Study on Mother Wavelet Optimization Framework Based on Changepoint Detection of Hydrological Time Series

Jiqing Li[1], Jing Huang[1], Lei Zheng[1], Wei Zheng[1]

[1]School of Water Resources and Hydropower Engineering, North China Electric Power University, Beijing, 102206, China

Correspondence to: Jing Huang (jinghuang23@163.com)

Abstract. Hydrological time series (HTS) is the key basis of water conservancy project planning and construction. However, under the influence of climate change, human activities and other factors, the consistency of HTS has been destroyed and cannot meet the requirements in mathematical statistics. Series dividing and wavelet transform are effective methods to reuse and analyse HTS. However, they are limited by the changepoint detection and mother wavelet (MWT) selection, and are difficult to apply and promote in practice. To address these issues, we constructed a potential changepoint set based on cumulative anomaly method, Mann-Kendal test and wavelet changepoint detection. Then, the degree of change before and after the potential changepoint was calculated with the Kolmogorov-Smirnov test, and the changepoint detection criteria was proposed. Finally, the optimization framework were proposed according to the detection accuracy of MWT, and continuous wavelet transform was used to analyse HTS evolution. We used Pingshan Station and Yichang Station in the Yangtze River as study cases. The result shows: (1) Changepoint detection criteria can quickly locate potential changepoints, determine the change trajectory and complete the division of HTS. (2) MWT optimal framework can select the MWT that conforms to HTS characteristics and ensure the accuracy and uniqueness of the transformation. This study analyses the HTS evolution and provides a better basis for hydrological and hydraulic calculation, which will improve the design flood estimation and the operation scheme preparation.

1 Introduction

Under the multiple influences of human activities, atmospheric circulation and other factors, the original evolution of river runoff is featured with randomness, fuzziness, nonlinearity, non-stationarity and multi-time scale variation, which breaks the consistency in the "three properties" of hydrological time series (formed by the time arrangement of hydrological elements such as rainfall and runoff, HTS) (Chen et al., 2021; Fang and Shao, 2022). Independent and identically distributed (IID) is an assumption of mathematical statistics in hydrological and hydraulic calculation (Mat Jan et al., 2020). When the series cannot meet the IID, analysing its internal evolution and division will help to improve the accuracy and decision-making of the hydrological forecasting and operation scheme preparation by the mathematical model (Li et al., 2021).

In stochastic hydrology, HTS consists of deterministic components and stochastic components. The analysis of its evolution involves period, trend and changepoint (Sanaa et al., 2022). Period and trend mainly focus on deterministic components,

while changepoint detection is used to explain the stochastic components caused by various random and uncertain factors (Dang et al., 2021). Changepoint detection determines the starting and ending points of period and trend division, thus is the key to analyse HTS evolution (Şen, 2021). However, affected by feature uncertainty, changepoint detection has become a complex problem because the extent, number and occurrence time of changepoints must be determined at the same time (Zhao et al., 2019). T-test, two-sample Kolmogorov-Smirnov (K-S) test and Shapiro-Wilk test are commonly used quantitative methods for series variation. In particular, the K-S test can calculate the degree of change by indicators such as asymptotic significance (two-tailed, p), therefore it is widely used (Jia et al., 2022).

Commonly used changepoint detection methods include graphical methods (cumulative anomaly method, etc.), parametric methods (sliding T-test, Lee-Heghinian test, etc.) and nonparametric methods (ordered clustering method, Mann-Kendall test, wavelet changepoint detection, etc.). Graphical methods have the advantages of simple calculation and intuitive results, but the detection accuracy are low. Parametric methods assume that the series to be analysed obeys a known distribution, which have certain limitations (Liu et al., 2022). Nonparametric methods have higher detection accuracy, but are easily affected by factors such as parameter settings and series marginal effects (Stasolla and Neyt, 2019). Malki et al. (2022) used machine learning to compare the gap between historical data and forecasts from real-time monitoring data to determine whether the consistency of IoT energy consumption data has changed. Shi et al. (2022) constructed a single changepoint test based on the covariance, cumulative sum and likelihood ratio of forecast residuals to detect the potential changepoint in time series. Corradin et al. (2022) constructed a Bayesian nonparametric multivariate changepoint detection method by combining prior distributions with multivariate kernels, and argued that the posterior probability of most changepoints should be lower than the posterior estimate. Xie et al. (2022) calculated the fitted local trend line based on the piecewise linear representation algorithm and the Akaike information criterion to realize changepoint detection and series division, and classified changepoints into three categories with the help of slope and intercept. Changepoint detection is of great significance to series division and is the basis for making full use of HTS to carry out more research. It can be seen that there is no unified standard to determine the changepoint of HTS. Therefore, this is a field worthy of further study.

After the changepoint detection, the period and trend of HTS can be further explored. These methods include cumulative anomaly method, Mann-Kendall (M-K) test, continuous wavelet transform (CWT) and mode decomposition (empirical, extreme point symmetric, etc.) (Oliveira-Júnior et al., 2022; Qin et al., 2021). Among them, CWT has a relatively complete theoretical system, which can comprehensively analyse the evolution of HTS and reveal its localization characteristics in in time domain (time variation) and frequency domain (frequency and amplitude variation), so it has been widely used in hydrology (Zerouali et al., 2022). However, the analysis results of CWT highly depend on the selection of mother wavelet (MWT). Moradi et al. (2022) optimized MWT by comparing the similarity of cross-correlation function, signal-to-noise ratio, and mean standard error between the denoised series and the original. Benhassine et al. (2021) determined the optimal MWT by comparing the minimum mean square error between the original image and the denoised. Strömbergsson et al. (2019) proposed and verified the validity of using the Shannon entropy of the wavelet coefficients as the index for selecting MWT.

However, changepoint detection has not been explored by scholars to optimize the MWT that conforms to the series characteristics.

To solve the above problems, we proposed the changepoint detection criteria based on cumulative anomaly method, M-K test, wavelet changepoint detection and K-S test, which can detect the consistency of HTS and complete a reasonable division. Furthermore, based on the detection accuracy, a MWT optimal framework that conforms to series characteristics was proposed, and the evolution analysis was summarized by CWT. This work pioneeringly proposed an efficient way to optimize the MWT based on variance and changepoint detention. Using the optimal MWT in CWT is helpful in catching the HTS evolution accurately and fully mining its information, which provides a feasible way to use inconsistent measured data for hydrological and hydraulic calculations.

2 Methodology

To solve the problems of incomplete changepoint detection and non-unique MWT optimization, we followed the process of potential changepoint set construction, changepoint determination, MWT optimization and evolution analysis, then proposed the changepoint detection criteria and the MWT optimization framework, as shown in Figure 1.

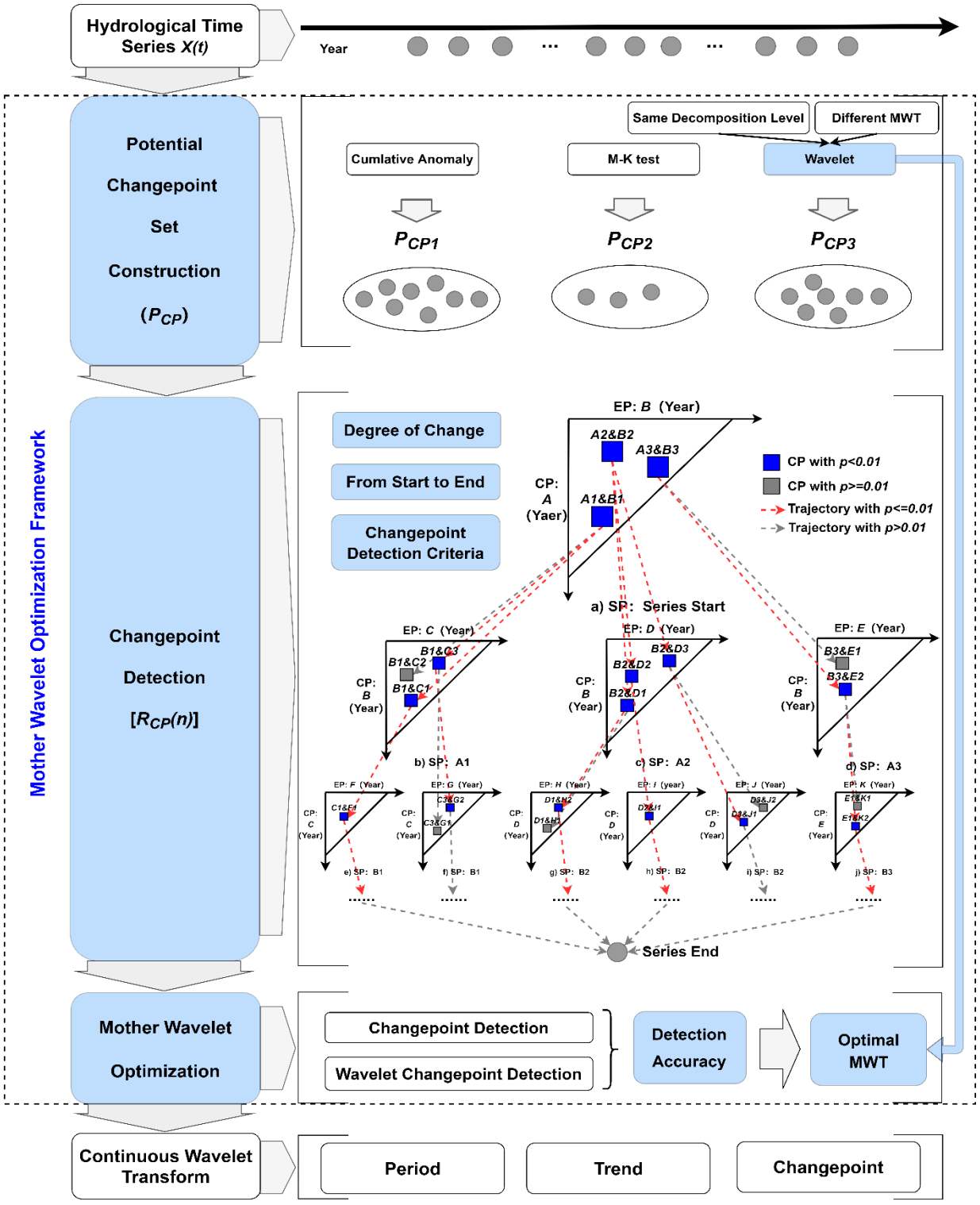


Figure 1: Study framework and main modules of MWT optimization

2.1 Wavelet Transform and Changepoint Detection

Wavelet transform can be divided into Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). Its essence is to reveal the similarity between the HTS to be analysed and the MWT. Therefore, the selection of MWT is a key factor affecting the accuracy of wavelet transform. MWT ($\varphi(t)$) is a wave of finite length and zero mean, with irregularity and asymmetry. The 16 commonly used MWT systems are shown in Table 1 (Moradi, 2022; Nielsen, 2001).

Table 1: Properties and application range of commonly used MWT systems

MWT System	Symbol	Properties and Application Range[1]				
		Orthogonality	Biorthogonality	Symmetry	CWT	DWT
Haar	haar	√	√	√	√	√
Daubechies	db2, db3, db4, db5, db6, db7, db8, db9, db10	√	√	√*	√	√
Biorthogonal	bior1.1, bior1.3, bior1.5, bior2.2, bior2.4, bior2.6, bior2.8, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8	—	√	—	√	√
Coiflets	coif1, coif2, coif3, coif4, coif5	√	√	√*	√	√
Symlets	sym2, sym3, sym4, sym5, sym6, sym7, sym8	√	√	√*	√	√
Morlet	morl	—	—	√	√	—
Mexican Hat	mexh	—	—	√	√	—
Meyer	meyr	√	√	√	√	√*
Gaussian	gaus1, gaus2, gaus3, gaus4, gaus5, gaus6, gaus7, gaus8	—	—	√	√	—
Dmeyer	dmey	—	—	√	—	√
ReverseBior	rbio1.1, rbio1.3, rbio1.5, rbio2.2, rbio2.4, rbio2.6, rbio2.8, rbio3.1, rbio3.3, rbio3.5, rbio3.7, rbio3.9, rbio4.4, rbio5.5, rbio6.8	—	√	√	√	√
Complex Gaussian	cgau1, cgau2, cgau3, cgau4, cgau5, cgau6, cgau7, cgau8	—	—	√	—	—
Complex Morlet	cmor1-1.5, cmor1-1, cmor1-0.5, cmor1-0.1	—	—	√	—	—
Frequency B-Spline	fbsp1-1-1.5, fbsp1-1-1, fbsp1-1-0.5, fbsp2-1-1, fbsp2-1-0.5, fbsp2-1-0.1	—	—	√	—	—
Fejer-Korovkin	fk4, fk6, fk8, fk14, fk18, fk22	√	√	√*	√	√
Shannon	shan1-1.5, shan1-1, shan1-0.5, shan1-0.1, shan2-3	—	—	√	—	—

Note 1: "√" means has this property. "√*" means approximately having this property. "—" means does not have this property.

2.1.1 Continuous Wavelet Transform (CWT)

CWT can be used to determine whether there is periodicity in HTS, and identify the main time scales and their local trends. Let $L^2(R)$ denote the measurable square-integrable functions on the real axis. If HTS $X(t)$ $(t=1,2,...,T)$ is a CWT in $L^2(R)$, which can be expressed as:

$$W_X(a,b) = \int_{-\infty}^{+\infty} X(t)\varphi_{a,b}^*(t)dt \tag{1}$$

$$\varphi_{a,b}(t)=\frac{1}{\sqrt{a}}\varphi\left(\frac{t-b}{a}\right)\ a,b\in R,\ a\neq 0 \tag{2}$$

Where, $W_X(a,b)$ is the coefficient of CWT. $\varphi^*_{a,b}(t)$ is the complex conjugate function of $\varphi_{a,b}(t)$. t is the time. a is the time scale factor, which reflects the period length of MWT. b is the time position factor, which reflects the translation of MWT in time.

The multi-time scale variation in wavelet transform refers to the multi-level structure and localized features of $X(t)$ in the time domain, which is usually analysed with the help of the real part or modulus-square contour map of CWT coefficients. HTS evolution of a certain year on different time scales can be observed by vertically intercepting the contour map. At a certain period, the HTS evolution over time can be observed by horizontally intercepting the contour map. In addition, the positive wavelet coefficient corresponds to the wet season. The negative wavelet coefficient corresponds to the dry season. The wavelet coefficient is zero, which corresponds to the transition point of wet and dry. The larger the absolute value of the wavelet coefficient, the more obvious its change.

2.1.2 Discrete Wavelet Transform (DWT)

Since the measured HTS is usually discrete, by discretizing Eq.1, we can get:

$$W_X(j,b)=\int_{-\infty}^{+\infty}X(t)\varphi^*_{j,b}(t)dt \tag{3}$$

$$\varphi_{j,b}(t)=a_0^{\frac{-j}{2}}\varphi\left(a_0^{-j}t-kb_0\right) \tag{4}$$

Where, $W_X(j,b)$ is the coefficient of DWT. a_0 and b_0 are both constants. j $(j=1,2,...,J)$ is the decomposition level.

Both $W_X(a,b)$ and $W_X(j,b)$ are the values output by $X(t)$ through the unit impulse response filter, which can reflect the evolution of $X(t)$ in the time domain and frequency domain at the same time. In practical applications, it is often decomposed with the help of dyadic DWT, i.e. $a_0=2$ and $b_0=1$, Eq.4 can be expressed as:

$$\varphi_{j,b}(t)=2^{\frac{-j}{2}}\varphi\left(2^{-j}t-k\right) \tag{5}$$

According to the dyadic DWT, the theoretical maximum value J of decomposition level j is:

$$J=\left[\log_2\left(T_{X(t)}\right)\right] \tag{6}$$

Where, $[\cdot]$ represents rounding operation. $T_{X(t)}$ represents the length of the $X(t)$.

2.1.3 Wavelet Changepoint Detection

Variance is one of the important parameters to detect whether HTS has fundamentally changed. Wavelet changepoint detection is based on the Maximal Overlap Discrete Wavelet Transform (MODWT). By calculating the variance of wavelet coefficients to be analysed one by one (Strömbergsson et al., 2019), the number and location of changepoint at Confidence Level 95% can be determined through the MATLAB software toolbox.

(1) MODWT multi-resolution analysis

Decompose $X(t)$ into T-dimensional column vectors W_1, W_2, …, W_J and V_J. Where W_J is calculated from the MODWT wavelet coefficient of $X(t)$ within $\tau_j \Delta t$. V_J consists of $\tau_{j+1}\Delta t$ and higher dimensional MODWT scaling coefficients. $X(t)$ can be expressed as:

$$X = \sum_{j=1}^{J} D_j + S_j \tag{7}$$

Where, $D_j = W_{j^{F^k}}^{F} h_j^{*}$ $(k = 0,1,...,T-1)$ is the j^{th} maximal-overlap detail. $S_j = V_{j^{F^k}}^{F} g_j^{*}$ is the j^{th} maximal-overlap smooth. h_j and g_j are the high-frequency filter and the low-frequency filter, respectively. F is a $T \times T$ dimensional matrix that cyclically shifts h_j by one unit.

(2) MODWT variance decomposition

After a series of decompositions are performed on the variance of $X(t)$ part by part, on the premise that the wavelet coefficient is stable, it can be expressed as:

$$\|X\|^2 = \sum_{j=1}^{J} \|W_j\|^2 + \|V_j\|^2 \tag{8}$$

Based on the above decomposition, the evolution of wavelet coefficient variance of $X(t)$ with time in different time scales can be obtained, and the point where the variance changes can be recorded as the changepoint. It is worth noting that the MWT used for changepoint detection needs to be biorthogonal (see Table 1).

2.2 Traditional Changepoint Detection Method

Changepoint detection has always been a significant issue in hydrology. However, except for the deterministic runoff changes caused by human activities such as large-scale river regulation, reservoir construction or operation (seasonal and

above regulation capacity), there exist many uncertain factors, such as whether there is a change point in HTS, how many changepoints exist, and the specific occurrence time of each changepoint. Therefore, it is necessary to integrate multiple detection methods. The main methods used in this study are as follows.

2.2.1 Cumulative Anomaly Method

Cumulative anomaly is a graphic method. The cumulative anomaly value of $X(t)$ at a certain time can be expressed as:

$$JP\left[X(t)\right]=\sum_{t=1}^{N}\left[X(t)-\overline{X}\right] \tag{9}$$

Where, $JP[\cdot]$ is the cumulative anomaly value of $X(t)$. T and $\overline{X}$ are the length and mean of $X(t)$, respectively.

The cumulative anomaly curve can be obtained by drawing the cumulative anomaly value in chronological order. According to the curve fluctuation, the change trend and potential changepoint of HTS can be identified. If the cumulative anomaly value is greater than 0, it indicates that the HTS is in an uptrend, otherwise, the HTS is in a downtrend. The point that changes the trend can be regarded as the potential changepoint.

2.2.2 Mann-Kendall（M-K）Test

The M-K test analyses the number, location, trend and significance of changepoints in HTS by setting a Confidence Level α and calculating statistics (U_{F_k} and U_{B_k}). Statistics U_{F_k} of $X(t)$ is calculated as follows:

$$U_{F_k}\left[X(t)\right]=\frac{S_k^{X(t)}-E\left[S_k^{X(t)}\right]}{\sqrt{Var\left[S_k^{X(t)}\right]}} \tag{10}$$

Where, $U_{F_k}\left[X(t)\right]$ is the statistical series of $X(t)$ calculated in order. $S_k^{X(t)}$ is the rank sum of Time k in $X(t)$, which is the cumulative value of the numbers at Time k greater than Time i $(1\le k\le i)$. $E\left[S_k^{X(t)}\right]$ and $Var\left[S_k^{X(t)}\right]$ are the mean and variance of $S_k^{X(t)}$, respectively.

When $U_{F_k}\left[X(t)\right]>0$, $X(t)$ shows an upward trend, it shows a downward trend. The statistic $U_{B_k}\left[X(t)\right]$ is obtained by repeating Eq.10 in the reverse order. Draw $U_{F_k}\left[X(t)\right]$ and $U_{B_k}\left[X(t)\right]$ in the same figure. If the two statistics intersect within the Confidence Interval $U_{0.05}=\pm1.96$ (Confidence Level 95%), the time corresponding to the intersection is the changepoint of $X(t)$.

2.2.2 Kolmogorov-Smirnov (K-S) Test

The K-S test can determine whether the distributions of the two series are the same according to the maximum vertical distance between the two empirical distributions. The empirical distribution of $X(t)$ is:

$$F_n[X(t)] = \frac{1}{T}\sum_{t=1}^{T} I^n_{[-\infty,T]}[X(t)] \tag{11}$$

Where, $I^n_{[-\infty,T]}[X(t)]$ is the indicator function of $X(t)$.

Original hypothesis H_0: $F_1[X(t)] = F_2[X(t)]$, that is, the empirical distribution of the two series is consistent. Alternative hypothesis H_1: $F_1[X(t)] \neq F_2[X(t)]$, that is, the empirical distribution is inconsistent. To quantify the difference between the empirical distributions, a maximum difference D is proposed, calculated as:

$$D = \sup_{-\infty<X(t)<\infty} \left|F_1[X(t)] - F_2[X(t)]\right| \tag{12}$$

$D_{T,\alpha}$ is used to represent the rejection domain when the series capacity is T at Significant Level α. When $D \geq D_{T,\alpha}$, reject H_0, otherwise, accept H_0. To further quantify the significance of the difference, p is introduced to concretize α. The value of α is usually 95% or 99%, and the corresponding p is 0.05 and 0.01. If $p \leq 0.01$, it indicates that the determination result is strong and H_0 should be rejected, that is, the two series obey different distributions and are not consistent. If $0.01 \leq p \leq 0.05$, the determination result is weak. In this case, p is considered to be marginal, and H_0 is usually rejected. If $p > 0.05$, H_0 is acceptable.

2.3 Changepoint Detection Criteria

Based on the changepoint detection results of various methods, the potential changepoint set $P_{CP}(n)$ $(n = 1, 2, ..., N)$ of HTS is constructed with deduplication and sorting. To determine the changepoint, it is necessary to further calculate the degree of change (p) before and after potential changepoints with the help of the K-S test. At Confidence Level 99%, first, record the starting point and ending point of $X(t)$ as $P_{CP}(0)$ and $P_{CP}(N+1)$ respectively, and arrange the potential changepoint set in chronological order. Secondly, take $P_{CP}(0)$ as the starting point and $P_{CP}(1)$ as the change point, and use K-S test to successively calculate the p of the end point from $P_{CP}(2)$ to $P_{CP}(N+1)$. Finally, the changepoint and its trajectory (connection of change points) of $X(t)$ are determined according to the changepoint detection criteria:

Criterion ①: Before and after the changepoint of $X(t)$, $p < 0.01$.

Criterion ②: The changepoint can realize the continuous division of $X(t)$ from $P_{CP}(0)$ to $P_{CP}(N+1)$.

Criterion ③: The trajectory contains the largest number ($m = 1, 2, ..., M$) of changepoints.

Criterion ④: The p of $M-1$ in the trajectory be the minimum value.

2.4 MWT Optimization Framework

By comparing $R_{CP}(n)$ and the results of wavelet changepoint detection, a MWT that conforms to HTS characteristics can be selected. The MWT optimization framework includes the construction of potential changepoint set, changepoint detection and optimal MWT determination. Among them, the potential changepoint set is built to improve the efficiency of changepoint detection, and the specific optimization steps are as follows:

Optimization Step (1): Select candidate wavelet with the highest changepoint detection accuracy.

Optimization Step (2): When two or more candidate wavelets have the same detection accuracy, the MWT or the MWT system with the highest frequency in different statistic series (length, flow, etc.) of the same hydrological station is selected as the optimal one.

After optimization, we can perform CWT according to the MWT conforming to HTS characteristics and analyse its evolution. For DWT, HTS can be more accurately decomposed and reconstructed, providing a good basis for hydrological forecasting and reservoir operation scheme formulation.

3 Data and Study Area

The Yangtze River originates from the southwest of the Tanggula Mountains on the Qinghai-Tibet Plateau. Its main stream flows through central China from west to east, with a total length of about 6,300 km, and the total catchment area is 1.8 million km^2, accounting for about 18.8% of the total area of China. The main stream from Yibin to Yichang is called the upstream, with a length of about 4,504 km and an area of about 1 million km^2. With the superposition and collection of upstream floods to the Yichang Hydrological Station (Yichang Station), it tends to form a process of high peaks and large volumes (Shuhui et al., 2021). The Pingshan Hydrological Station (Pingshan Station) on the Jinsha River controls about half of catchment area and one-third of the flood season average flow of Yichang Station, and is the basic source of upstream flooding. Therefore, exploring the runoff evolution at Pingshan Station and Yichang Station will help to scientifically arrange the watershed storage space to alleviate the frequent floods in flood seasons and water shortages in dry seasons in the middle and lower Yangtze River. The overview of the upper Yangtze River is shown in Figure 2, and the hydrological parameters of the tow stations are shown in Table 2.

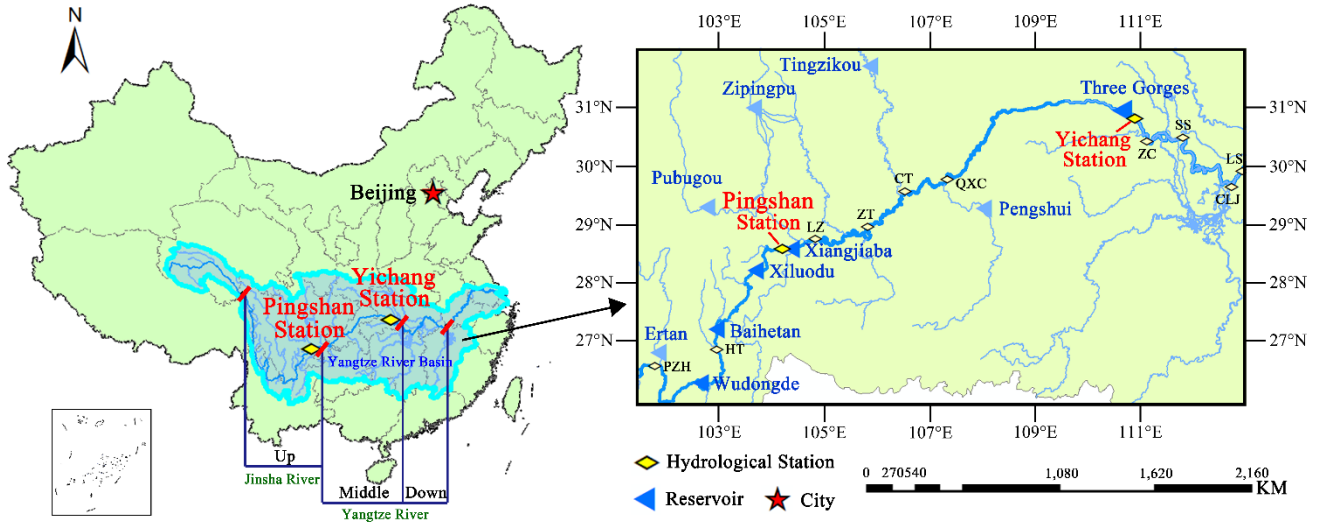


Figure 2: Location of the study area

Table 2: Main hydrological parameters of Pingshan Station and Yichang Station

River Hydrological Station		Jinsha Pingshan	Yangtze Yichang
Catchment Area	Area (km^2) Proportion (%)	485,099 48.2	1,005,501 100
Annual Average Water Volume	Volume (10^8 m^3) Proportion (%)	1,147 33.6	3,410 100
Annual Distribution of Runoff	Flood Season (month) Flow (m^3/s) Proportion (%)	6-11 44,850 81.34	5-10 127,700 78.67

The flood season of Pingshan Station is from June to November, and the flood season of Yichang Station is from May to October. The three months with the largest flow on the two stations are both from July to September (accounting for 49.96% and 54.18% of the year, respectively). In 2012, Pingshan Station was moved down 24 km to Xiangjiaba Hydrological Station. In addition, the runoff of Pingshan Station should consider the influence of the upstream Ertan Reservoir (seasonal regulation, water storage in May 1998), and Yichang Station should consider the Three Gorges Reservoir (annual regulation, water storage in June 2003). Combined the above factors, the measured runoff data of Pingshan Station (1950-2011) and Yichang Station (1950-2016) were used to test the applicability of the changepoint detection framework and the MWT optimization framework proposed in this study, and the runoff evolution of the two stations was analysed by CWT.

4 Results and Discussion

The statistical series of the two stations used in the study includes: Pingshan annual mean runoff series (Pingshan Annual Series, PAS), Pingshan 6-11 mean runoff series (Pingshan Flood Season Series, PFSS), Yichang annual mean runoff series

(Yichang Annual Series, YAS) and Yichang 5-10 mean runoff series (Yichang Flood Season Series, YFSS), collectively referred to as 4-Series.

4.1 Construction of Potential Changepoint Set

The cumulative anomaly method, M-K test and wavelet changepoint detection were used to detect the potential changepoints in the 4-Series. At the same time, by comparing the annual series and the flood season series at the same station, we further analysed the sensitivity of the three methods to the variation of flow amplitude and the influence of flood season on the annual series.

4.1.1 Results of Cumulative Anomaly Method and M-K Test

The points causing the trend change can be regarded as potential changepoints, and the detection results of the cumulative anomaly method are shown in Figure 3. At Confidence Level 95% (the upper and lower critical lines are ±1.96), the intersection of U_{F_k} and U_{B_k} is a potential changepoint, and the M-K test results are shown in Figure 4. Potential changepoints in the two figures were marked in red.

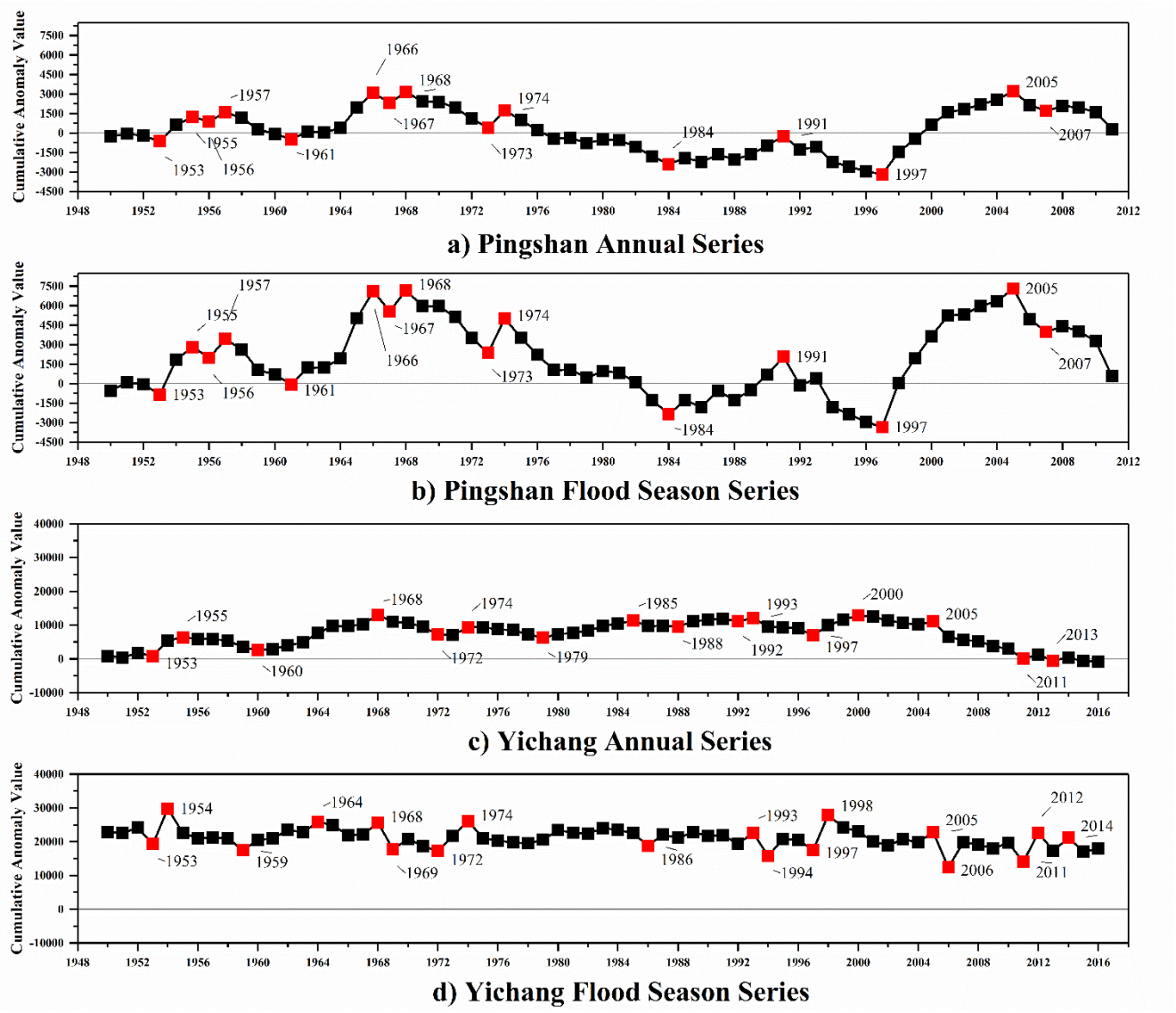


Figure 3: Potential changepoints of the cumulative anomaly method at Pingshan Station and Yichang Station

The number of potential changepoints of 4-Series detected by the cumulative anomaly method is 15, 15, 16 and 18 (Figure 3). However, the number detected by the M-K test is 2, 2, 0 and 0 (Figure 4). In addition, there are differences in the potential changepoint detection results between the annual series and the flood season series, indicating that the cumulative anomaly method has certain response ability to flow changes. However, the consistent rate of potential changepoints in Pingshan Station is 100%, while Yichang Station is 37.5% and 33.33%, respectively. This means that the response ability can only be reflected when the flow variation reaches a certain extent.

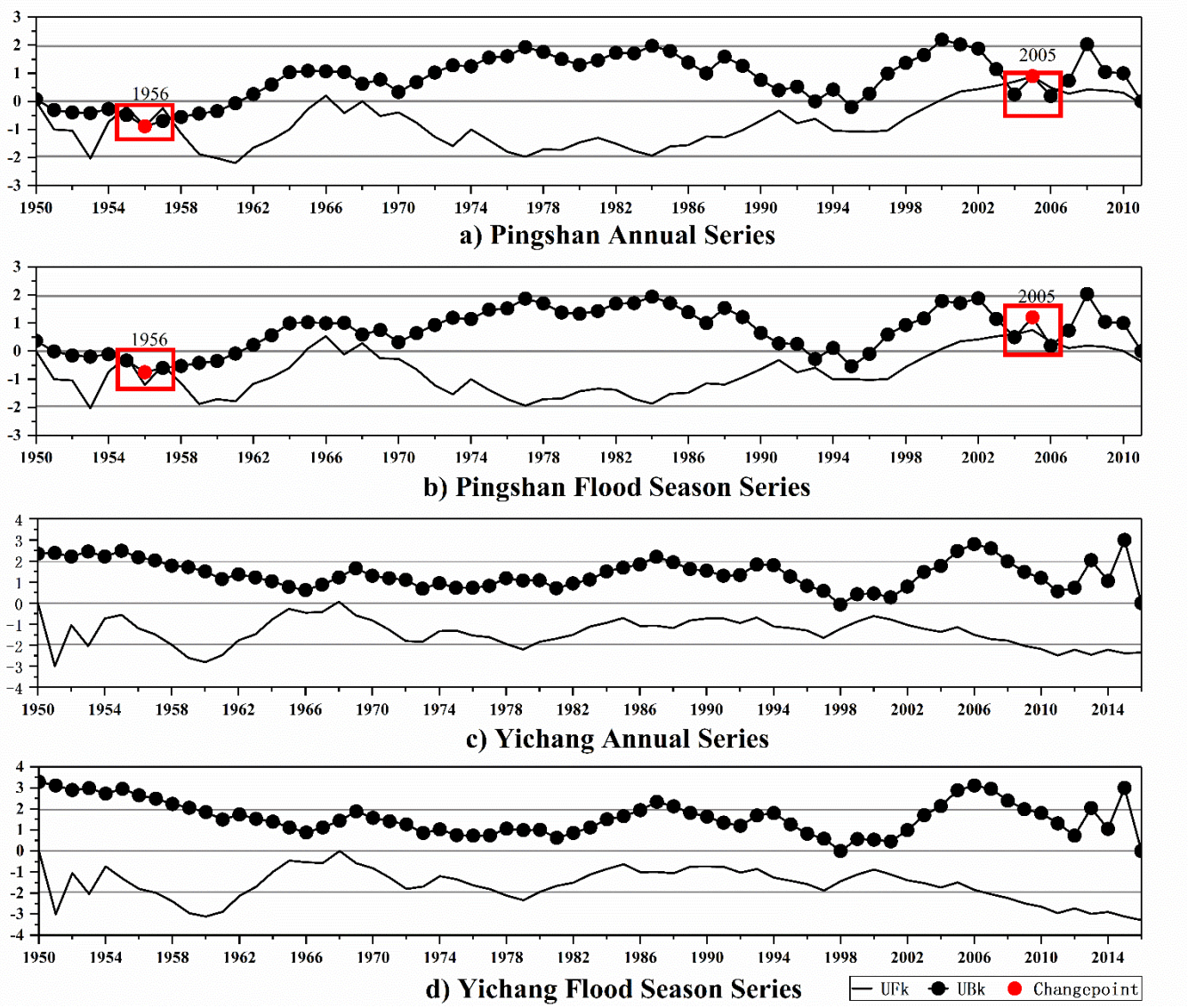


Figure 4: Potential changepoints of the M-K test at Pingshan Station and Yichang Station

The changepoint detection results of M-K test at Pingshan Station (Figure 4a-b) are concentrated around 1956 and 2005. During the same time scale, the intersection of the flood season series is slightly later than the annual series, but the amplitude of U_{F_k} and U_{B_k} is lower, which indirectly reflects the flood season in Pingshan Station is relatively gentle, but the difference between the wet and dry seasons of the year is obvious. The YFSS is the opposite. In addition, the detection results of M-K test for 4-Series are basically consistent, insensitive to flow variation. The detected number of potential changepoints is small. It can be included that the cumulative anomaly method is more suitable for constructing the potential changepoint set of HTS. A more accurate locating of the changepoint needs other methods.

4.1.2 Results of Wavelet Changepoint Detection

Among the 16 commonly used MWT systems, 8 of them satisfy the biorthogonality (59 MWT in total). In this study, 59 MWT were used to detect the potential changepoints of 4-Series one by one, and the number of decomposition layers used is 5. However, only 5 MWT systems can detect the changepoints of 4-Series, as shown in Table 3.

Table 3: Wavelet changepoint detection results of biorthogonal MWT at Pingshan Station and Yichang Station (decomposition layers is 5)

MWT Systems	Symbol	PAS[1]			PFSS[2]		YAS[3]			YFSS[4]	
Daubechies	db2	1999	1985		1999		1996	1975	1961	1977	1975
	db3	——			1985		1968			——	
	db4	1999	1995	1992	1999	1992	1962			1960	
	db5	——			2000	1963	——			——	
	db6[3]	2000	1965		2000	1965	**2002[3]**			1972	
	db7	——			——		1962			2000	
	db8[12]	**1998[1]**	1992		**1998[2]**	1991	2004			2005	
	db9	1965			1964		1966			1998	
	db10	1983	1959		——		1992	1965		1994	1967
Symlets	sym2	1999	1985		1999		1996	1975	1961	1977	1975
	sym3	——			1985		1968			——	
	sym4	1996	1990		1996		1959			1959	
	sym5	——			1983		2003			——	
	sym6	1989	1963		1962		1969			2005	
	sym7	1967			——		——			——	
	sym8	1989			——		1998			1999	
Coiflets	coif1	——			——		1968	1961		——	
	coif2	1990	1960		1964		1971			2005	1972
	coif3	——			——		1966			1993	
	coif4	1993	1992		1993	1990	1990			——	
	coif5	1968			1968		1998	1985		1969	
Dmeyer	dmey	1969	1966		1968	1965	——			——	
Fejer-Korovkin	fk4	1996			1996		1995	1971		1975	1969
	fk6	——			——		1968			——	
	fk8	**1998[1]**	1992	1990	**1998[2]**	1989	1961			1984	1959
	fk14[4]	——			2000		1966			**2003[4]**	
	fk18	——			1966		2000			1992	
	fk22	——			1959		——			1983	

Note: The changepoint and the optimal MWT are marked with the same number (in the upper right corner) as the series.

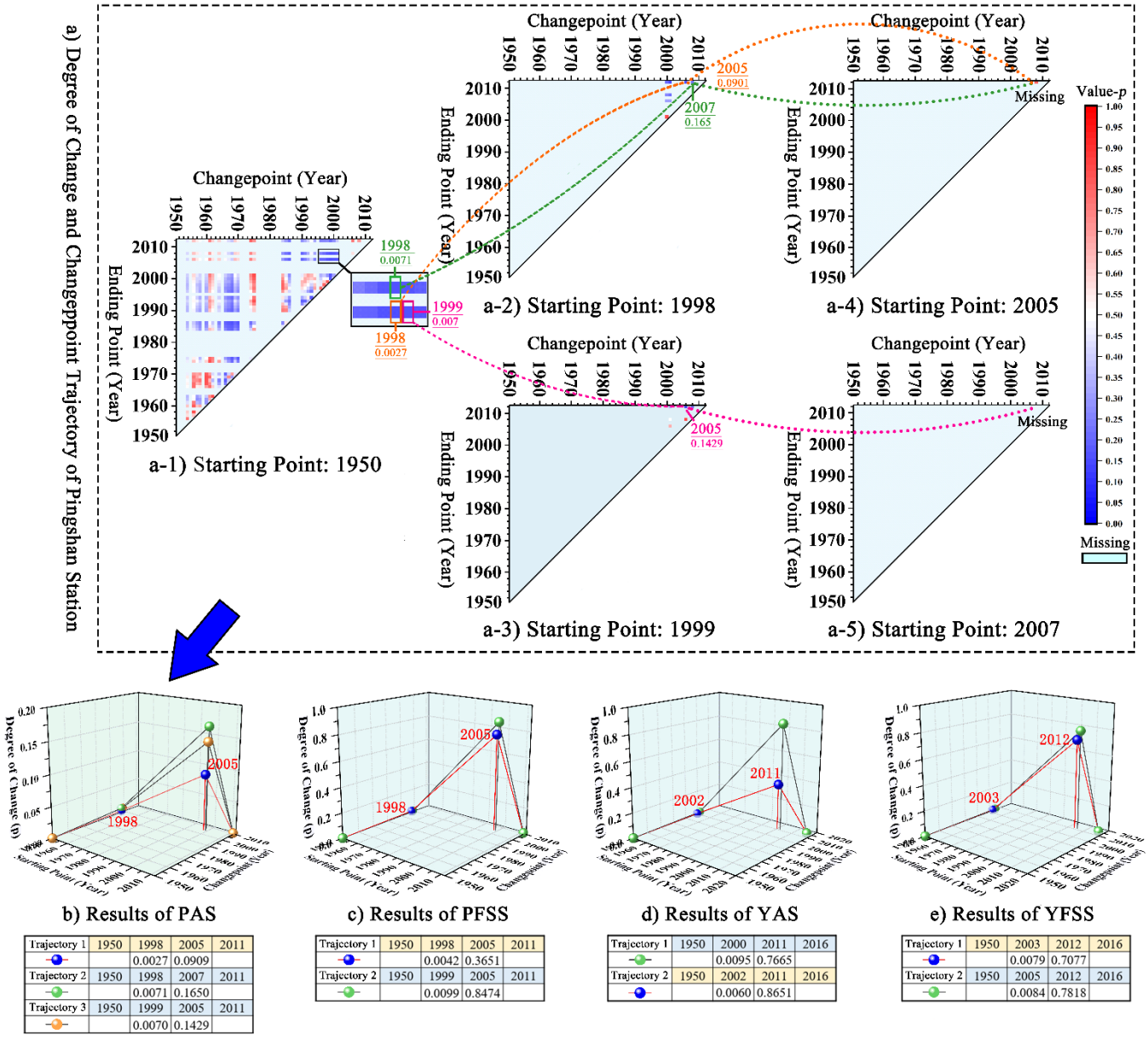
From Table 3, the number of potential changepoints detected by a single MWT is between 1 and 3. The top two potential changepoint of the PAS are 1992 and 1999, the PFSS are 1999 and 2000, the YAS are 1961 and 1968, and the YFSS are

1975 and 2005. The number of 4-Series of changepoints detected is 19, 18, 19 and 17 respectively. Compared with the cumulative anomaly method and M-K test, the wavelet changepoint detection has the highest contribution to the construction of the potential changepoint set, followed by the cumulative anomaly method.

As the MWT changes, the detection results are quite different. For the same hydrological station and the same MWT, there is also a difference in the detection results between the annual series and the flood season series, indicating that the wavelet changepoint detection is very sensitive to the flow variation of HTS. Furthermore, the detection results of Pingshan Station are concentrated in 1959-2000, while Yichang Station are concentrated in 1959-2004. Compared with the series length used in the study (Pingshan 1950-2011 and Yichang 1950-2016), the detection results are susceptible to marginal effects, and the potential changepoints at both ends of the series (before and after 10 years) may be ignored.

4.2 Results of Changepoint Detection

We deduplicated and sorted the above detection results as potential changepoint sets for each series, with capacities of 31, 30, 31, and 28, respectively. The degree of change (p) before and after each potential changepoint was calculated by the K-S test. Traditional changepoint detection often adopts the method of traversal series. Take PAS as an example (62 years in total), because the starting point, changepoint and end point are changing, its p -value is calculated $\sum_{n=1}^{60}\sum_{i=1}^{n} i$ =35990 times. After constructing the potential changepoints set, the number of calculation is reduced to $\sum_{n=1}^{29}\sum_{i=1}^{n} i$ =4060, and the efficiency is improved by 88.72%, and the calculation results are shown in Figure 5a. The changepoint trajectories (marked with red lines and blue dots) and alternative trajectories of 4-Series were determined according to the Detection Criteria in Section 2.3, as shown in Figure 5b-c.



Trajectory 1	1950	1998	2005	2011
		0.0027	0.0909	
Trajectory 2	1950	1998	2007	2011
		0.0071	0.1650	
Trajectory 3	1950	1999	2005	2011
		0.0070	0.1429	

Trajectory 1	1950	1998	2005	2011
		0.0042	0.3651	
Trajectory 2	1950	1999	2005	2011
		0.0099	0.8474	

Trajectory 1	1950	2000	2011	2016
		0.0095	0.7665	
Trajectory 2	1950	2002	2011	2016
		0.0060	0.8651	

Trajectory 1	1950	2003	2012	2016
		0.0079	0.7077	
Trajectory 2	1950	2005	2012	2016
		0.0084	0.7818	

Figure 5: Changepoint trajectory of Pingshan Station and Yichang Station (Confidence Level 99%)

For PAS, The starting point of the changepoint trajectory is 1950. We need to find the grid point with $p < 0.01$ in Figure 5a-1. Then, with the changepoint as the starting point and the ending point as the changepoint, find the grid point with $p < 0.01$ until 2011. At Confidence Level 99%, there are 3 points in Figure 5a-1 that meet the requirements of Criterion ①, namely 1950-1998-2005 (Trajectory 1), 1950-1998-2007 (Trajectory 2) and 1950-1999-2005 (Trajectory 3), and p is shown in Figure 5b. It can be seen that the Criterion ① can effectively narrow the selection range of changepoints from

many potential points. Criterion ② requires further search extending to 2011, which can fully explore the changepoint and ensure the continuity of the trajectory. When there are multiple alternative trajectories with inconsistent number of changepoints, Criterion ③ requires to select the one with the most points, which helps to divide the series in detail. Figure 5b~e shows all alternative trajectories that meet the requirements of the above 3 detection criteria. According to Criterion ④, select the year with small p of the first $M-1$ changepoints one by one, which can make the series before and after the changepoint have a large degree of change.

Based on the changepoint detection criteria, the year in which the series consistency has changed due to human factors (water storage of large reservoirs, etc.) can be determined (Figure 5b~e red line). The changepoint trajectory of PFSS is consistent with PAS, while YFSS lags behind YAS by one year. The reason could be related to the interannual variation of runoff. The flood season of Pingshan Station is from June to November, accounting for 81.34% of the annual average runoff. The upstream Ertan Reservoir (water storage in May 1998) has seasonal regulation capacity, so it can have a direct impact on PFSS, which is divided into 1950-1997, 1998-2004 and 2005-2011. However, the flood season of Yichang Station is from May to October, and the runoff in May accounts for 7.1% of the year. The annual mean runoff from 2001 to 2004 is 13154.73 m^3/s, 12454.25 m^3/s, 12991.84 m^3/s and 13115.10 m^3/s respectively. The monthly mean runoff in flood season from 2001 to 2004 is 20010.98 m^3/s, 18895.22 m^3/s, 20690.22 m^3/s and 19841.30 m^3/s respectively. For hydrological regime, 2002 is a year with less water inflow, while 2003 is the opposite. However, affected by the Three Gorges Reservoir, the water inflow in 2002 is closer to 2003-2010 in the flood season series, while the annual series is closer to 1950-2001. It indirectly shows that the changepoint detection framework proposed in this study considers the influence of both human factors and hydrological regime on the series. The HTS division results of Pingshan Station and Yichang Station are shown in Figure 5b~e. Dividing series helps ensure consistency of HTS and provides a basis for better information mining through statistical analysis methods.

4.3 Results of MWT Optimization

Based on the changepoint trajectories, the detection accuracy of the three methods was calculated, and the MWT optimization can be completed according to the optimization framework in Section 2.4. The screening process is shown in Table 3, and the optimization results of MWT are shown in Table 4.

Table 4: Changepoint and optimal MWT of Pingshan Station and Yichang Station (Confidence Level 99%)

Detection Method	Cumulative Anomaly		M-K Test		Wavelet Changepoint Detection		Optimal MWT
	Accuracy	Contribution1	Accuracy	Contribution1	Accuracy	Contribution[1]	
PAS	6.67%	48.39%	50%	6.45%	50%	61.29%	db8
PFSS	6.67%	50%	50%	6.67%	50%	60%	db8、fk8
YAS	6.25%	51.62%	0	0	50%	32.26%	db6
YFSS	5.56%	64.29%	0	0	50%	60.71%	fk14

Note 1: Contribution refers to the percentage of changepoints provided by the detection method for the potential changepoint set.

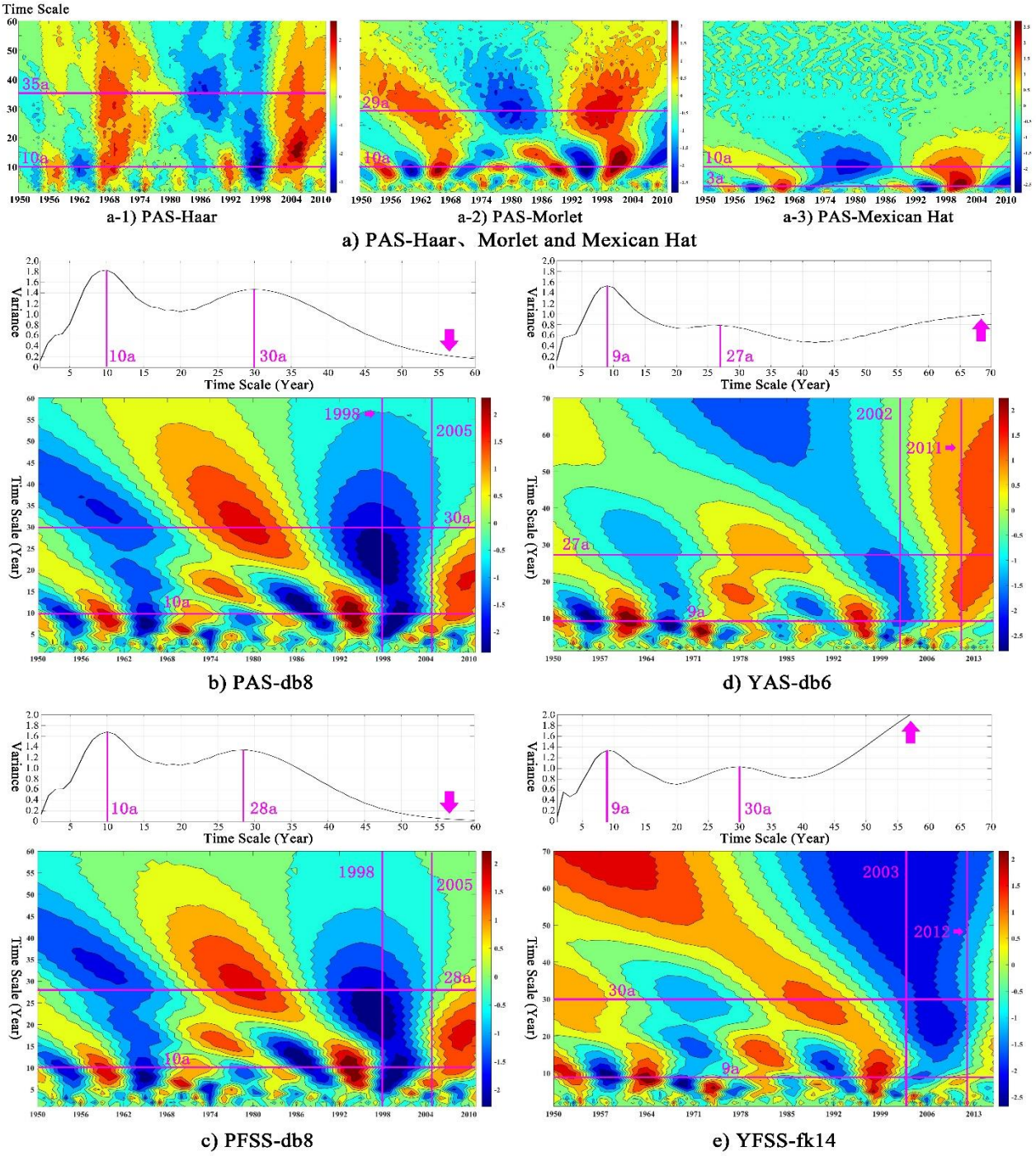
Combining the MWT optimization results in Table 3 and Table 4, it is found that the changepoint is the key to series division, and Optimization Step (1) can quickly locate the MWT that conforms to the series characteristics. For Pingshan Station, the

annual series of MWT meeting Optimization Step (1) is db8, and the flood season series are db8 and fk8. The Optimization Step (2) is selected according to the runoff physical cause at the same station, which makes it easier to analyse the evolution of the two series from the time-frequency space of the same MWT. Therefore, the optimal MWT of PFSS is db8.

When the optimal MWT of the series is determined, the accuracy of wavelet changepoint detection is generally higher than the cumulative anomaly method and the M-K test (Table 4). Except for YAS, the contribution rate of wavelet changepoint detection to the overall potential changepoint is also higher than both of them. The results show that the MWT optimization framework proposed in this study can accurately screen the optimal MWT of each series. The wavelet transform based on the MWT conforming to the series characteristics is helpful to improve the rationality of the analysis.

4.4 Analysis of HTS Evolution Based on CWT

Based on the optimization results of MWT in Table 4, the evolution of 4-Series was analysed by CWT. To further explore the influence of MWT, Haar, Morlet and Mexican Hat (referred to as 3 common wavelets) were used in CWT of PAS, as shown in Figure 6a. The analysis results of the optimal MWT are shown in Figure 6b~e.



Note: The pink reference line is the main period and the changepoint year of the series, which can be judged according to the marked content.

Figure 6: Results of CWT at Pingshan Station and Yichang Station (Wavelet Variance and Real Part Contour Map, Confidence Level 99%)

The 3 common wavelets have great differences in the analysis results of the main periods of PAS, namely 10a and 35a, 10a and 29a, and 3a and 10a (Figure 6a). Furthermore, they frequently alternate between wet and dry in the short time period, and exhibit a distinct "Wet-Dry-Wet" evolution over the long time period. Compared with Figure 6b, the CWT of 3 common wavelets is relatively scattered in the time scale of 0 to 60a, and the Morlet and Mexican Hat wavelets show a wet period after 1998, which does not reflect the regulation effect of the Ertan Reservoir on Pingshan Station, and the accuracy of the analysis results is questionable. According to historical records, during the flood season in June 1998, a basin wide flood occurred in the middle and lower Yangtze River due to continuous heavy rain in Dongting Lake and Panyang lake below Yichang Station (Zhang et al., 2021). From the time scale (Figure 6b-c), Pingshan Station and Yichang Station suffer continuous dry years, which is consistent with the actual situation. Based on the analysis of integrated moisture transport, land-falling atmospheric rivers geometric metrics and large-scale climatic circulations, Ayantobo et al. (2022) believed that the extreme rainfall in the Yangtze River Basin had a declining period after 1999, which was consistent with the analysis results of this study. We believe that optimizing the MWT that conform to series characteristics based on the changepoint detection is a suitable approach.

According to the analysis, the main periods of PAS are 10a and 30a, and the flood season series are 10a and 29a. The long-period scale of flood season is slightly earlier than the annual series, indicating that the annual adjustment of Pingshan Station has a certain buffer capacity. On the short-period scale 10a, the two series show the phenomenon of frequent alternation of wat and dry seasons, but the consecutive dry seasons from 1926 to 1968 and 1998 to 2004 have a serious impact on the series. Especially after 1998, due to the operation of Ertan Reservoir, the runoff reduction in the annual series is larger than that in flood season, so attention should be paid to the annual water demand of river channels and cities along the route. From 2005 to 2011, Pingshan Station had the wet season, and attention should be paid to flood control and flood resource utilization. The main periods of YAS are 9a and 27a, and the main periods of flood season series are 9a and 31a. Similarly, Yichang Station frequently alternates between wet and dry on the short-period scale. The annual series shows the evolution of "Wet-Dry-Wet-Dry-Wet" on the long-period scale, while the flood season series shows "Wet-Dry-Wet-Dry". After 2002-2003, YFSS did not enter the wet season as the annual series, indicating that the operation of the Three Gorges reservoir has a large reduction in the flood season. On the premise of ensuring the storage of the downstream reservoir at the end of the flood season, it is helpful to adjust the annual and interannual distribution of the runoff in the Yangtze River and improve the utilization efficiency of water resources.

5 Conclusion

Hydrological Time Series (HTS) is the basis of water conservancy project planning and construction. However, under the multiple effects of human activities and other factors, the consistency of HTS is destroyed. It is necessary to analyse its evolution to ensure the rationality of hydrological and hydraulic calculation. Wavelet transform is one of the widely used analysis tools of evolution in hydrology, but the its analysis accuracy is closely related to Mother Wavelet (MWT). To solve

these two problems, with the help of cumulative anomaly method, Mann-Kendall (M-K) test and wavelet changepoint detection, we proposed the changepoint detection criteria and a MWT optimization framework in this study, and took Pingshan Station and Yichang Station of the Yangtze River as study cases to test their effectiveness. The main conclusions are as follows:

(1) Changepoint detection criteria: Based on the three changepoint detection methods, a potential changepoints set of HTS is constructed, which can make up for the limitations of a single method affected by factors such as parameter settings and marginal effects, and improve the calculation efficiency. In addition, with the help of Kolmogorov-Smirnov (K-S) test, we proposed the detection criteria to quickly confirm the changepoint trajectory from the beginning to the end of HTS. While ensuring the uniqueness of the result, the changepoint formed by the combined action of multiple factors can be accurately identified to complete the series division.

(2) MWT optimization framework: Based on the changepoint detection accuracy of wavelet changepoint detection, the MWT consistent with the series characteristics can be selected to ensure the accuracy of wavelet transform to analyse the HTS evolution and provide a good basis for hydrological and hydraulic calculation.

It is found that the changepoints of the Pingshan annual series and the Pingshan flood season series both are 1998 and 2005, the Yichang annual series are 2002 and 2011, and the Yichang flood season series are 2003 and 2012. In addition, the optimal MWT of 4-Series are db8, db8, db6 and fk8 respectively. The Ertan reservoir has a greater impact on the annual runoff of Pingshan Station, while the Three Gorges Reservoir only reduces the runoff of the Yichang Station to a large extent during the flood season. Limited by the data, we did not explore the evolution of the two stations after 2017. It is also found that the wavelet changepoint detection is not sufficient enough to detect the potential changepoint of 10 years before and after the series.

Acknowledgements

The authors would like to give special thanks to the anonymous reviewers.

Acronym List

Order	Acronym	Full name	Order	Acronym	Full name
1	HTS	Hydrological Time Series	7	DWT	Discrete Wavelet Transform
2	MWT	Mother Wavelet	8	MODWT	Maximal Overlap Discrete Wavelet Transform
3	IID	Independent and identically distributed	9	PAS	Pingshan Annual Series
4	K-S	Kolmogorov-Smirnov	10	PFSS	Pingshan Flood Season Series
5	M-K	Mann-Kendall	11	YAS	Yichang Annual Series
6	CWT	Continuous Wavelet Transform	12	YFSS	Yichang Flood Season Series

Authors' Contributions

Jiqing Li: Conceptualization, Validation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition

Jing Huang: Conceptualization, Methodology, Software, Formal analysis, Resources, Writing - Original Draft, Visualization

Lei Zheng: Methodology, Software, Formal analysis, Data Curation

Wei Zheng: Software, Validation, Investigation, Visualization

Funding

This study is financially supported by the National Natural Science Foundation of China (No. 52179014) and the National Key R&D Program of China (2022YFC3002702-4).

Data Availability Statement

Data for this study can be downloaded from the Yangtze River Hydrological Network (http://www.cjh.com.cn/). In this study, the wavelet changepoint detection is based on the Matlab (R2020b) toolbox, and the rest of the codes (PyCharm 2021.2.2) are available from the corresponding author upon reasonable request.

Compliance with Ethical Standard

Declaration The authors confirm that this article is original research and has not been published or presented previously in any journal or conference.

Conflict of Interest None.

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

References

Ayantobo, O. O., Wei J. and Wang G.: Climatology of landfalling atmospheric rivers and its attribution to extreme precipitation events over Yangtze River Basin, J. Atmos Res, 270, 106077, doi:10.1016/j.atmosres.2022.106077, 2022.

Benhassine, N. E., Boukaache, A. and Boudjehem, D.: Medical image denoising using optimal thresholding of wavelet coefficients with selection of the best decomposition level and mother wavelet, J. Int J Imag Syst Tech, 31, 1906-20, doi:10.1002/ima.22589, 2021.

Chen, Y., Paschalis, A., Wang, L. and Onof, C.: Can we estimate flood frequency with point-process spatial-temporal rainfall models?, J. J Hydrol, 600, 126667, doi:10.1016/j.jhydrol.2021.126667, 2021.

Corradin, R., Danese, L. and Ongaro, A.: Bayesian nonparametric change point detection for multivariate time series with missing observations, J. Int J Approx Reason, 143, 26-43, doi:10.1016/j.ijar.2021.12.019, 2022.

Dang, C., Zhang, H., Singh, V. P., Zhi, T., Zhang, J. and Ding, H.: A statistical approach for reconstructing natural streamflow series based on streamflow variation identification, J. Hydrol Res 52, 1100-15, doi:10.2166/nh.2021.180, 2021.

Fang, L. and Shao, D.: Application of Long Short-Term Memory (LSTM) on the Prediction of Rainfall-Runoff in Karst Area, J. Frontiers in Physics, 9, doi:10.3389/fphy.2021.790687, 2022.

Jia, B., Zhou, J., Tang, Z., Xu, Z., Chen, X. and Fang, W.: Effective stochastic streamflow simulation method based on Gaussian mixture model, J. J Hydrol, 605, 127366, doi:10.1016/j.jhydrol.2021.127366, 2022.

Li, J., Huang, J., Chu, X. and Lund, J. R.: An Improved Peaks-Over-Threshold Method and its Application in the Time-Varying Design Flood, J. Water Resour Manag, 35, 933-48, doi:10.1007/s11269-020-02758-3, 2021.

Liu, W., Wen, J., Chen, J., Wang, Z., Lu, X., Wu, Y., et al.: Characteristic analysis of the spatio-temporal distribution of key variables of the soil freeze-thaw processes over the Qinghai-Tibetan Plateau, J. Cold Reg Sci Technol, 197, 103526, doi:10.1016/j.coldregions.2022.103526, 2022.

Malki, A., Atlam, E. and Gad, I.: Machine learning approach of detecting anomalies and forecasting time-series of IoT devices, J. Alexandria Engineering Journal, 61, 8973-86, doi:10.1016/j.aej.2022.02.038, 2022.

Mat Jan, N. A., Shabri, A. and Samsudin, R.: Handling non-stationary flood frequency analysis using TL-moments approach for estimation parameter, J. J Water Clim Change, 11, 966-79, doi:10.2166/wcc.2019.055, 2020.

Moradi, M.: Wavelet transform approach for denoising and decomposition of satellite-derived ocean color time-series: Selection of optimal mother wavelet, J. Adv Space Res, 69, 2724-44, doi:10.1016/j.asr.2022.01.023, 2022.

Nielsen, M.: On the Construction and Frequency Localization of Finite Orthogonal Quadrature Filters, J. J Approx Theory, 108, 36-52, doi:10.1006/jath.2000.3514, 2001.

Oliveira-Júnior, J. F. D., Correia, F. W., Monteiro, L. D. S., Shah, M., Hafeez, A., Gois, G. D., et al. Urban rainfall in the Capitals of Brazil: Variability, trend, and wavelet analysis, J. Atmos Res, 267, 105984, doi:10.1016/j.atmosres.2021.105984, 2022.

Qin, Y., Sun, X., Li, B. and Merz, B.: A nonlinear hybrid model to assess the impacts of climate variability and human activities on runoff at different time scales, J. Stoch Env Res Risk A, 35, 1917-29, doi:10.1007/s00477-021-01984-4, 2021.

Sanaa, H., Gab, A., Ukkola, A. M., Martin, D. K., Andy, P., et al.: Reconciling historical changes in the hydrological cycle over land, J. NPJ climate and atmospheric science, 5, 1-9. doi:10.1038/s41612-022-00240-y, 2022.

Şen, Z.: Jump point identification in hydro-meteorological time series by crossing methodology, J. Theor Appl Climatol 144, 769-77, doi:10.1007/s00704-021-03576-2, 2021.

Shi, X., Gallagher, C., Lund, R. and Killick, R.: A comparison of single and multiple changepoint techniques for time series data, J. Comput Stat Data An, 170, 107433, doi:10.1016/j.csda.2022.107433, 2022.

Wang, S. H., Su, B. R., Wang, Y. Q., Wang, Y. J., Zhu, J. Q. and Fu, J.: Change analysis of runoff and sediment in the Three Gorges Reservoir Region in recent 16 years, J. Science of Soil and Water Conservation, 19, 69-78, doi:10.16843/j.sswc.2021.01.009, 2021 (In Chinese)

Stasolla, M. and Neyt, X.: Enhanced Morphological Filtering for Wavelet-Based Changepoint Detection, J. IEEE, 56-60, doi:10.1109/SITIS.2019.00021, 2019.

Strömbergsson, D., Marklund, P., Berglund, K., Saari, J. and Thomson, A.: Mother wavelet selection in the discrete wavelet transform for condition monitoring of wind turbine drivetrain bearings, J. Wind Energy, 22, 1581-92, doi:10.1002/we.2390, 2019.

Xie, Y., Liu, S., Huang, S., Fang, H., Ding, M., Huang, C., et al.: Local trend analysis method of hydrological time series based on piecewise linear representation and hypothesis test, J. J Clean Prod, 339, 130695, doi:10.1016/j.jclepro.2022.130695, 2022.

Zerouali, B., Chettih, M., Abda, Z., Mesbah, M., Santos, C. A. G., Brasil, N. R. M.: A new regionalization of rainfall patterns based on wavelet transform information and hierarchical cluster analysis in northeastern Algeria, J. Theor Appl Climatol, 147, 1489-510, doi:10.1007/s00704-021-03883-8, 2022.

Zhang, Y., Fang, G., Tang, Z., Wen, X., Zhang, H., Ding, Z., et al.: Changes in Flood Regime of the Upper Yangtze River, J. Frontiers in Earth Science, 9, doi:10.3389/feart.2021.650882, 2021.

Zhao, Y. H., Yu, B. K., Qu, P., Li, S., Zhan, D. Q. and Wang, X. Q.: Analysis of runoff variation characteristics in Yishuhe River Basin. IOP conference series, J. Earth and environmental science, 344, 12080, doi:10.1088/1755-1315/344/1/012080, 2019.