water

MDPI

Article

Daily Streamflow Prediction Using Multi-State Transition SB-ARIMA-MS-GARCH Model

Jin Zhao [1,2], Jianhui Shang [1,2], Qun Ye [3], Huimin Wang [4], Gengxi Zhang [4], Feng Yao [1] and Weiwei Shou [1,2,*]

[1] Power China Huadong Engineering Corporation Limited, Hangzhou 311122, China; zhao_j15@hdec.com (J.Z.); shang_jh@hdec.com (J.S.); yao_f@hdec.com (F.Y.)
[2] Huachen Electric Power Control Corporation Limited, Hangzhou 310053, China
[3] Zhejiang XPro Info Technology Limited, Hangzhou 311122, China; yeq@xproinfo.com
[4] College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou 225009, China; wanghuimin@yzu.edu.cn (H.W.); gengxizhang@yzu.edu.cn (G.Z.)
* Correspondence: shou_ww1@hdec.com; Tel.: +86-15088765052

Abstract

Under the combined influences of climate change and anthropogenic activities, the variability of basin streamflow has intensified, posing substantial challenges for accurate prediction. Although Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models characterize volatility in time series, many previous studies have neglected changes in series structure, leading to inaccurate identification of the form of volatility. Building on tests for structural breaks (SBs) in time series, this study first removes the series mean using an Autoregressive Integrated Moving Average (ARIMA) model and then incorporates Markov-switching (MS) to develop a multi-state MS-GARCH model. An asymmetric MS-GARCH (MS-gjrGARCH) variant is also incorporated to describe the volatility of streamflow series with SBs. Daily streamflow data from five hydrological stations in the middle reaches of the Yellow River are used to compare the predictive performance of SB-ARIMA-MS-GARCH, SB-ARIMA-MS-gjrGARCH, ARIMA-GARCH, and ARIMA-gjrGARCH models. The results show that daily streamflow exhibits SBs, with the number and timing of breakpoints varying among stations. Standard GARCH and gjrGARCH models have limited ability to capture runoff volatility clustering, whereas MS-GARCH and MS-gjrGARCH effectively characterize volatility features within individual states. The multi-state switching structure substantially improves daily streamflow prediction accuracy compared with single-state volatility models, increasing R^2 by approximately 5.8% and NSE by approximately 36.3%.The proposed modeling framework offers a robust new tool for streamflow prediction in such changing environments, providing more reliable evidence for water resource management and flood risk mitigation in the Yellow River basin.

Keywords: Yellow River Basin; daily streamflow; structural breaks; multi-state transition; time-series analysis

Check for updates

Academic Editors: Md Jahangir Alam, Monzur Imteaz and Faisal Anwar

Received: 21 November 2025
Revised: 6 January 2026
Accepted: 9 January 2026
Published: 16 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Streamflow is a fundamental component of the basin water cycle [1]. Accurate streamflow prediction is essential for flood control, drought mitigation, and scientific water-resource management [2,3], providing theoretical support for hydraulic engineering. Under the dual influences of climate change and intensified anthropogenic activities, streamflow series often display pronounced non-stationarity and volatility [4], posing major challenges to current prediction approaches.

Streamflow prediction models generally adopt process-driven or data-driven approaches [5]. Process-driven models must consider numerous factors affecting hydrological processes [6], and they commonly suffer from parameter uncertainty and limited generalizability. In contrast, data-driven models require only the mathematical characterization of time series, are straightforward to implement, and offer strong advantages in forecasting and time-series analysis [7–11], leading to widespread adoption in hydrology. Time-series analysis (TSA) and the advancement of time-series methodologies have remained central research themes in hydrological modeling [12–15]. Traditional models such as the Autoregressive Integrated Moving Average (ARIMA) model primarily describe the conditional mean of a series (i.e., first-moment behavior). However, hydrological time series are strongly influenced by changing environmental conditions and often exhibit pronounced volatility, i.e., conditional heteroscedasticity (ARCH effects). ARIMA alone cannot capture this volatility, which reduces predictive accuracy [13]. To address ARCH effects, researchers introduced the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model [16,17], which corrects ARIMA residuals and improves prediction performance [18,19].

Despite these advances, most existing studies implicitly assume constant model parameters across the full modeling period, disregarding structural changes in the data and consequently misidentifying ARCH/GARCH effects [20,21]. Structural breaks (SBs), which are unexpected changes in regression parameters over time, undermine model robustness and increase prediction errors. Time-series models that explicitly account for structural changes can therefore improve forecasting performance [22]. Multi-state switching approaches have been widely applied across disciplines, and among them, Markov-switching (MS) models are commonly used in hydrological time-series research. MS models estimate the transition probabilities between states and the smoothed probability of each state, thereby improving streamflow prediction [23–25]. MS-GARCH models, which integrate MS with GARCH, capture regime-dependent volatility through a discrete latent variable and rapidly adapt to changes in unconditional volatility levels, outperforming single-state GARCH models [26,27]. Hydrological studies have rarely considered SBs explicitly (e.g., threshold autoregression), yet under a changing climate, multi-state switching models that detect SBs at each station and characterize volatility within each state show strong potential for improving runoff prediction accuracy.

In this study, to accurately model volatility in different streamflow states, we integrate MS with GARCH to construct MS-GARCH-type streamflow prediction models that incorporate SBs and state transitions. After identifying SBs in daily streamflow series and determining the number of breakpoints n, we remove the mean component and assume the volatility structure differs on either side of these breakpoints. We then construct an MS-GARCH model with n 1 states to forecast daily streamflow. Both symmetric GARCH and asymmetric gjrGARCH formulations are considered, resulting in an MS-gjrGARCH model. These two multi-state models are compared with widely used single-state GARCH and gjrGARCH models based on daily streamflow from five hydrological stations in the middle reaches of the Yellow River. This work addresses critical gaps in hydrological time-series modeling and offers a new approach for streamflow prediction under climate change.

2. Methods

2.1. Single-State Models

2.1.1. ARIMA-GARCH Model

The ARIMA and GARCH models were used to represent the conditional mean and conditional variance of daily runoff series, respectively. Their general forms are

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \sum_{i-1}^{V} \varphi_i \varepsilon_{t-i}^2 + \sum_{j=1}^{M} \theta_j \sigma_{t-j}^2 \end{cases} \quad (2)$$

where y_t is the stationary streamflow series after first differencing, c and ω are constants, α_i and β_j are the ARIMA coefficients, ε_t is the ARIMA residual, η_t is white noise, φ_i and θ_j are the GARCH coefficients, and σ_t^2 is the conditional variance of ε_t. When $\theta > \varphi$, fluctuations in daily streamflow exhibit strong persistence, meaning that current-period volatility is strongly influenced by past volatility.

2.1.2. ARIMA–gjrGARCH Model

To construct the ARIMA–gjrGARCH model, the conditional variance equation in the GARCH model formulation is replaced by the asymmetric gjrGARCH variance equation. Equation (2) becomes

$$\sigma_t^2 = \omega + \sum_{i-1}^{V} \varphi_i \varepsilon_{t-i}^2 + \sum_{j=1}^{M} \theta_j \sigma_{t-j}^2 + \gamma \varepsilon_{t-1}^2 d_{t-1} \quad (3)$$

$$d_t = \begin{cases} 1, \varepsilon_t < 0 \\ 0, other \end{cases} \quad (4)$$

where d_t is a dummy indicator distinguishing positive from negative shocks in the variance equation. γ is the asymmetry parameter; $\gamma > 0$ indicates that volatility is asymmetrically influenced mainly by negative shocks, while $\gamma = 0$, reduces the model to the symmetric GARCH formulation. In streamflow time series, positive and negative shocks correspond to model underestimation and overestimation, respectively. The gjrGARCH model also relaxes the non-negativity constraints imposed on parameters in the symmetric GARCH formulation. Parameters of both ARIMA–GARCH and ARIMA–gjrGARCH models are estimated using maximum likelihood.

2.2. Multi-State Models

After identifying SBs in the streamflow series, a multi-state Markov-switching GARCH (MS-GARCH) model is applied to the ARIMA residuals from Equation (1), yielding the SB-ARIMA-MS-GARCH model. The MS framework allows analysis of time series with SBs or multiple regimes by assuming an unobserved state k at time t, driven by a hidden Markov chain. The model estimates state-dependent location (μ_k), scale (σ_k), and smoothed state probabilities ($\xi_{k,t}$). The general form of a two-state MS model is

$$\varepsilon_t \sim \Phi(\mu_k, \sigma_k) = p_{1,1} \cdot \xi_{k=1,t} + p_{2,2} \cdot \xi_{k=2,t} \quad (5)$$

where $p_{1,1}$ and $p_{2,2}$ are the fixed transition probabilities for each state k. As a hidden Markov chain, the probability transition matrix $\Pi(k \times k)$ for transitions from a given state (k = 1) to the next state (k = 2) can be expressed as

$$\Pi = \begin{bmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{bmatrix}, p_{i,j} = P(k_{t+1} = j | k_t = i, \mu_k, \sigma_k, \varepsilon_t) \quad (6)$$

To estimate the state transition probability matrix Π during the filtering process, a Gaussian kernel is used to filter the stationary probabilities of the sequence ($\xi_{k,t}$) generated under each state k:

$$p_{k,t} = \frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-1}{2}\left(\frac{\varepsilon_t}{\sigma_t}\right)^2} \quad (7)$$

Thus, the filtering probabilities, location (μ_k) and scale (σ_k) parameters, and the transition probability matrix (Π) are obtained. To prevent abrupt changes in the state probability

at time t, the filtered probability for state k is smoothed as a supplementary estimate. This yields the smoothed state probabilities $\xi_{k,t}$ for each state and the series, which can be arranged as matrix $P = [\xi_{1,t}, \xi_{2,t}]$. Finally, the state transition probability matrix at time t (Π) and the smoothed state probability matrix (P) are used to predict the probability of each state at time $t + 1$:

$$\begin{bmatrix} \xi_{1,t+1} \\ \xi_{2,t+1} \end{bmatrix} = \Pi \times \begin{bmatrix} \xi_{1,t} \\ \xi_{2,t} \end{bmatrix} \tag{8}$$

Therefore, the general MS-GARCH model takes the form:

$$\varepsilon_t | (s_t = k, \mathcal{J}_{t-1}) \sim \mathcal{D}(0, h_{k,t}) \tag{9}$$

where $\mathcal{J}_{t-1}$ denotes the information set of observations up to time $t - 1$, and $\mathcal{D}(0, h_{k,t})$ is a zero-mean distribution with state-dependent variance $h_{k,t}$. The latent variable s_t, defined on the discrete space $\{1, \ldots, K\}$, evolves according to the state transition probability matrix in Equation (6). The conditional variance of y_t follows Equation (2) (GARCH) or Equations (3) and (4) (gjrGARCH). Conditional on $s_t = k$, $h_{k,t}$ is formed by a state-dependent vector of functions of past observations and parameters θ_k. $h_{k,t}$ can be expressed as

$$h_{k,t} \equiv h(y_{t-1}, h_{k,t-1}, \theta_k) \tag{10}$$

To ensure the conditional variance function $h(\cdot)$ is positive, we assume $h_{k,1} \equiv \bar{h}_k (k = 1, \ldots, K)$, where $\bar{h}_k$ denotes the fixed initial variance for state k, i.e., the unconditional variance. Its general form is

$$h_{k,t} \equiv \omega_k + \varphi_k \varepsilon_{t-1}^2 + \theta_k h_{k,t-1} \tag{11}$$

where $\omega_k > 0$, $\varphi_k > 0$, $\theta_k > 0$, and $\varphi_k + \theta_k < 1$. $\Theta_k \equiv (\omega_k, \varphi_k, \theta_k)'$ can be estimated using maximum likelihood. A standard normal distribution is assumed for ARIMA residuals when constructing the models.

Prior to modeling, the Augmented Dickey–Fuller (ADF) test and Bai–Perron test [28] were employed to assess stationarity and SBs of the runoff series, respectively. The Ljung–Box (LB) test [29] and Lagrange multiplier (LM) test [30] were used to examine independence of ARIMA residuals and the presence of ARCH effects, respectively.

2.3. Evaluation Metrics

The evaluation metrics included Mean Absolute Error (MAE), Relative Error (RE), the coefficient of determination (R^2), and Nash–Sutcliffe Efficiency (NSE) with the Peak Difference Index (PDIFF) used to evaluate peak-flow prediction performance [31–33].

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{Q}_i - Q_i \right| \tag{12}$$

$$RE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Q_i - \hat{Q}_i}{Q_i} \right| \tag{13}$$

$$R^2 = \frac{\sum_{i=1}^{n} (Q_i - \overline{Q})(\hat{Q}_i - Q_i)}{\sqrt{\sum_{i=1}^{n} (Q_i - \overline{Q})^2 \sum_{i=1}^{n} (\hat{Q}_i - \overline{Q})^2}} \tag{14}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n} (Q_i - \overline{Q})^2} \tag{15}$$

$$PDIFF = max(Q_i) - max(\hat{Q}_i) \tag{16}$$

where Q_i and $\hat{Q}_i$ denote the observed and predicted streamflow values, respectively; $\overline{Q}$ and $\widetilde{Q}$ are the mean values of the observed and predicted daily streamflow series, respectively; and n is the length of the testing period.

3. Case Studies

3.1. Study Area and Data Analysis

The middle reaches of the Yellow River Basin constitute an important agricultural and pastoral production zone as well as an energy production base in China. This river segment is 1234.6 km long, accounting for approximately 45.7% of the total drainage area of the Yellow River Basin [34]. The segment includes the largest tributary in the basin, the Wei River, which is approximately 818 km long and has two primary tributaries: the Jing River and the Beilu River.

This study used daily streamflow series from five hydrological stations in the middle reaches of the Yellow River Basin (Fugu, Suide, Lintong, Xianyang, and Tongguan) to construct and evaluate models. The predictive performances of SB-ARIMA-MS-gjrGARCH, SB-ARIMA-MS-GARCH, ARIMA-gjrGARCH, and ARIMA-GARCH models were compared. Among these stations, Fugu and Tongguan are located on the main stream of the Yellow River, whereas Suide, Lintong, and Xianyang are tributary stations (Figure 1). The selection of these five stations in the middle reaches of the Yellow River Basin was based on the following three criteria: ① The middle reaches of the Yellow River Basin are the primary source of sediment for the Yellow River and a critical area for soil and water conservation. We selected two mainstem stations to control the integrated water-sediment signals of the upper and lower boundaries of the middle reaches. ② To capture the spatial heterogeneity of the Loess Plateau, three tributary stations were selected. These tributaries represent different sub-catchment characteristics, including varying degrees of vegetation restoration and human interventions. ③ All five stations are National Key Hydrological Stations managed by the Yellow River Conservancy Commission (YRCC). The datasets provide high-quality, long-term continuous records with high temporal resolution, ensuring the reliability of our analysis of the water-sediment relationship in this complex region.

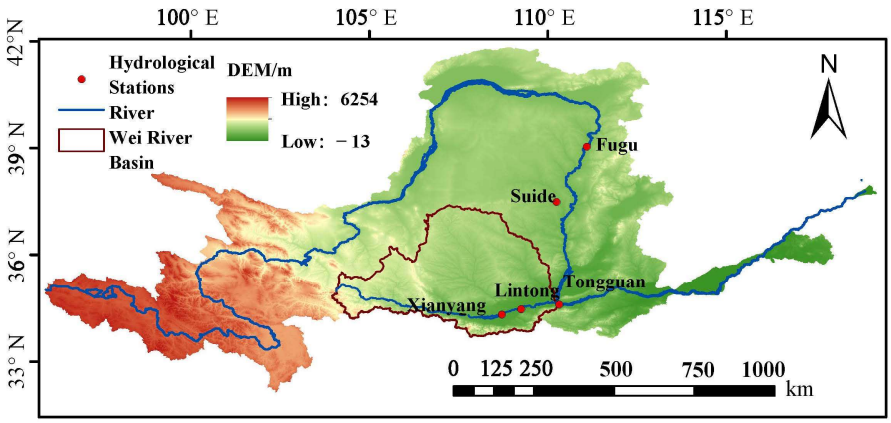


Figure 1. Location map of the study area and hydrological stations.

Table 1 summarizes the statistical characteristics of the daily streamflow series at each station. As shown, the mean, median, and standard deviation of streamflow at the Fugu and Tongguan stations were significantly higher than those at the three tributary stations, indicating higher average daily runoff and stronger variability at these two sites. By contrast, the Suide station exhibited relatively low mean daily runoff and smaller fluctuations. The skewness and kurtosis coefficients indicated that the probability density

functions of daily runoff at all stations had pronounced right tails and peak values higher than those of a normal distribution.

Table 1. Statistical characteristics of daily streamflow series.

Stations	Period	Mean (m^3/s)	Median (m^3/s)	Standard Deviation	Skewness	Watershed Area (km^2)
Fugu	1 January 2005–31 December 2015	592.624	485.0	443.515	1.962	404,000
Suide	1 January 1990–31 December 2003	4.260	2.2	16.922	19.394	3906
Lintong	1 January 1985–31 December 2007	159.717	89.6	239.144	6.285	97,299
Xianyang	1 January 1985–31 December 2013	86.546	44.9	151.37	7.344	46,827
Tongguan	1 January 2005–31 December 2015	793.917	650.0	560.504	2.705	682,141

The record lengths of the daily streamflow series differ among stations. In this study, the final year of each station's record was used as the prediction period, while the remaining data formed the training set for model development. Because prediction accuracy typically decreases as the forecast horizon lengthens, a rolling forecast with a one-day lead time was adopted, meaning the forecast horizon was set to 1 day [35].

3.2. Data Stationarity Test and SB Identification

Before model construction, it is necessary to examine the stationarity of the series and identify whether SBs exist. The training-period data for each station were analyzed using the ADF test and the Bai–Perron test, with results reported in Table 2. The maximum number of allowable breaks was set to 5. The original daily streamflow series at all stations failed the stationarity test; Table 2 therefore reports the results for the first-order differenced series, which satisfied the stationarity requirement for modeling.

Table 2. Stationarity and structural break tests of daily streamflow series.

Station	Augmented Dickey–Fuller Test		Structural Break Test	
	Statistical Value	*p* Value	Number of Breaks	Break Timings
Fugu	−18.901	0.01	2	31 May 2010; 25 September 2011
Suide	−27.842	0.01	2	22 July 1994; 11 August 1996
Lintong	−27.6687	0.01	4	20 April 1988; 8 August 1991; 25 November 1994; 15 July 2003
Xianyang	−27.5544	0.01	2	27 November 1993; 14 July 2003
Tongguan	−17.34	0.01	4	8 May 2006; 27 March 2008; 4 July 2010; 29 September 2011

The SB tests indicated that all five stations exhibited structural breakpoints in their daily streamflow series, implying structural inconsistencies during the training period. The number and timing of breaks, and hence the number of structural states, differed among stations.

3.3. Daily Streamflow Prediction Model Construction

3.3.1. ARIMA Model Validity Testing

An ARIMA model was first constructed in accordance with the modeling requirements. Regarding the ARIMA model selection, we utilized the auto.arima function from the forecast package in R. This algorithm automatically identifies the optimal model structure (p, d, q) by conducting a stepwise search and selecting the candidate that minimizes the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). This automated process ensures that the chosen model achieves the best balance between goodness-of-fit and model parsimony, avoiding the subjectivity of manual selection. The

selected model orders and residual test results are presented in Table 3. The *p*-values of the Ljung–Box (LB) test for the residuals at all five stations were less than 0.05, indicating that the null hypothesis of residual independence could be rejected and that the ARIMA models effectively extracted the conditional mean of the runoff series. However, the squared residuals exhibited significant autocorrelation, and the LM test rejected the null hypothesis that the residuals did not have ARCH effects, confirming the presence of conditional heteroscedasticity in the ARIMA residuals. Therefore, volatility models were required to capture and remove these ARCH effects.

Table 3. ARIMA model order and p-values of residual tests.

Stations	Orders	Residual LB Test	Square Residual LB Test	LM Test
Fugu	5, 1, 5	0.992	0	0
Suide	8, 1, 9	0.068	0	0
Lintong	5, 1, 6	0.068	0	0
Xianyang	8, 1, 8	0.280	0	0
Tongguan	5, 1, 5	0.662	0	0

Note: LB test = Ljung–Box test; LM test = Lagrange Multiplier test.

3.3.2. Volatility Regime Analysis

Table 4 reports the estimated parameters of the GARCH, gjrGARCH, MS-GARCH, and MS-gjrGARCH models for the five hydrological stations. For the single-state GARCH and gjrGARCH models, $\theta > \varphi$, indicating that daily streamflow volatility exhibited strong persistence and was predominantly influenced by past volatility. The asymmetry parameter γ showed that volatility at Xianyang was primarily influenced by negative shocks, whereas volatility at the other stations was mainly influenced by positive shocks.

Table 4. Parameter estimates of volatility models.

Stations	Models	ω	$t\varphi$	θ	γ
Fugu	GARCH	1971.130	0.153	0.826	
	gjrGARCH	1679.633	0.187	0.812	−0.141
	MS-GARCH	174.895	0.024	0.882	-
	MS-gjrGARCH	3654.320	0.102	0.002	0.0001
Lintong	GARCH	31.498	0.234	0.765	-
	gjrGARCH	42.534	0.268	0.732	−0.608
	MS-GARCH	33,636.221	0.670	0.000	-
	MS-gjrGARCH	1004.384	0.792	0.001	0.000
Suide	GARCH	0.366	0.134	0.865	-
	gjrGARCH	0.280	0.113	0.879	−0.989
	MS-GARCH	323.029	0.999	0.000	-
	MS-gjrGARCH	0.044	0.001	0.874	0.251
Xianyang	GARCH	12.063	0.124	0.783	-
	gjrGARCH	13.669	0.11	0.848	0.043
	MS-GARCH	0.187	0.000	1.000	-
	MS-gjrGARCH	0.315	0.001	0.999	0.000
Tongguan	GARCH	2026.588	0.107	0.866	-
	gjrGARCH	1877.925	0.144	0.890	−0.177
	MS-GARCH	1957.358	0.00	0.569	-
	MS-gjrGARCH	1870.612	0.000	0.580	0.001

The multi-state MS-GARCH and MS-gjrGARCH models revealed distinct volatility regimes within the streamflow series of each station. For Fugu and Xianyang the MS-GARCH model indicated strong persistence in volatility for all regimes. The MS-gjrGARCH

model further showed volatility clustering in regime 1 at Fugu and in the latter two regimes at Xianyang. At Suide, the MS-GARCH model identified strong volatility clustering in regime 3, while the MS-gjrGARCH model indicated strong persistence across all regimes. At Lintong, except the fact that the MS-gjrGARCH model indicated strong persistence in regimes 3 and 4, volatility in the other regimes was mainly driven by external shocks rather than persistent effects. At Tongguan, the MS-GARCH model indicated volatility clustering in regime 3, but with relatively low intensity; otherwise, volatility primarily exhibited strong persistence driven by past shocks. In the multi-state MS-gjrGARCH models, all γ estimates were non-negative, implying no significant negative asymmetry in streamflow volatility across states at any station.

3.3.3. Volatility State Transition Analysis

Figure 2 presents the state transition probabilities obtained from the multi-state MS-GARCH and MS-gjrGARCH models for the five stations. In each transition matrix, rows correspond to the current time t and columns to the next time $t + 1$. Large $p_{1,1}$, $p_{2,2}$, $p_{3,3}$ values in both the symmetric MS-GARCH and asymmetric MS-gjrGARCH models indicate that volatility states at Fugu and Xianyang tends to remain in the same state at the next time step. At Suide, the MS-GARCH model yields relatively small differences in transition probabilities among states, whereas the MS-gjrGARCH model shows that volatility in states 1 and 3 has a high probability of transitioning to state 3 at $t + 1$, while state 2 is more likely to transition to state 1. For Lintong and Tongguan, the MS-gjrGARCH model shows that volatility to remain in the current state during transitions, and the MS-GARCH model produces similar patterns, except for the fact that state 2 tends to transition to state 1 at Lintong and state 3 tends to transition to state 4 at Tongguan.

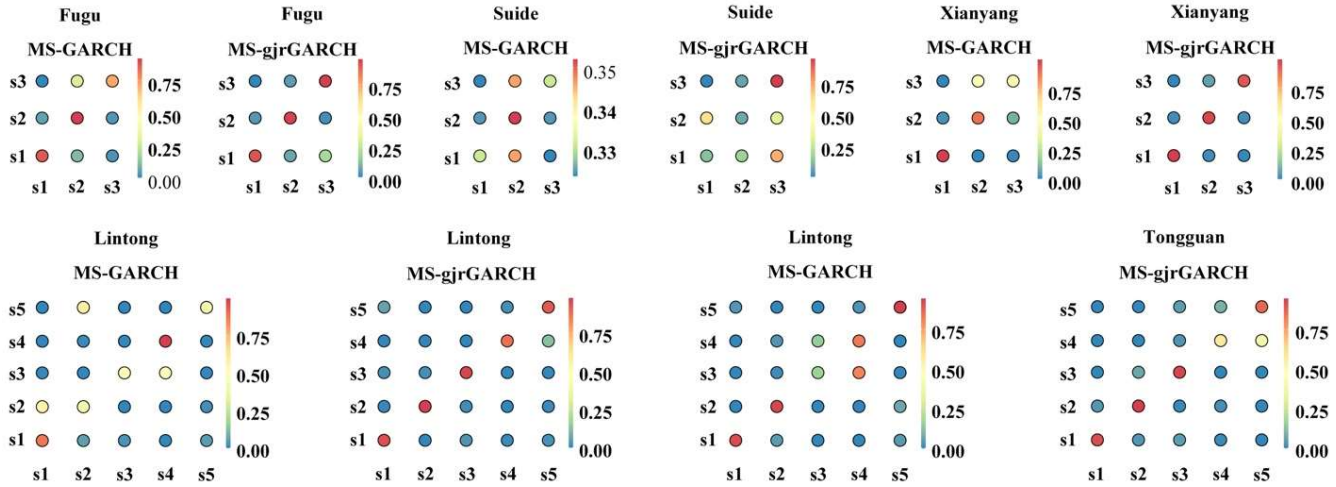


Figure 2. Regime-switching transition matrices of the Markov model.

3.4. Model Prediction Results

Figure 3 shows the daily streamflow predictions of the four models at the selected hydrological stations. Single-state volatility models tend to overestimate streamflow volatility, whereas multi-state switching models better reproduce the temporal patterns of streamflow variability and produce forecasts that align more closely with observations.

Figure 4 shows the prediction performance metrics for each station. For the single-state models, MAE and RE are similar for ARIMA-gjrGARCH and ARIMA-GARCH, indicating comparable average error magnitudes. However, ARIMA-gjrGARCH achieves higher R^2 and NSE values, implying that the asymmetric model provides a better overall fit than the symmetric ARIMA-GARCH. This advantage is not consistently observed for the two

multi-state models. Specifically, SB-ARIMA-MS-GARCH yields a notably higher RE at Suide than the asymmetric SB-ARIMA-MS-gjrGARCH, while their performance at the other stations is comparable.

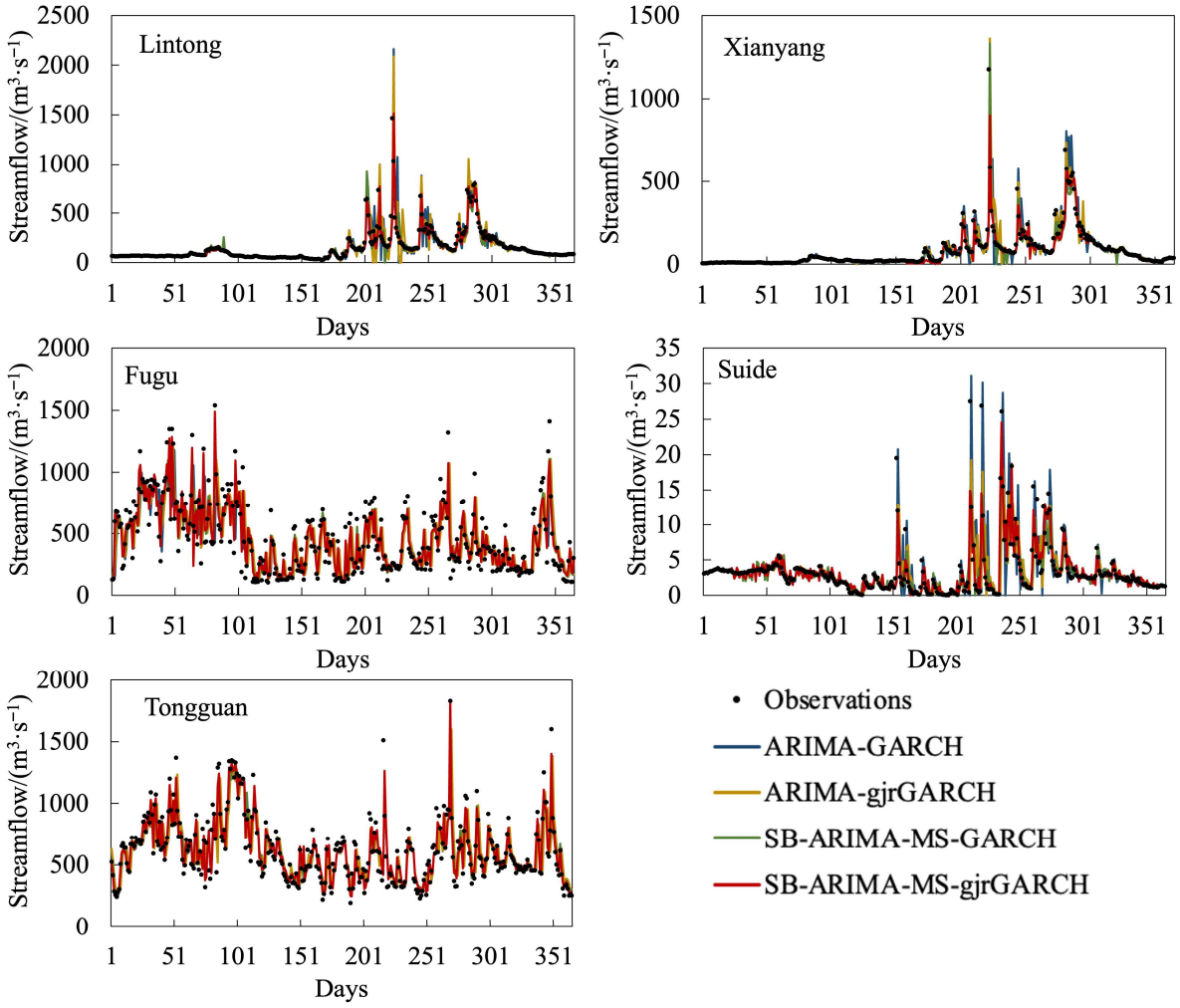


Figure 3. Model prediction results.

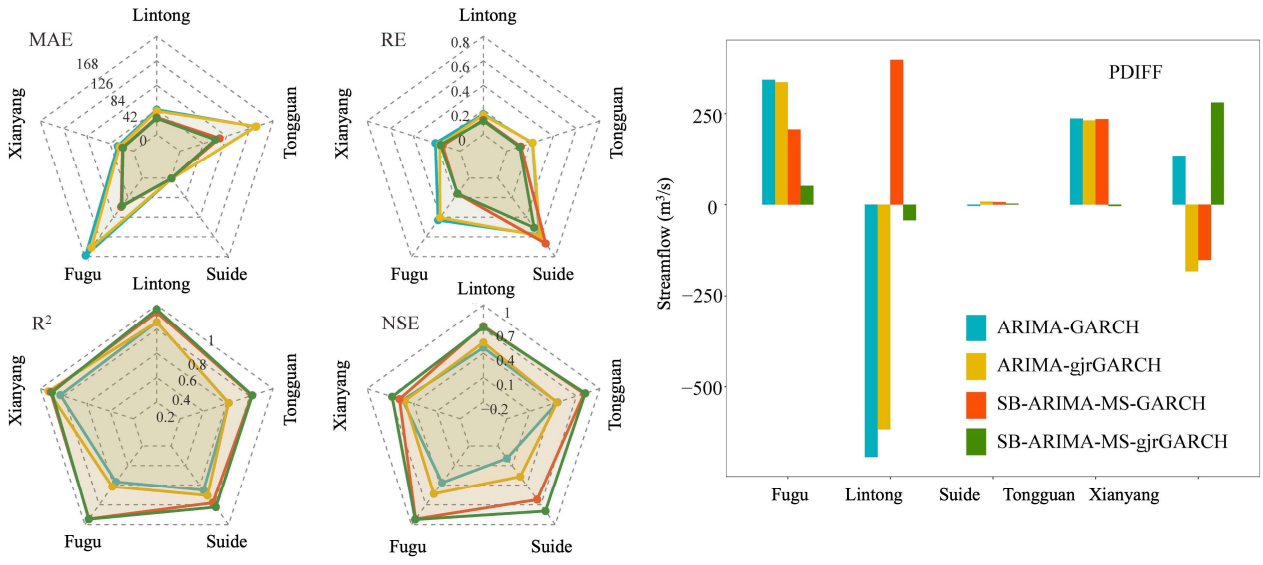


Figure 4. Evaluation metrics of the four models.

Peak-difference results indicate that, except at Xianyang, SB-ARIMA-MS-gjrGARCH more accurately predicts streamflow peaks than SB-ARIMA-MS-GARCH. This may be because the SB-ARIMA-MS-gjrGARCH volatility exhibits short persistence at Xianyang, and peak changes at this station are mainly driven by short-term low prior streamflow values, which may lead to larger peak-flow errors.

Overall, the asymmetric SB-ARIMA-MS-gjrGARCH performs slightly better than its symmetric counterpart. Comparing single- and multi-state models, results from all five stations indicates that multi-state models incorporating SBs and multiple volatility regimes outperform single-state models, especially in their asymmetric form. Among the dimensionless metrics, R^2 describes the collinearity between predictions and observations, while NSE indicates the fit relative to the 1:1 line [36]. Figure 4 demonstrates strong collinearity and close fit between multi-state model predictions and observations, indicating that these models possess strong daily streamflow prediction ability.

4. Discussion

This study integrates multi-state switching with volatility models to improve the predictive performance of single-state volatility models. SB tests showed breakpoints at all stations, likely linked to climate change and land-use change. The IPCC Fifth Assessment Report noted a linear 0.85 °C increase in global mean surface temperature from 1880 to 2012 and stated that mean temperatures during 2003–2012 were 0.78 °C higher than those during 1850–1900 [36]. Located in China's arid–semiarid northwest, the Weihe Basin has a relatively fragile environment that can amplify the impacts of warming [37].

Much of the middle Yellow River traverses the Loess Plateau, where large-scale reforestation and grassland restoration have been implemented since 2002 to mitigate severe soil erosion [38]. These ecological restoration measures have substantially altered regional land-use patterns and affected basin streamflow, producing SBs in streamflow series. Breakpoints at Suide, Lintong, and Xianyang are concentrated in the 1990s and around 2003. As mainstream stations on the Yellow River, Fugu and Tongguan also exhibit SBs in recent years, largely due to intensified water consumption and hydraulic infrastructure construction.

Single-state volatility models that ignore SBs struggle to represent the evolving influence of past shocks on current volatility and fail to capture volatility clustering in streamflow series. Because single-state GARCH and gjrGARCH models are calibrated over long training periods, while volatility clustering often occurs over relatively short intervals, such models may neglect such features and yield streamflow volatility patterns that deviate from reality, thereby degrading predictive performance. By contrast, multi-state models, fit after identifying SBs at each station, describe regime-dependent volatility more accurately. Simulations show that streamflow volatility differs substantially across periods at the same station and that MS-GARCH and MS-gjrGARCH effectively characterize volatility clustering. Moreover, these models allow transitions among volatility regimes; the transition probability matrix (Equation (6)) and the smoothed state-probability matrix (Equation (7)) can be used to compute the probabilities of different volatility states at the next time step, giving MS-GARCH and MS-gjrGARCH greater flexibility in prediction.

The symmetric GARCH variance equation (Equation (2)) includes only squared terms and therefore ignores potential asymmetry in the effects of positive versus negative shocks ε_{t-i} on current volatility. In the ARIMA mean model (Equation (1)), ε_t represents unexpected deviations of daily streamflow from the conditional mean; under strong climate change and human influences, streamflow variability increases and may respond differently to positive and negative ε_t [39]. The gjrGARCH model (Equation (3)) introduces a dummy variable d_t to capture asymmetry in the conditional variance, enabling the model to

represent lower or higher conditional variance induced by positive or negative shocks. This flexibility enables better characterization of streamflow volatility, and, in turn, improves prediction performance. However, the present study assumes normally distributed residuals, while previous research suggests that skewed distributions may fit daily streamflow series more appropriately [4]. Future work could extend the proposed models to incorporate skewed or heavy-tailed distributions, thereby further exploring their applicability to hydrological time-series prediction.

5. Conclusions

This study underscores the importance of accounting for structural changes and volatility regimes in hydrological forecasting. By integrating structural break (SB) detection with a Markov-switching (MS) framework, we developed a series of multi-state GARCH-family models to capture the complex dynamics of daily streamflow in the middle reaches of the Yellow River. The primary findings and their implications are as follows:

(1) The identification of multiple structural breaks across all five stations confirms that streamflow in the middle reaches of the Yellow River is highly non-stationary. This highlights that traditional stationary models are insufficient for capturing the long-term evolutionary shifts driven by environmental and anthropogenic changes in this region.

(2) Compared to single-state models, the multi-state MS-GARCH and MS-gjrGARCH models provide a more nuanced characterization of volatility. By allowing the model parameters to switch between different regimes, these frameworks effectively capture "volatility clustering" and the alternating periods of high and low flow variability that are characteristic of the Loess Plateau.

(3) The integration of structural breaks and asymmetric volatility (SB-ARIMA-MS-gjrGARCH) significantly improves predictive accuracy and model reliability. Specifically, the asymmetric model's ability to distinguish between different directions of flow shocks proves essential for the middle reaches of the Yellow River, where flow extremes are frequent.

In conclusion, under the combined influences of climate change and intensive human activities, the hydrological series of the Yellow River exhibits abnormal volatility and evolving structures. The proposed modeling framework offers a robust new tool for streamflow prediction in such changing environments, providing more reliable evidence for water resource management and flood risk mitigation in the Yellow River basin.

Author Contributions: Conceptualization, J.Z. and H.W.; methodology, J.Z. and H.W.; software, J.S.; validation, Q.Y.; formal analysis, Q.Y.; investigation, J.S.; resources, J.S.; data curation, Q.Y.; writing—original draft preparation, J.Z. and H.W.; writing—review and editing, J.Z., G.Z. and W.S.; visualization, J.S.; supervision, W.S.; project administration, J.Z. and W.S.; funding acquisition, F.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 42401390.

Data Availability Statement: The raw data supporting the conclusions of this paper will be made available by the authors on request.

Conflicts of Interest: Authors Jin Zhao, Jianhui Shang and Weiwei Shou were employed by the company Power China Huadong Engineering Corporation Limited, as well as by Huachen Electric Power Control Corporation Limited. Author Feng Yao was employed by the company Power China Huadong Engineering Corporation Limited. Author Qun Ye was employed by the company Zhejiang XPro info Technology Limited. The remaining authors declare that the research was conducted in the

absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Lei, X.; Wang, H.; Yang, M.; Gui, Z. Advances in hydro-meteorological forecast under changing environment. *J. Hydraul. Eng.* **2018**, *49*, 9–18. (In Chinese) [CrossRef]
2. Seo, S.B.; Kim, Y.-O.; Kim, Y.; Eum, H.I. Selecting climate change scenarios for regional hydrologic impact studies based on climate extremes indices. *Clim. Dyn.* **2019**, *52*, 1595–1611. [CrossRef]
3. Wang, H.; Song, S.; Zhang, G. Daily runoff simulation based on seasonal long-term memory DAR model. *J. Hydraul. Eng.* **2023**, *54*, 686–695. (In Chinese) [CrossRef]
4. Wang, H.; Song, S.; Zhang, G.; Ayantobo, O.O.; Guo, T. Stochastic volatility modeling of daily streamflow time series. *Water Resour. Res.* **2023**, *59*, e2021WR031662. [CrossRef]
5. Wang, W.; Ma, J. Review on some methods for hydrological forecasting. *Adv. Sci. Technol. Water Resour.* **2005**, *25*, 56–60. (In Chinese)
6. Liu, Y.; Wang, W.; Liu, Y.; Ling, Z.; Liu, Q. Advances in multi-source data assimilation approach and application in simulation and forecast of hydrological model. *J. Hohai Univ. (Nat. Sci.)* **2021**, *49*, 483–491. (In Chinese) [CrossRef]
7. Wang, Z.; Si, Y.; Chu, H. Daily streamflow prediction and uncertainty using a long short-term memory (LSTM) network coupled with bootstrap. *Water Resour. Manag.* **2022**, *36*, 4575–4590. [CrossRef]
8. Li, F.; Wang, Z.; Qiu, J. Long-term streamflow forecasting using artificial neural network based on preprocessing technique. *J. Forecast.* **2019**, *38*, 192–206. [CrossRef]
9. Liang, H.; Huang, S.; Meng, E.; Huang, Q. Runoff prediction based on multiple hybrid models. *J. Hydraul. Eng.* **2022**, *51*, 112–125. (In Chinese) [CrossRef]
10. Zhou, T.; Jin, J.; Li, R.; Ji, C.; Li, J. Performance optimization analysis for inflow prediction using wavelet Support Vector Machine. *J. Hydroelectr. Eng.* **2017**, *36*, 45–55. (In Chinese) [CrossRef]
11. Sun, Y.; Niu, J.; Sivakumar, B. A comparative study of models for short-term streamflow forecasting with emphasis on wavelet-based approach. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 1875–1891. [CrossRef]
12. Wang, H.; Gao, X.; Qian, L.; Yu, S. Uncertainty analysis of hydrological processes based on ARMA-GARCH model. *Sci. China Technol. Sci.* **2012**, *55*, 2321–2331. (In Chinese) [CrossRef]
13. Fathian, F.; Fard, A.F.; Ouarda, T.B.M.J.; Dinpashoh, Y.; Mousavi Nadoushani, S.S. Modeling streamflow time series using nonlinear SETAR-GARCH models. *J. Hydrol.* **2019**, *573*, 82–97. [CrossRef]
14. Mehdizadeh, S.; Fathian, F.; Safari, M.J.S.; Adamowski, J.F. Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: A local and external data analysis approach. *J. Hydrol.* **2019**, *579*, 124225. [CrossRef]
15. Fathian, F.; Fakheri-Fard, A.; Modarres, R.; van Gelder, P.H.A.J.M. Regional scale rainfall–runoff modeling using VARX–MGARCH approach. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 999–1016. [CrossRef]
16. Wang, H.; Song, S.; Zhang, G.; Ayantobo, O.O. Time-varying effects of precipitation and temperature on daily streamflow determined using the DCC model. *J. Water Resour. Plan. Manag.* **2023**, *149*, 05023012. [CrossRef]
17. Zha, X.; Xiong, L.; Guo, S.; Kim, J.S.; Liu, D. AR-GARCH with exogenous variables as a postprocessing model for improving streamflow forecasts. *J. Hydrol. Eng.* **2020**, *25*, 04020036. [CrossRef]
18. Guo, T.; Song, S.; Ma, W. Point and interval forecasting of groundwater depth using nonlinear models. *Water Resour. Res.* **2021**, *57*, E2021WR030209. [CrossRef]
19. Pandey, P.K.; Tripura, H.; Pandey, V. Improving prediction accuracy of rainfall time series by hybrid SARIMA–GARCH modeling. *Nat. Resour. Res.* **2019**, *28*, 1125–1138. [CrossRef]
20. Hillebrand, E. Neglecting parameter changes in GARCH models. *J. Econom.* **2005**, *129*, 121–138. [CrossRef]
21. Mikosch, T.; Stărică, C. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Rev. Econ. Stat.* **2004**, *86*, 378–390. [CrossRef]
22. Teterin, P.; Brooks, R.; Enders, W. Smooth volatility shifts and spillovers in U.S. crude oil and corn futures markets. *J. Empir. Finance* **2016**, *38*, 22–36. [CrossRef]
23. Wang, W.; Zhang, J.; Zheng, Y.; Xu, S. Runoff prediction based on BP-SVM model of Markov Chain. *Water Resour. Hydropower Eng.* **2020**, *11*, 78–84. (In Chinese) [CrossRef]
24. Li, W.; Wang, X.; Pang, S.; Guo, H. A runoff prediction model based on nonhomogeneous markov chain. *Water Resour. Manag.* **2022**, *36*, 1431–1442. [CrossRef]
25. Zhou, X.P.; Zhu, C.; Li, T.X. Annual runoff prediction based on least square support vector machines-Markov chain combined model. *Adv. Mater. Res.* **2012**, *418*, 2114–2117. (In Chinese) [CrossRef]
26. Chan, J.Y.-L.; Phoong, S.W.; Phoong, S.Y.; Cheng, W.K.; Chen, Y.L. The Bitcoin halving cycle volatility dynamics and safe haven-hedge properties: A MSGARCH approach. *Mathematics* **2023**, *11*, 698. [CrossRef]

27. Segnon, M.; Gupta, R.; Wilfling, B. Forecasting stock market volatility with regime-switching GARCH-MIDAS: The role of geopolitical risks. *Int. J. Forecast.* **2024**, *40*, 29–43. [CrossRef]
28. Bai, J.; Perron, P. Computation and analysis of multiple structural change models. *J. Appl. Econom.* **2003**, *18*, 1–22. [CrossRef]
29. Ljung, G.M.; Box, G.E.P. On a measure of lack of fit in time series models. *Biometrika* **1978**, *65*, 297–303. [CrossRef]
30. Engle, R.F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **1982**, *50*, 987–1007. [CrossRef]
31. Dawson, C.W.; Abrahart, R.J.; See, L.M. HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environ. Model. Softw.* **2007**, *22*, 1034–1052. [CrossRef]
32. Liu, C.; Cui, N.; Gong, D.; Hu, X.; Feng, Y. Evaluation of seasonal evapotranspiration of winter wheat in humid region of East China using large-weighted lysimeter and three models. *J. Hydrol.* **2020**, *590*, 125388. [CrossRef]
33. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]
34. Zhang, J.; Wang, G.; He, R.; Liu, C. Variation trends of runoffs in the Middle Yellow River basin and its response to climate change. *Adv. Water Sci.* **2009**, *20*, 153–158. (In Chinese) [CrossRef]
35. Cui, Z.; Zhou, Y.L.; Guo, S.L.; Wang, J.; Ba, H.; He, S. A novel hybrid XAJ-LSTM model for multi-step-ahead flood forecasting. *Hydrol. Res.* **2021**, *52*, 1436–1454. [CrossRef]
36. Qin, D.; Stocker, T. Highlights of the IPCC Working Group I Fifth Assessment Report. *Prog. Inquisitiones Mutat. Clim.* **2014**, *10*, 1–6. (In Chinese) [CrossRef]
37. Ren, L.; Shen, H.; Yuan, F.; Zhao, C.; Yang, X.; Zheng, P. Hydrological drought characteristics in the Weihe catchment in a changing environment. *Adv. Water Sci.* **2016**, *27*, 492–500. (In Chinese) [CrossRef]
38. Xue, J. Impacts of a new round of returning farmland to forests and grasslands project on land use efficiency—Empirical study in Yellow River Basin. *Bull. Soil Water Conserv.* **2023**, *43*, 155–164. (In Chinese) [CrossRef]
39. Song, S. Stochastic modeling research progress of hydrological non-stationary heteroscedasticity series in the changing environment of watershed. *J. Water Resour. Water Eng.* **2023**, *34*, 1–6. (In Chinese) [CrossRef]

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