., Zhang, Y.H., Yin, Z.K., Ding, G.B., Wang, J., Xu, Y., 2021. Daily runoff forecasting using ensemble empirical mode decomposition and long short-term memory. Frontiers in Earth Science, 9. https://doi.org/10.3389/feart.2021.621780

Zha, X., Xiong, L., Guo, S., Kim, J.S., Liu, D., 2020. AR-GARCH with Exogenous Variables as a Postprocessing Model for Improving Streamflow Forecasts. Journal of Hydrologic Engineering, 25: 04020036.

Zhong, M., Zhang, H.R., Jiang, T., Guo, J., Zhu, J.X., Wang, D.G., Chen, X.H., 2023. A hybrid model combining the Cama-flood model and deep learning methods for streamflow Prediction. Water Resources Management, 37(12): 4841-4859. https://doi.org/10.1007/s11269-023-03583-0