

## Research Article

# Study of the Flood Frequency Based on Normal Transformation in Arid Inland Region: A Case Study of Manas River in North-Western China

Changlu Qiao <sup>1,2</sup>, Guotao Cai <sup>1,3</sup>, Yanxue Liu <sup>1,2</sup>, Junfeng Li <sup>1,2</sup> and Fulong Chen <sup>1,2</sup>

<sup>1</sup>College of Water Conservancy & Architectural Engineering, Shihezi University, Shihezi 832003, Xinjiang, China

<sup>2</sup>Key Laboratory of Modern Water-Saving Irrigation of Xinjiang Production & Construction Group, Shihezi 832000, Xinjiang, China

<sup>3</sup>Xinjiang Hami Pumped Storage Power Company Limited, Hami 839000, Xinjiang, China

Correspondence should be addressed to Guotao Cai; [caiguotao@stu.shzu.edu.cn](mailto:caiguotao@stu.shzu.edu.cn) and Yanxue Liu; [liuyanxue@stu.shzu.edu.cn](mailto:liuyanxue@stu.shzu.edu.cn)

Received 20 May 2022; Revised 19 June 2022; Accepted 22 June 2022; Published 13 July 2022

Academic Editor: Xingsi Xue

Copyright © 2022 Changlu Qiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Flood disaster is one of the natural disasters which cause the most serious economic losses, the most casualties, and the greatest social impact. Flood frequency analysis is very important for reducing flood disaster. In this paper, based on the flood data of Manas River and tools of Box–Cox and Johnson normal transformation, the nonparametric statistical method for flood frequency analysis is studied in order to analyze the adaptability between it and the rivers in arid region of north-western China. The calculation result of the fitness index is divided into two parts: high flood discharge and low flood discharge. One of the two evaluation indexes has an advantage in fitting, and the number of advantages of the three methods in each part has been counted. After analysis, for the flood peak discharge frequency of rivers in arid region of north-western China, the frequency curve of Johnson transformation fits best with empirical data. The high flood discharge advantage is 6, and the low flood discharge is 4. For the flood volume frequency of rivers in arid region of north-western China, Box–Cox transform fits well with empirical data at the high flood discharge frequency curve, and its advantage is 12; Johnson transformation has a better fit between the low flood discharge frequency curve and empirical data, and its advantage is 12. Therefore, it is the way of improving the precision of flood frequency analysis to use the method of P-III distribution and normal transformation comprehensively.

## 1. Introduction

In recent years, due to the impact of global climate change and human activities, hydrologic extreme disasters had increased greatly in most countries and regions, resulting in economic, ecological, and even life and property heavy losses [1]. As a hydrologic extreme event, flood disaster is one of the natural disasters which cause the most serious economic losses, the most casualties, and the greatest social impact [2]. As one of the main methods to accurately estimate the design value of hydrologic variables, flood frequency analysis is very important for reducing flood disaster. At present, flood frequency analysis has been widely used in the field of hydrologic design [3]. The current flood frequency analysis methods in China are mainly divided into two categories, the parametric

statistical method and the nonparametric statistical method. The parameter statistical method is based on the presumed flood frequency distribution form, and the parameters of the population distribution are obtained through the samples, and the design value under the specified frequency is obtained from the population distribution [4]. In the case of reasonable distribution form assumption, the parameter statistical method can obtain more information from the samples and the calculation results are more accurate. However, if it is unreasonable, the calculation accuracy will be reduced. The nonparametric statistical method avoids the error caused by the assumed distribution form and directly calculates the design value of the specified frequency according to the measured flood samples. The nonparametric statistical method is more flexible and robust than the parametric

statistical method. It is also a research hotspot at present and provides another research way for hydrologic frequency analysis [5–10]. The normal transformation transforms the original skewed distribution of the sample into the normal distribution and then uses its inverse transformation to calculate the design value of hydrologic variables (normal quantile) under the specified frequency. This process does not involve the content of parameter calculation method and takes normal distribution as the intermediate medium. It belongs to the nonparametric statistical method in theory [11–13]. Relevant researches have shown that, after the normal transformation of single variable, the original skewed distribution to the normal distribution is a one-to-one monotonic increasing relationship, and the serry obtained by using the normal transformation can retain the sample information of the original skew distribution more completely [14, 15]. The research of Chen and Song [16] also pointed out that there is a good fitting effect between the design value obtained by the normal transformation and the measured serry. So, the normal transformation can be used in hydrologic frequency analysis.

The application of the normal transformation in the calculation of flood frequency in arid region is less. Johnson transformation is mostly used in quality management statistics [17] and processing nonnormal statistics problems [18]. Box–Cox transformation is often used to improve the skewness and heteroscedasticity of linear regression, and it is more suitable for hydrology than Johnson transformation, such as the research of Liang and Dai [11] and Li et al. [19] and others. As a typical river in arid region of north-western China, hydrologic extreme events often occur at Manasi River. In this paper, based on the flood data of Manas River and tools of Box–Cox and Johnson normal transformation, the nonparametric statistical method for flood frequency analysis is studied in order to analyze the adaptability between it and the rivers in arid region of north-western China.

## 2. Materials and Methods

**2.1. Box–Cox Transformation.** Box–Cox transformation is a normal transformation model proposed by Box and Cox in 1964. The model is [20–25]

$$Y = \begin{cases} \frac{X^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log X, & \lambda = 0. \end{cases} \quad (1)$$

The inverse transformation of the model is

$$X = \begin{cases} (1 + \lambda Y)^{1/\lambda}, & \lambda \neq 0, \\ e^Y, & \lambda = 0, \end{cases} \quad (2)$$

where  $X = \{x_1, x_2, \dots, x_n\}$  is the original serry to be transformed;  $Y = \{y_1, y_2, \dots, y_n\}$  is the output serry after transformation;  $\lambda$  is the transformation parameter,  $\lambda \in [-5, +5]$ . The method to determine the best  $\lambda$  is when  $\lambda \in [-5, +5]$ , the optimal value of  $\lambda$  is the one where the minimum standard deviation of the  $Z$  serry defined by equation (3) is obtained

[26]. When  $\lambda = 0$ , the transformation is logarithmic transformation,  $\lambda = -1$  is the reciprocal transformation, and  $\lambda = 0.5$  is the square root transformation.

$$Z = \begin{cases} \frac{X^\lambda - 1}{\lambda g^{\lambda-1}}, & \lambda \neq 0, \\ g \cdot \ln(X), & \lambda = 0, \end{cases} \quad (3)$$

where  $g$  is the geometric mean of the original serry,  $X$  is the original data serry to be transformed, and  $Z$  is the output serry after transformation.

Box–Cox normal transformation requires that each item of the serry to be transformed is greater than 0. That is  $x_i > 0$ . Each item of hydrologic serry is greater than 0, so it meets the transformation requirements.

**2.2. Johnson Transformation.** Johnson transformation is a normal transformation model based on three distribution curves proposed by Johnson in 1949. The model is shown in Table 1 [27–30].

Chou et al. [32–34].  $X = \{x_1, x_2, \dots, x_n\}$  is the original serry to be transformed;  $Y = \{y_1, y_2, \dots, y_n\}$  is the output serry after transformation;  $\epsilon$  and  $\gamma$  are position control parameters;  $\lambda$  and  $\eta$  are scale parameters and are generally positive. In the Johnson normal transformation, the calculation of the parameters to be estimated is based on the method proposed by Hill et al. [31] and Chou et al. [32–34].

**2.3. Study Area.** The Manas River is located in the northern foot of the middle section of Tianshan and on the southern edge of the Junggar Basin. It is the largest river on the northern slope of the Tianshan. It originates from the Erenhabirga Mountains on the northern slope of the Tianshan. It is about 324 km in length, and the drainage area is about 5156 km<sup>2</sup> [35]. Kenswat Hydrologic Station is mid-stream of Manas River, which was built in 1955. The station controls the flow of Manas River (Figure 1). The hydrologic data have been compiled and reviewed by the Hydrology and Water Resources Bureau with reliable accuracy [36].

**2.4. Data Acquisition and Processing.** In this paper, the measured flood data of Kenswat Hydrologic Station with the longest observation time of Manas River are used as the measured serry. Kenswat reservoir is one kilometer upstream of Kenswat Hydrological Station. The construction of the reservoir started on August 7, 2009, and officially began to impound on December 6, 2014. Since the impoundment of the reservoir, the consistency of hydrologic data of Kenswat Hydrologic Station had been destroyed. Therefore, the data period selected in this paper is from 1955 to 2014.

According to the data analysis of the station, the average annual flood peak discharge of the Manas River is 356 m<sup>3</sup>/s. The measured maximum flood peak discharge is 1095 m<sup>3</sup>/s (August 2, 1999), the second is 758 m<sup>3</sup>/s (July 28, 1966), and the third is 735 m<sup>3</sup>/s (July 18, 1996). The annual maximum sampling method is used to select flood peak discharge

TABLE 1: Johnson transformation model.

| Type | Transformation model  | Inverse transformation model   | Parameter constraints  | X constraints                             |
|------|---|--|--|---|
| SB   | $Y = \gamma + \eta \ln(X - \varepsilon/\lambda + \varepsilon - X)$  | $X = \varepsilon + (\lambda + \varepsilon)e^{Y/\eta}/1 + e^{Y/\eta}$ | $\eta, \lambda > 0$<br>$-\infty < \gamma < +\infty$<br>$-\infty < \varepsilon < +\infty$<br>$\eta > 0$ | $\varepsilon < X < \varepsilon + \lambda$ |
| SL   | $Y = \gamma + \eta \ln(X - \varepsilon)$                            | $X = e^{Y/\eta} + \varepsilon$                                       | $-\infty < \gamma < +\infty$<br>$-\infty < \varepsilon < +\infty$<br>$\eta, \lambda > 0$               | $X > \varepsilon$                         |
| SU   | $Y = \gamma + \eta \operatorname{arcsinh}(X - \varepsilon/\lambda)$ | $X = \lambda \sinh(Y - \gamma/\eta) + \varepsilon$                   | $-\infty < \gamma < +\infty$<br>$-\infty < \varepsilon < +\infty$                                      | $-\infty < X < +\infty$                   |
| Note |   | $\operatorname{arcsinh} x = \ln[x + (x^2 + 1)^{0.5}]$                |  |   |

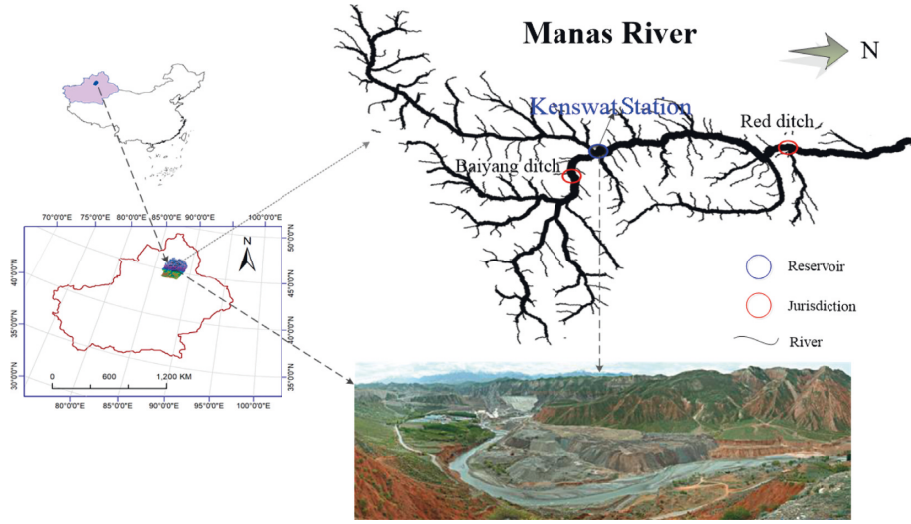


FIGURE 1: Locations of Manas river and Ken Swat station.

samples, the unified sample method is adopted, and the Weibull equation is used to determine the empirical frequency. Because the Manas River is a small river, when selecting flood volume samples, 1 d, 3 d, and 5 d are used as the flood volume calculation period.

### 3. Results and Discussions

#### 3.1. Detection and Correction of the Mutation

##### 3.1.1. Detection of the Mutation

(1) *Flood Peak Discharge Serry*. In order to ensure the accuracy of detecting the mutation, three methods are selected to detect the mutation at the same time, and the mutation point is finally determined by comprehensive analysis.

According to calculation and analysis, the skipping mutation point of Lee–Heghinan test is 1995, and the skipping mutation point of ordered clustering test and sliding  $T$  test is 1993. Combining other literature and test results, the flood peak discharge serry mutation point is determined to be 1993, and so the serry is divided into two subseries from the mutation point. The subseries before the mutation point shows a downward trend, and its average value is  $334.08 \text{ m}^3/\text{s}$ . The subseries after the mutation point also shows a downward trend, and its average value is  $442.87 \text{ m}^3/\text{s}$ . The average value of the two subseries is quite

different, showing skipping change. According to calculation and analysis, the main mutation type of flood peak discharge serry is identified as skipping mutation, and the results of detecting the mutation are shown in Figures 2(a) and 2(b).

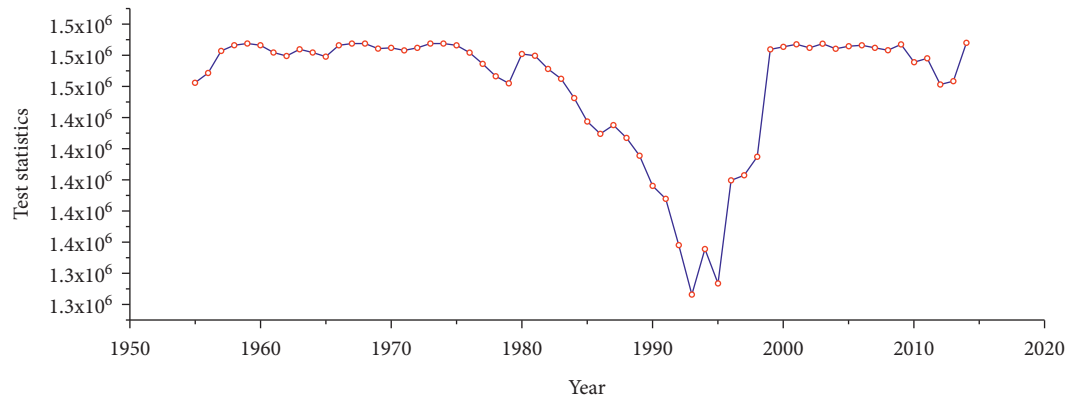
(2) *Flood Volume Serry*. The method of detecting the mutation of the flood volume serry is the same as that of the flood peak discharge. The test results are shown in Figures 2(c)–2(h).

It can be seen from Figures 2(c)–2(h) that the maximum 1 d, 3 d, and 5 d flood volume serry mutation points are all in 1993. It is consistent with the test results of flood peak serry. So, it can be determined that the main mutation type of Manas River flood serry is skipping mutation and the mutation point is in 1993.

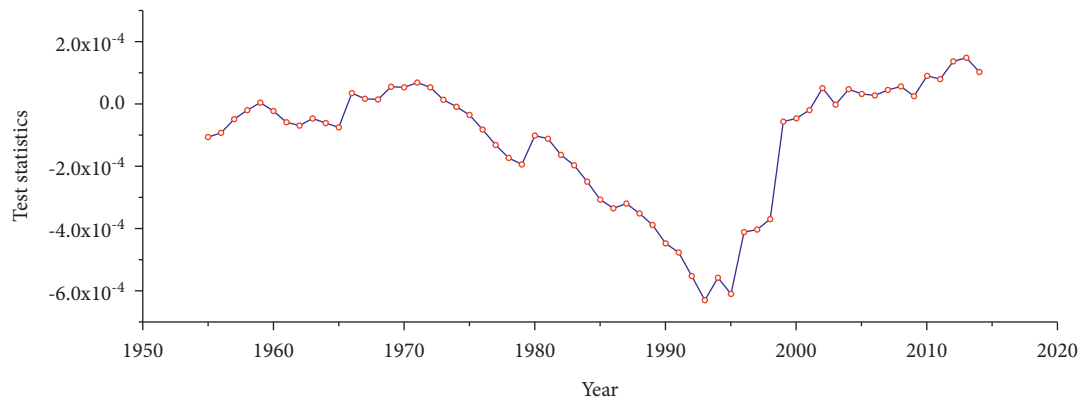
##### 3.1.2. Correction of the Mutation

(1) *Flood Peak Discharge Serry*. In this paper, the decomposition synthesis theory proposed by Xie Ping is adopted as the method of correcting the skipping mutation, and the calculation process is based on the previous researches [37–39].

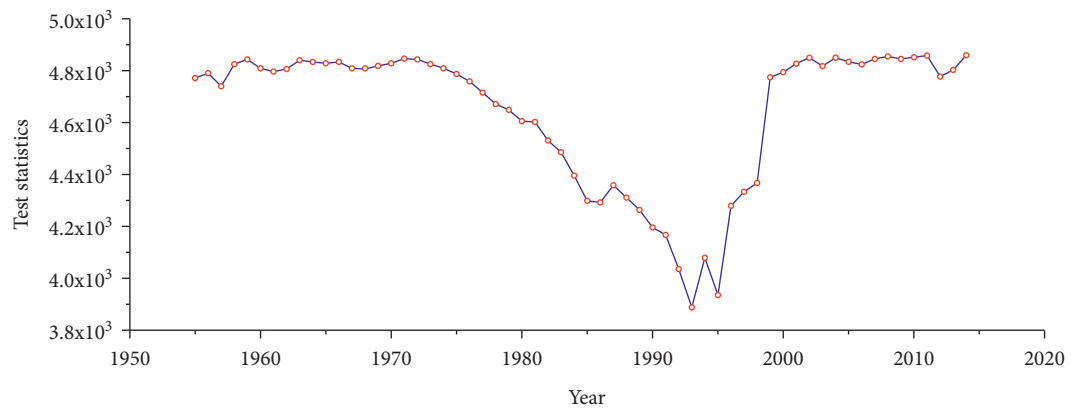
According to the decomposition synthesis theory proposed by Xie Ping, the equation of correcting the skipping mutation is



(a)

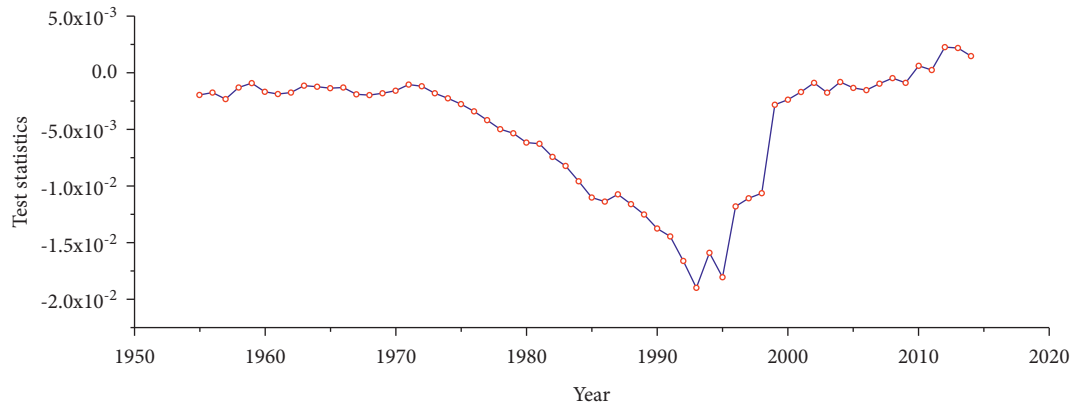


(b)

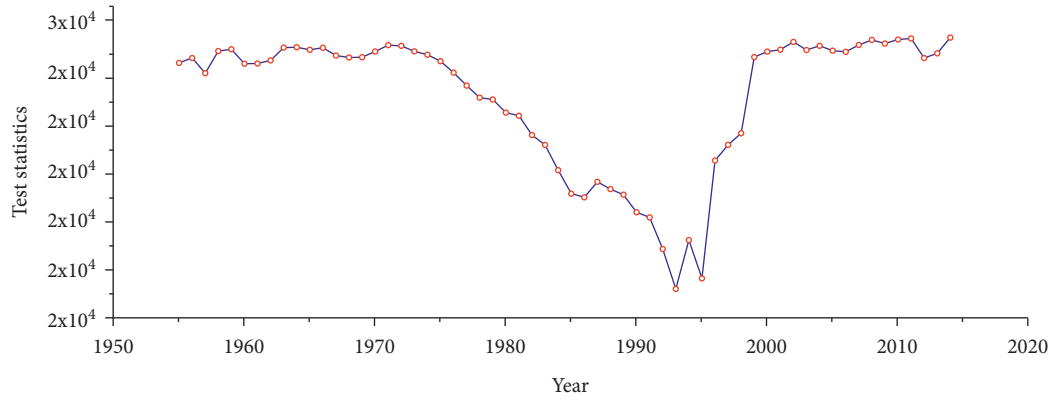


(c)

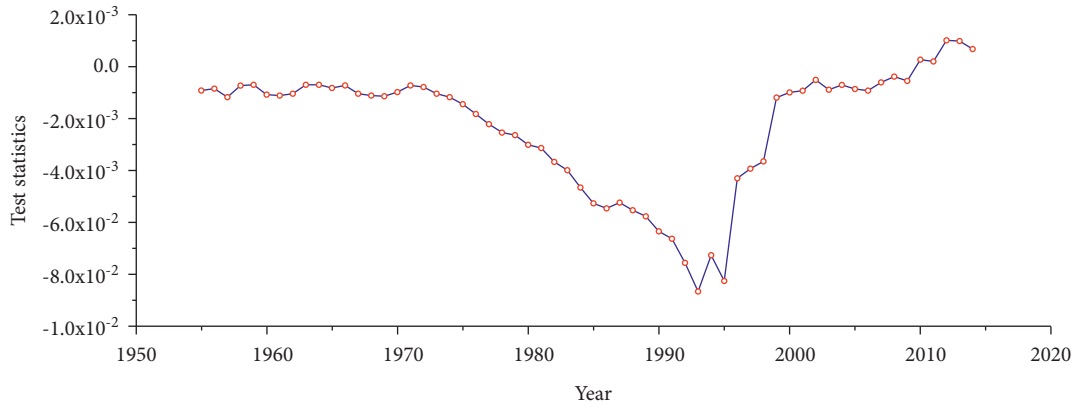
FIGURE 2: Continued.



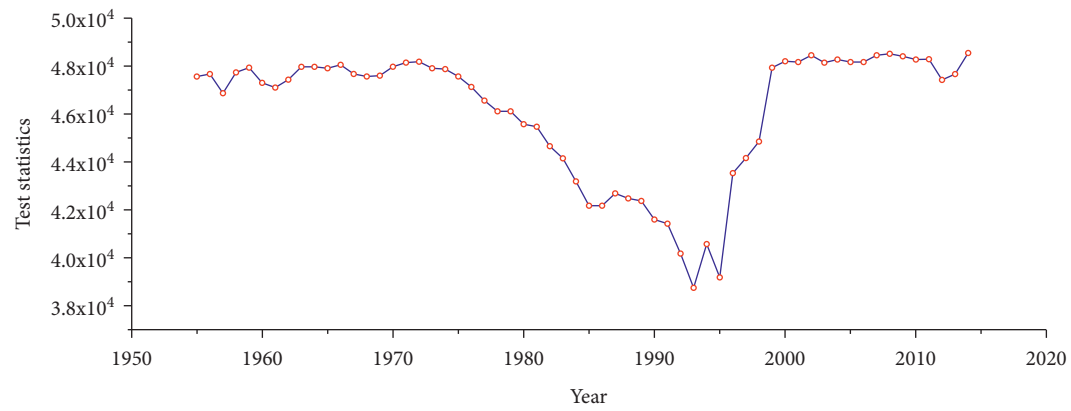
(d)



(e)



(f)



(g)

FIGURE 2: Continued.

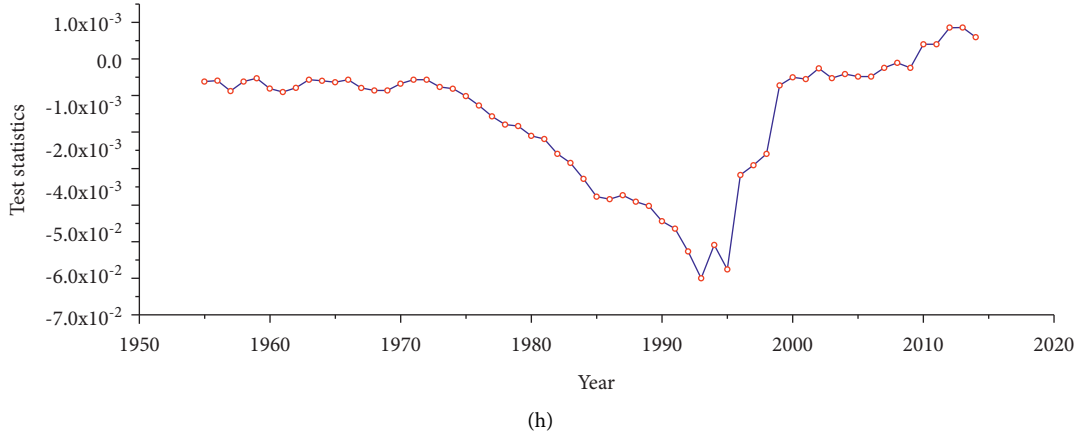


FIGURE 2: Results of detecting the mutation of the flood serry. (a) and (b) The results of ordered cluster test and sliding  $T$  test of the annual maximum flood peak discharge serry, (c) and (d) the results of the ordered clustering test and the sliding  $T$  test of the maximum 1 d flood volume, (e) and (f) the results of ordered clustering test and sliding  $T$  test with the maximum 3 d volume, and (g) and (h) the results of ordered clustering test and sliding  $T$  test with the maximum 5 d volume.

$$x'_k = x_k - 108.79, \quad (k = j + 1, j + 2, \dots, n), \quad (4)$$

where  $x'_k$  is the item of the serry after correcting the mutation;  $x_k$  is the item of the serry before correcting the mutation; and  $j$  is the year of mutation (here is 1993).

Based on the subseries before the mutation point, the subseries after the mutation point is corrected by equation (4). The corrected synthetic serry is detected again, and no mutation point is found, so the corrected serry meets the premise of consistency. The corrected flood peak discharge serry is shown in Figure 3(a).

(2) *Flood Volume Serry*. The method of correcting the mutation of the flood volume serry is the same as that of the flood peak discharge. The corrected results are shown in Figures 3(b)–3(d).

### 3.2. Normal Detection and Transformation

#### 3.2.1. Normal Detection

(1) *Flood Peak Discharge Serry*. The normal detection is carried out on the flood peak discharge serry after correcting the mutation. The detection methods are the nonparametric Shapiro–Wilk test ( $W$  test) and normal P–P diagram method. The normal P–P test results are shown in Figure 4(a), and the  $W$  test results are shown in Table 2.

According to Figure 4(a) and Table 2, the normality of the flood peak discharge serry is not significant, so the normal transformation cannot be directly used to calculate the flood peak discharge frequency. Therefore, the skewed serry is transformed into the normal serry by the methods of Box–Cox transformation and Johnson transformation, and the normal test results of the transformed flood peak discharge serry are shown in Figures 5(a) and 5(b).

It can be seen from Figures 5(a) and 5(b) and the results of normal transformation for the corrected flood peak

discharge serry that the effect of the Box–Cox normal transformation is poor, and the effect of Johnson normal transformation is better.

(2) *Flood Volume Serry*. For the normal test of the flood volume serry, the method is the same as that of the flood peak discharge.

#### (I) Maximum 1 d flood volume serry

The normal test is performed on the maximum 1 d flood volume after correcting the mutation, and the test results are shown in Figure 4(b) and Table 2. The normal test results of the maximum 1 d flood volume serry after normal transformation are shown in Figure 5(c) and 5(d).

#### (II) Maximum 3 d flood volume serry

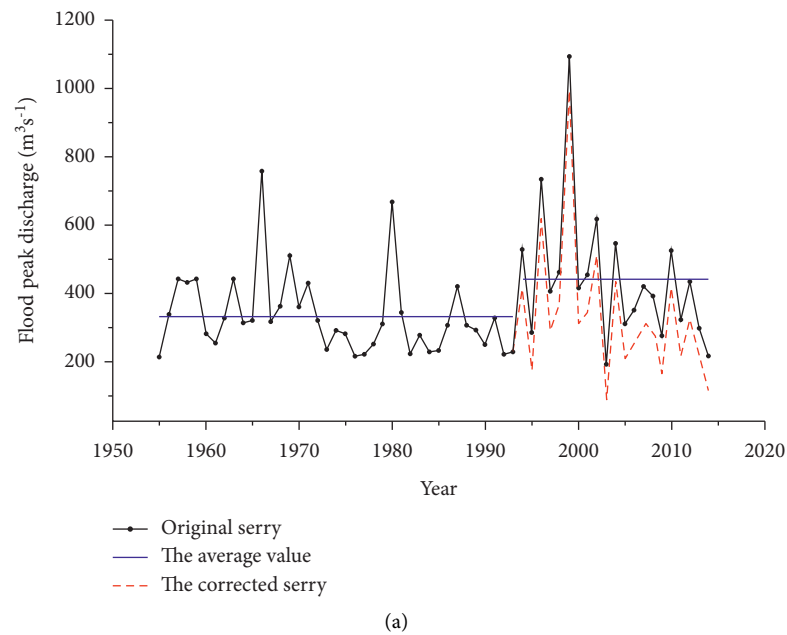
The normal test is performed on the maximum 3 d flood volume after correcting the mutation, and the test results are shown in Figure 4(c) and Table 2. The normal test results of the maximum 3 d flood volume serry after normal transformation are shown in Figure 5(e) and 5(f).

#### (III) Maximum 5 d flood volume serry

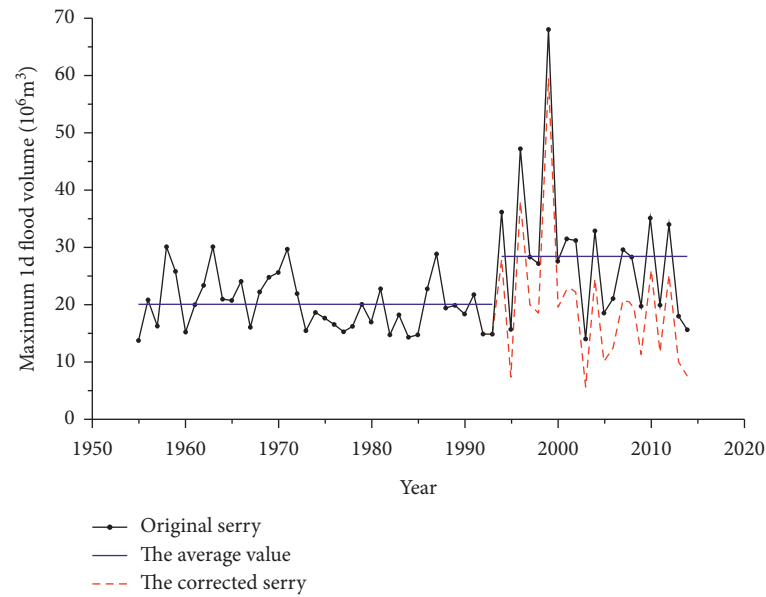
The normal test is performed on the maximum 5 d flood volume after correcting the mutation, and the test results are shown in Figure 4(d) and Table 2. The normal test results of the maximum 5 d flood volume serry after normal transformation are shown in Figures 5(g) and 5(h).

#### 3.2.2. Normal Transformation

(1) *Flood Peak Discharge Serry*. After Box–Cox normal transformation for the corrected flood peak discharge serry, the new serry passed the normal test. The  $P$ -value of 95% confidence interval of  $W$  test is equal to 0.06. The mean value of the new serry is 13.78, and the standard deviation is 1.998.



(a)



(b)

FIGURE 3: Continued.

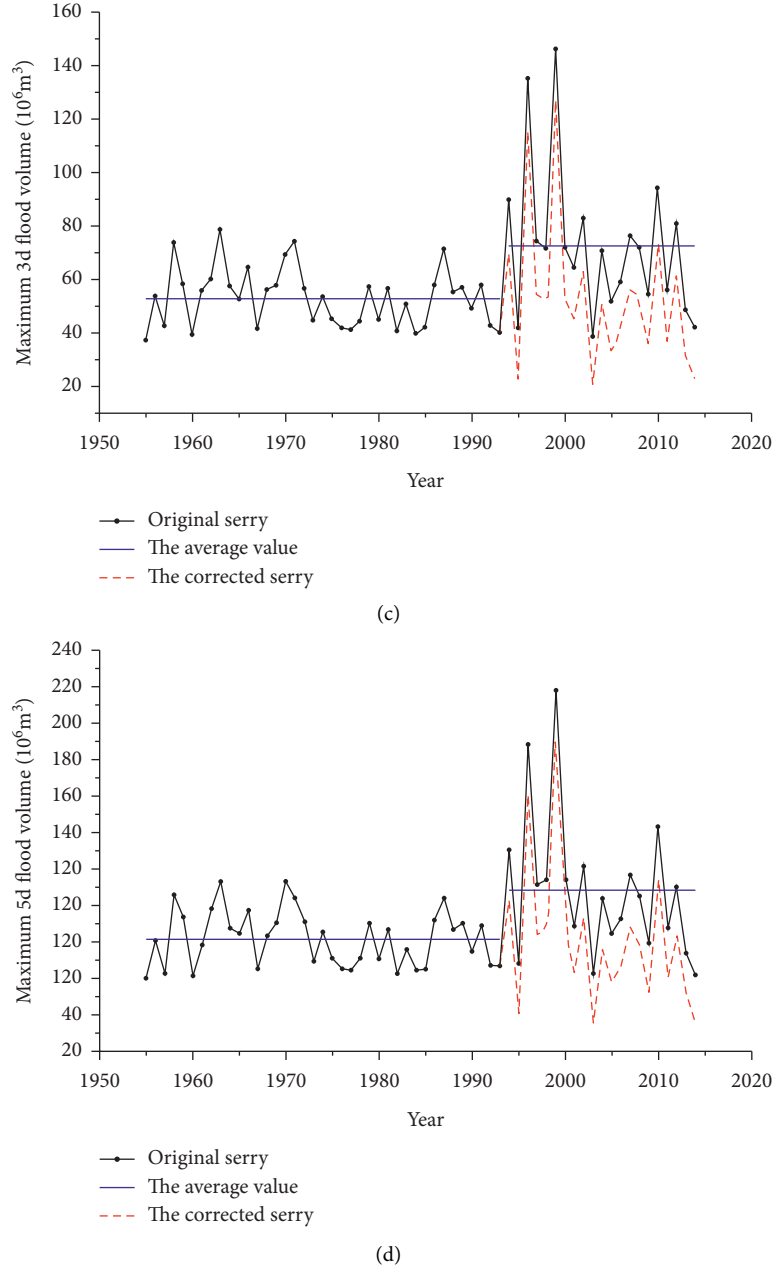


FIGURE 3: Results of correcting the mutation of flood serry. (a) The correction result of the annual maximum flood peak discharge mutation, (b) the correction result of the maximum 1 d flood volume mutation, (c) the correction result of the maximum 3 d flood volume mutation, and (d) the correction result of the maximum 5 d flood volume mutation.

By comparing  $S_B$ ,  $S_L$ , and  $S_U$ , the three types of Johnson normal transformation for the corrected flood peak discharge serry, we found that  $S_U$  transformation is the best. After  $S_U$  type of Johnson normal transformation for the corrected flood peak discharge serry, the new serry passed the normal test. The  $P$ -value of 95% confidence interval of  $W$  test is equal to 0.87. The mean value of the new serry is 0.06259, and the standard deviation is 0.9161. The parameters of the  $S_U$ -type Johnson normal transformation for the corrected flood peak serry are as follows:

$\gamma = 0.831159$ ,  $\eta = 1.12611$ ,  $\varepsilon = 221.007$ ,  $\lambda = 92.8704$ . So, the optimal equation of  $S_U$ -type Johnson normal transformation is

$$Y = -0.831159 + 1.12611 \operatorname{arc} \sinh \left( \frac{X - 221.007}{92.8704} \right). \quad (5)$$

The best inverse normal transformation is

$$X = 92.8704 \sinh \frac{Y + 0.831159}{1.12611} + 221.007. \quad (6)$$



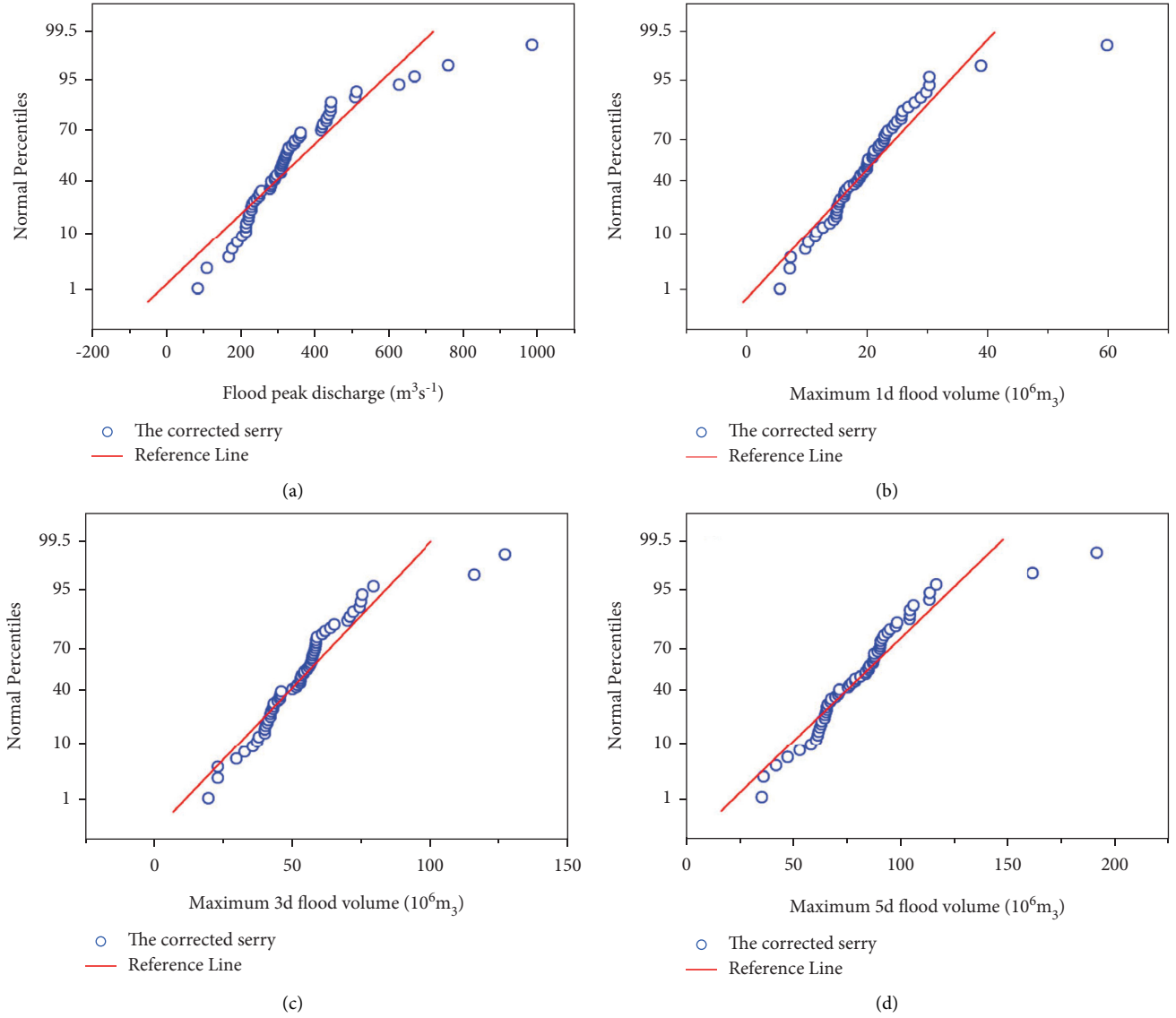


FIGURE 4: Normal P-P test results of the flood serry. (a) The normal test chart of the annual maximum flood peak discharge after correcting the mutation, (b) the normal test chart of the maximum 1 d flood volume after correcting the mutation, (c) the normal test chart of the maximum 3 d flood volume after correcting the mutation, and (d) the normal test chart of the maximum 5 d flood volume after correcting the mutation.

TABLE 2: *W* test results of flood.

| Test variable            | DF | Statistic | <i>P</i> -value       | Decision at level (5%) |
|--------------------------|----|-----------|-----------------------|------------------------|
| Flood peak discharge     | 60 | 0.84      | $1.76 \times 10^{-6}$ | Reject normality       |
| Maximum 1 d flood volume | 60 | 0.86      | $6.17 \times 10^{-6}$ | Reject normality       |
| Maximum 3 d flood volume | 60 | 0.87      | $1.58 \times 10^{-5}$ | Reject normality       |
| Maximum 5 d flood volume | 60 | 0.88      | $2.2 \times 10^{-5}$  | Reject normality       |

## (2) Flood Volume Serry

### (I) Maximum 1 d flood volume serry

The normal transformation results of maximum 1 d flood volume serry are that the best transformation parameter of  $\lambda$  is equal to 0 for Box-Cox normal transformation and the best normal transformation

equations are as follows for Johnson normal transformation.

$$Y = -0.383021 + 1.5283 \operatorname{arcsinh}\left(\frac{X - 16.8127}{8.74075}\right), \quad (7)$$

$$X = 8.740775 \sinh \frac{Y + 0.383021}{1.5283} + 16.8127.$$

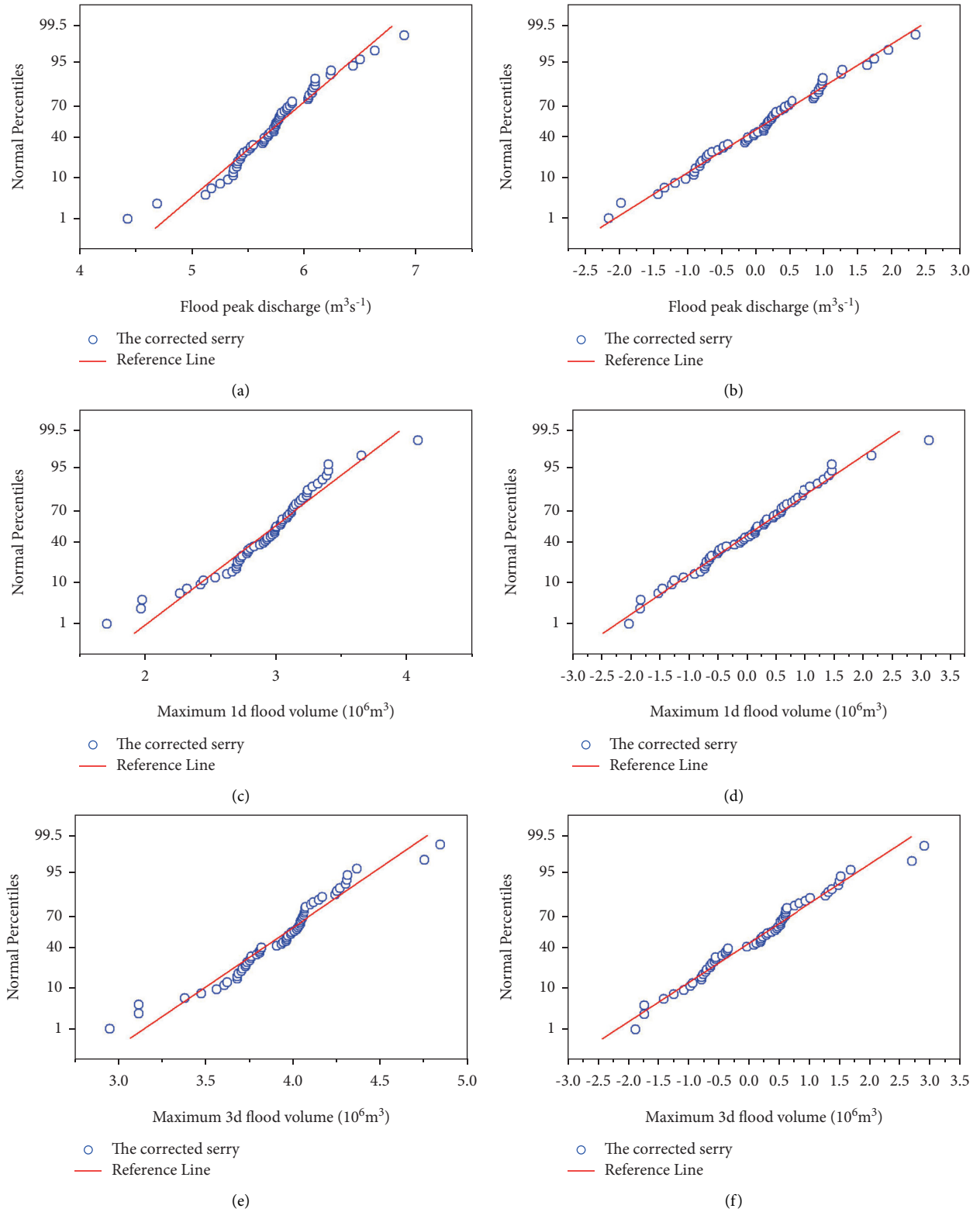


FIGURE 5: Continued.

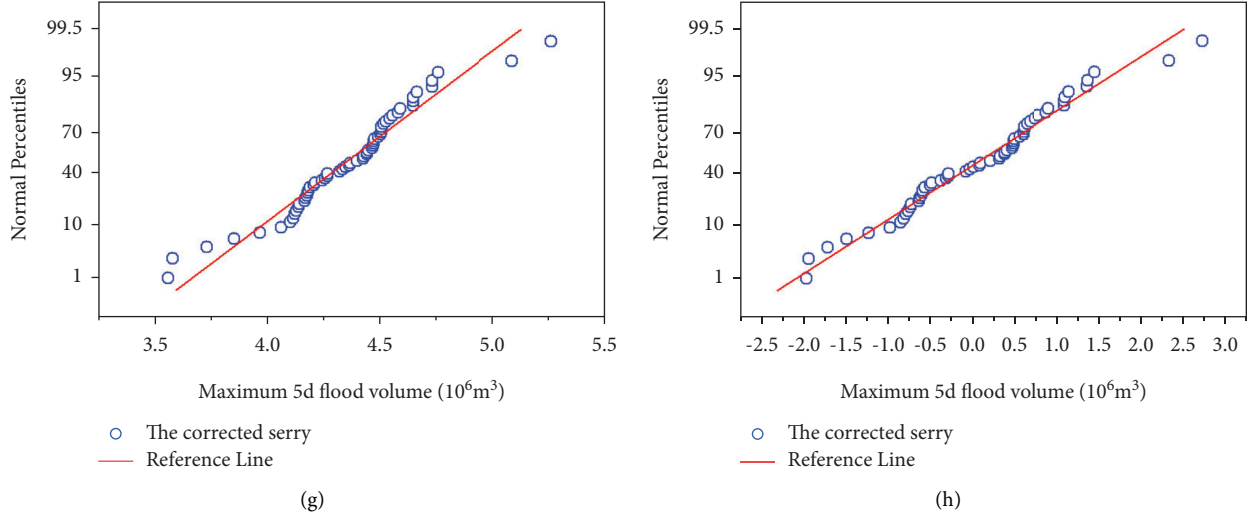


FIGURE 5: The normal test results of the transformed flood serry. (a) and (b) The Box–Cox transformation and Johnson transformation of the annual maximum flood peak discharge, (c) and (d) the Box–Cox transformation and Johnson transformation of the maximum 1 d flood volume, (e) and (f) the Box–Cox transform and Johnson transformation of the maximum 3 d flood volume, and (g) and (h) the Box–Cox transform and Johnson transformation of the maximum 5 d flood volume.

### (II) Maximum 3 d flood volume serry

The normal transformation results of maximum 3 d flood volume serry are that the best transformation parameter of  $\lambda$  is equal to 0 for Box–Cox normal transformation and the best normal transformation equations are as follows for Johnson normal transformation.

$$Y = -0.118591 + 1.36577 \operatorname{arc} \sinh \left( \frac{X - 48.5286}{17.466} \right), \quad (8)$$

$$X = 17.466 \sinh \frac{Y + 0.118591}{1.36577} + 48.5286.$$

### (III) Maximum 5 d flood volume serry

The normal transformation results of maximum 5 d flood volume serry are that the best transformation parameter of  $\lambda$  is equal to 0 for Box–Cox normal transformation and the best normal transformation equations are as follows for Johnson normal transformation.

$$Y = -0.54766 + 1.48425 \operatorname{arc} \sinh \left( \frac{X - 66.3067}{28.1313} \right), \quad (9)$$

$$X = 28.1313 \sinh \left( \frac{Y + 0.54766}{1.48425} \right) + 66.3067.$$

## 3.3. Flood Frequency

**3.3.1. Flood Peak Discharge Frequency Curve.** Using the inverse transformation model of Box–Cox and Johnson normal transformation given above, the design value (normal quantile) of normal distribution under specified frequency is used to deduce the design flood peak discharge corresponding to each frequency of the original distribution and draw the flood peak discharge frequency curve. At the

same time, based on P-III distribution, the design flood peak discharge under each frequency is calculated by using the Optimization Curve-Fitting Method [40, 41], and the P-III distribution of flood peak discharge frequency curve can be drawn. The design flood peak discharge under specified frequency of normal distribution obtained by two normal transformations is shown in Tables 3 and 4. The three kinds of flood peak discharge frequency curves are shown in Figure 6(a).

It can be seen from Figure 6(a) that the flood peak discharge frequency curve deduced by Johnson normal transformation fits the measured value best. The fitting goodness of Box–Cox transformation and P-III distribution cannot be seen directly, so further quantitative calculation is needed.

**3.3.2. Flood Volume Frequency Curve.** Based on Box–Cox normal transformation, Johnson normal transformation, and P-III distribution, the flood volume frequency curves of maximum 1 d, 3 d, and 5 d are deduced. And the results are shown in Figures 6(b)–6(d).

**3.4. Goodness-of-Fit Calculation.** Using the two evaluation indexes of Mean Square Error (MSE) and Residual Sum of Squares (RSS), in this paper, the goodness of fitness between the measured values of high flood discharge (10%, 30%, and 50%) and low flood discharge (10% and 30%) with the corresponding designed values of different methods are calculated [3, 42, 43]. The results are shown in Table 5.

According to the calculation results of RSS evaluation index of flood peak frequency, the order of goodness of fitness between the measured value and the designed value is Johnson transformation > Box–Cox transformation > P-III distribution. And the calculation result of MSE evaluation index is the same as the RSS evaluation index.

TABLE 3: The design flood peak discharge at specified frequency of normal distribution (Box-Cox).

| Box-Cox transformation |           |    |           |    |           |    |           |
|------------------------|-----------|----|-----------|----|-----------|----|-----------|
| 1                      | 4.8478894 | 16 | 5.4650952 | 31 | 5.7356066 | 46 | 6.0101125 |
| 2                      | 4.9688762 | 17 | 5.4852734 | 32 | 5.7521198 | 47 | 6.0319379 |
| 3                      | 5.0462946 | 18 | 5.5050398 | 33 | 5.7694154 | 48 | 6.0547928 |
| 4                      | 5.1051820 | 19 | 5.5244768 | 34 | 5.7862992 | 49 | 6.0786772 |
| 5                      | 5.1537744 | 20 | 5.5433372 | 35 | 5.8035948 | 50 | 6.1033852 |
| 6                      | 5.1953662 | 21 | 5.5618682 | 36 | 5.8208904 | 51 | 6.1301522 |
| 7                      | 5.2324282 | 22 | 5.5799874 | 37 | 5.8381860 | 52 | 6.1581546 |
| 8                      | 5.2653722 | 23 | 5.5981066 | 38 | 5.8558934 | 53 | 6.1886278 |
| 9                      | 5.2958454 | 24 | 5.6158140 | 39 | 5.8740126 | 54 | 6.2215718 |
| 10                     | 5.3238478 | 25 | 5.6331096 | 40 | 5.8921318 | 55 | 6.2586338 |
| 11                     | 5.3506148 | 26 | 5.6504052 | 41 | 5.9106628 | 56 | 6.3002256 |
| 12                     | 5.3753228 | 27 | 5.6677008 | 42 | 5.9295232 | 57 | 6.3488180 |
| 13                     | 5.3992072 | 28 | 5.6845846 | 43 | 5.9489602 | 58 | 6.4077054 |
| 14                     | 5.4220621 | 29 | 5.7018802 | 44 | 5.9687266 | 59 | 6.4851238 |
| 15                     | 5.4438875 | 30 | 5.7183934 | 45 | 5.9889048 | 60 | 6.6061106 |

TABLE 4: The design flood peak discharge at specified frequency of normal distribution (Johnson).

| Johnson transformation |            |    |            |    |           |    |           |
|------------------------|------------|----|------------|----|-----------|----|-----------|
| 1                      | -1.8931003 | 16 | -0.5200496 | 31 | 0.0817365 | 46 | 0.6924088 |
| 2                      | -1.6239501 | 17 | -0.4751607 | 32 | 0.1184721 | 47 | 0.7409621 |
| 3                      | -1.4517233 | 18 | -0.4311879 | 33 | 0.1569483 | 48 | 0.7918056 |
| 4                      | -1.3207210 | 19 | -0.3879480 | 34 | 0.1945084 | 49 | 0.8449394 |
| 5                      | -1.2126212 | 20 | -0.3459906 | 35 | 0.2329846 | 50 | 0.8999054 |
| 6                      | -1.1200951 | 21 | -0.3047661 | 36 | 0.2714608 | 51 | 0.9594519 |
| 7                      | -1.0376461 | 22 | -0.2644577 | 37 | 0.3099370 | 52 | 1.0217467 |
| 8                      | -0.9643581 | 23 | -0.2241493 | 38 | 0.3493293 | 53 | 1.0895381 |
| 9                      | -0.8965667 | 24 | -0.1847570 | 39 | 0.3896377 | 54 | 1.1628261 |
| 10                     | -0.8342719 | 25 | -0.1462808 | 40 | 0.4299461 | 55 | 1.2452751 |
| 11                     | -0.7747254 | 26 | -0.1078046 | 41 | 0.4711706 | 56 | 1.3378012 |
| 12                     | -0.7197594 | 27 | -0.0693284 | 42 | 0.5131280 | 57 | 1.4459010 |
| 13                     | -0.6666256 | 28 | -0.0317683 | 43 | 0.5563679 | 58 | 1.5769033 |
| 14                     | -0.6157821 | 29 | 0.0067079  | 44 | 0.6003407 | 59 | 1.7491301 |
| 15                     | -0.5672288 | 30 | 0.0434435  | 45 | 0.6452296 | 60 | 2.0182803 |

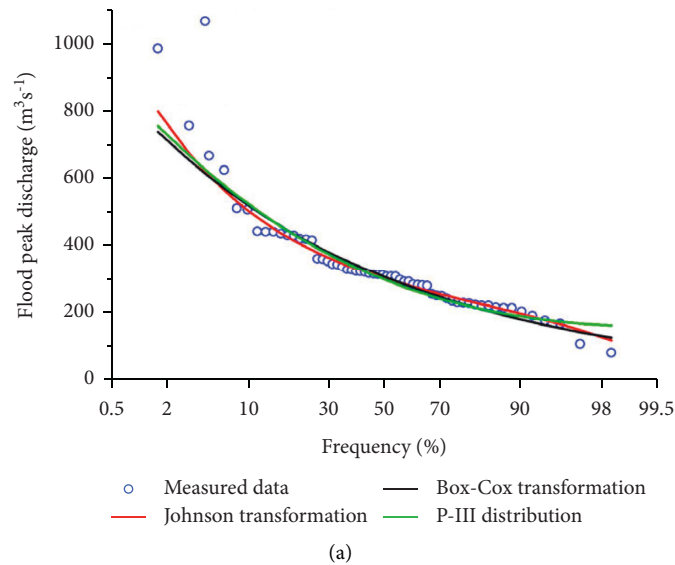


FIGURE 6: Continued.

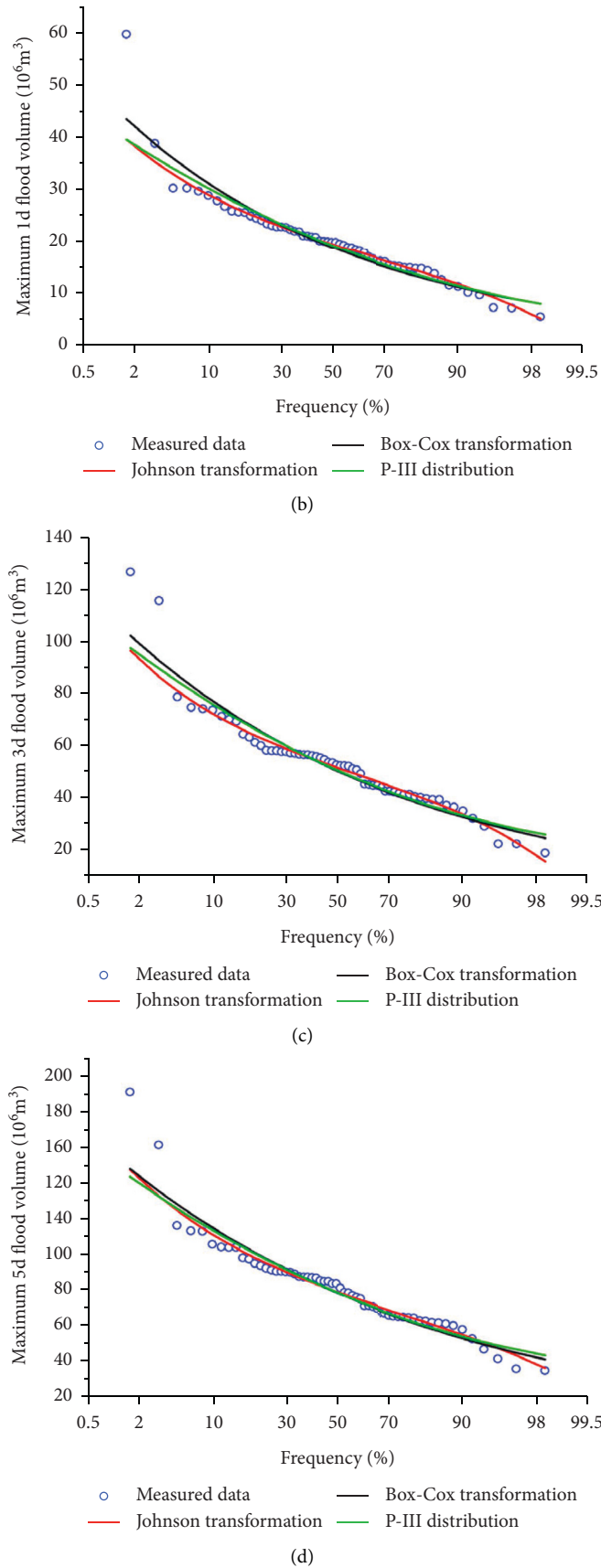


FIGURE 6: The three flood frequency curves. (a) The calculation result of the annual maximum flood peak discharge frequency, (b) the calculation result of the maximum 1 d flood volume frequency, (c) the calculation result of the maximum 3 d flood volume frequency, and (d) the calculation result of the maximum 5 d flood volume frequency.



TABLE 6: Statistical tables of goodness of fitness.

| Variables                | Method  | High | Low |
|--------------------------|---------|------|-----|
| Flood peak               | P-III   | 0    | 0   |
|                          | Box-Cox | 0    | 0   |
|                          | Johnson | 6    | 4   |
| Maximum 1 d flood volume | P-III   | 0    | 0   |
|                          | Box-Cox | 6    | 0   |
|                          | Johnson | 0    | 4   |
| Maximum 3 d flood volume | P-III   | 0    | 0   |
|                          | Box-Cox | 6    | 0   |
|                          | Johnson | 0    | 4   |
| Maximum 5 d flood volume | P-III   | 0    | 0   |
|                          | Box-Cox | 0    | 0   |
|                          | Johnson | 6    | 4   |

According to the calculation results of RSS evaluation index of flood volume frequency, for the maximum 1 d flood volume serry, the order of goodness of fitness between the measured value and the designed value is Box-Cox transformation > Johnson transformation > P-III distribution at high flood discharge (10%, 30%, and 50%), and Johnson transformation > Box-Cox transformation > P-III distribution at low flood discharge (10% and 30%), and the calculation results of MSE evaluation index are the same as the RSS evaluation index.

For the maximum 3 d flood volume serry, the order of goodness of fitness between the measured value and the designed value is Box-Cox transformation > P-III distribution > Johnson transformation at high flood discharge (10%, 30%, and 50%), and Johnson transformation > Box-Cox transformation > P-III distribution at low flood discharge (10% and 30%), and the calculation results of MSE evaluation index are the same as RSS evaluation index.

For the maximum 5 d flood volume serry, the order of goodness of fitness between the measured value and the designed value is Johnson transformation > Box-Cox transformation > P-III distribution both at high flood discharge (10%, 30%, and 50%) and low flood discharge (10% and 30%), and the calculation results of MSE evaluation index are the same as the RSS evaluation index.

The calculation results of goodness of fitness are divided into two parts: high flood discharge and low flood discharge. The number of advantages of the three methods in each part is counted. The results are shown in Table 6.

#### 4. Conclusion

In this paper, 60 years measured flood data of Manas River in north-western China are used to study the adaptability of flood frequency analysis based on normal transformation to Manas River, by comparing with the traditional P-III distribution. The conclusions are as follows:

- (1) For the flood peak discharge frequency, Johnson transformation is best at the fitting accuracy between frequency curve with empirical data among the three methods of Johnson transformation, Box-Cox transformation, and P-III distribution. So, Johnson

transformation is more excellent and adaptive than the traditional P-III distribution in analysis of the flood peak discharge frequency of Manas River.

- (2) For the flood volume frequency, Johnson transformation has strong adaptability to stable hydrologic serry, but Box-Cox transformation has strong adaptability to unstable high flood discharge of hydrologic serry. The main reason is that Johnson transformation is a multiparameter transformation and that Box-Cox transformation is a single-parameter transformation. The multiparameter transformation is more stable than the single-parameter one, so does the inverse transformation. Therefore, there is a higher fitting accuracy between all part of frequency curve with empirical data of Johnson transformation than Box-Cox transformation. But at the high flood discharge part of frequency curve, the fitting accuracy of Box-Cox transformation is higher than that of Johnson transformation.
- (3) The P-III distribution frequency analysis method has its unique advantages, and the calculation results are relatively stable. When the distribution of the original hydrologic serry is more consistent with the P-III distribution, more information can be obtained from the sample, and the calculation results are more accurate, but when the distribution of the original hydrologic serry is not consistent with the P-III distribution, it will cause more errors. The results of this paper show that the normal transformation method has certain advantages and rationality at flood frequency analysis of rivers in arid island region of north-western China. Therefore, the precision of flood frequency analysis can be improved by using P-III distribution and normal transformation comprehensively.

#### Data Availability

The data of Kenswat Hydrologic Station with the longest observation time of Manas River are provided by Kenswat Hydrological Station. The data period selected in this paper is from 1955 to 2014.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### Acknowledgments

This work was financially supported by the Xinjiang Production and Construction Corps (nos. 2022DB024 and 2021AA003) and is gratefully acknowledged.

#### References

- [1] I. Yu, M. L. A. Felix, L. H. Kim, and S. Jeong, "Analysis on flood frequency and water quality variations induced by abnormal climate," *Desalination and Water Treatment*, vol. 53, no. 9, pp. 2339–2353, 2015.

- [2] Q. Schiermeier, "Increased flood risk linked to global warming," *Nature*, vol. 470, no. 7334, 316 pages, 2011.
- [3] C. Q. Ye, X. H. Chen, Q. X. Shao, and J. M. Zhang, "A comparison study on distribution for flood frequency analysis focus on the observed flow data in the high flow part (in Chinese with English abstract)," *Journal of Hydraulic Engineering*, vol. 44, pp. 68–76, 2013.
- [4] Y. F. Chen, Y. Hou, P. Van Gelder, and S. Zhigui, "Study of parameter estimation methods for Pearson-III distribution in flood frequency analysis," *Extremes of the Extremes: Extraordinary Floods*, vol. 271, pp. 263–269, 2002.
- [5] H. Vittal, J. Singh, P. Kumar, and S. Karmakar, "A framework for multivariate data-based at-site flood frequency analysis: essentiality of the conjugal application of parametric and nonparametric approaches," *Journal of Hydrology*, vol. 525, pp. 658–675, 2015.
- [6] K. Adamowski, "A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies," *Journal of Hydrology*, vol. 108, pp. 295–308, 1989.
- [7] S. L. Guo, "Nonparametric estimation of flood frequencies," *Water Resources and Power*, vol. 9, pp. 324–332, 1991.
- [8] S. B. Song, "Research challenges and suggestions of hydrological frequency calculation (in Chinese with English abstract)," *Journal of Water Resources and Architectural Engineering*, vol. 17, pp. 12–18, 2019.
- [9] P. Vijay and W. G. Singh, "Strupczewski, on the status of flood frequency analysis," *Hydrological Processes*, vol. 16, pp. 3737–3740, 2002.
- [10] K. Adamowski, "Nonparametric kernel estimation of flood frequencies," *Water Resources Research*, vol. 21, no. 11, pp. 1585–1590, 1985.
- [11] Z. M. Liang and C. J. Dai, "Comparisons of two normalization approaches in hydrologic analytic calculation," *Hydroelectric Energy*, vol. 23, pp. 1–3+89, 2005.
- [12] J. A. Greenwood, J. M. Landwehr, N. C. Matalas, and J. R. Wallis, "Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form," *Water Resources Research*, vol. 15, no. 5, pp. 1049–1054, 1979.
- [13] A. I. Fleishman, "A method for simulating non-normal distributions," *Psychometrika*, vol. 43, no. 4, pp. 521–532, 1978.
- [14] X. Y. Chen and Y. K. Tung, "Investigation of polynomial normal transform," *Structural Safety*, vol. 25, no. 4, pp. 423–445, 2003.
- [15] W. Xu, "Efficiency of Bayesian probabilistic hydrological forecast system based on Box-Cox transformation (in Chinese with English abstract)," *Journal of Hydroelectric Engineering*, vol. 37, pp. 15–23, 2018.
- [16] D. F. Chen and S. B. Song, "Investigation of annual runoff frequency calculation based on polynomial normal transformation (in Chinese with English abstract)," *Journal of Hydroelectric Engineering*, vol. 38, pp. 58–69, 2019.
- [17] D. B. Zhuo and G. P. Xia, "Non-normal Processes Quality Control Using the Best Johnson Transformation," in *Proceedings of the 7th international conference on industrial management*, pp. 177–182, Japan, November 2004.
- [18] W. M. Luh and J. H. Guo, "Using Johnson's transformation and robust estimators with heteroscedastic test statistics: an examination of the effects of non-normality and heterogeneity in the non-orthogonal two-way ANOVA design," *British Journal of Mathematical and Statistical Psychology*, vol. 54, no. 1, pp. 79–94, 2001.
- [19] Y. Q. Li, J. Zhou, and X. Z. Zhou, "Stage-Discharge Curve Fitting Based on Box-Cox Transform," *Journal of Yangtze River Scientific Research Institute*, vol. 37, 2018.
- [20] J. Y. Yoon, J. Kim, and S. Song, "Comparison of parameter estimation methods for normal inverse Gaussian distribution," *Communications for Statistical Applications and Methods*, vol. 27, no. 1, pp. 97–108, 2020.
- [21] V. M. Guerrero and R. A. Johnson, "Use of the box-cox transformation with binary response models," *Biometrika*, vol. 69, no. 2, pp. 309–314, 1982.
- [22] N. Taylor, "Realised variance forecasting under Box-Cox transformations," *International Journal of Forecasting*, vol. 33, no. 4, pp. 770–785, 2017.
- [23] M. J. Gurka, L. J. Edwards, K. E. Muller, and L. L. Kupper, "Extending the Box-Cox transformation to the linear mixed model," *Journal of the Royal Statistical Society: Series A*, vol. 169, no. 2, pp. 273–288, 2006.
- [24] T. Hamasaki and S. Y. Kim, "Box and Cox power-transformation to confined and censored nonnormal responses in regression," *Computational Statistics & Data Analysis*, vol. 51, no. 8, pp. 3788–3799, 2007.
- [25] T. Proietti and H. Lütkepohl, "Does the Box-Cox transformation help in forecasting macroeconomic time series?" *International Journal of Forecasting*, vol. 29, no. 1, pp. 88–99, 2013.
- [26] E. J. J., "MINITAB: HandStudent book," *Technometrics*, vol. 20, no. 2, pp. 211–212, 1978.
- [27] Q. Y. Zhou, P. Tian, and Z. Y. Tian, "Estimation of non-normal process capability index based on johnson system of transformations (in Chinese with English abstract)," *Systems Engineering*, vol. 22, pp. 98–102, 2004.
- [28] J. F. Slikker and S. S. Shapiro, "The johnson system: selection and parameter estimation," *Technometrics*, vol. 22, no. 2, pp. 239–246, 1980.
- [29] A. Mateus and M. Tome, "Estimating the parameters of the johnson's S-B distribution using an approach of method of moments," *Numerical Analysis and Applied Mathematics Icnam 2011: International Conference on Numerical Analysis and Applied Mathematics*, vol. 1389, 2011.
- [30] H. J. H. Tuentert, "An algorithm to determine the parameters of  $S_B$ -curves in the johnson system of probability distributions by moment matching," *Journal of Statistical Computation and Simulation*, vol. 70, no. 4, pp. 325–347, 2001.
- [31] I. D. Hill, R. Hill, and R. L. Holder, "Algorithm AS 99: fitting johnson curves by moments," *Applied Statistics*, vol. 25, no. 2, p. 180, 1976.
- [32] Y. M. Chou, A. M. Polansky, and R. L. Mason, "Transforming non-normal data to normality in statistical process control," *Journal of Quality Technology*, vol. 30, no. 2, pp. 133–141, 1998.
- [33] S. T. Mandraccia, G. D. Halverson, and Y. M. Chou, "Control chart design strategies for skewed data," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 2876, pp. 196–205, 1996.
- [34] A. M. Polansky, Y. M. Chou, and R. L. Mason, "An algorithm for fitting Johnson transformations to non-normal data," *Journal of Quality Technology*, vol. 31, no. 3, pp. 345–350, 1999.
- [35] P. F. Li, G. Yang, X. L. He, F. D. Li, K. Yan, and Z. L. Wang, "Effects of drip irrigation on components of water cycle in arid inland areas: a case study of Manas river basin in north-western China," *International Journal of Agricultural and Biological Engineering*, vol. 12, no. 1, pp. 132–138, 2019.



- [36] H. B. Ling, H. L. Xu, W. Shi, and Q. Q. Zhang, "Regional climate change and its effects on the runoff of Manas River, Xinjiang, China," *Environmental Earth Sciences*, vol. 64, no. 8, pp. 2203–2213, 2011.
- [37] Z. M. Liang, Y. M. Hu, and J. Wang, "Advances in hydrological frequency analysis of non-stationary time series (in Chinese with English abstract)," *Advances in Water Science*, vol. 22, pp. 864–871, 2011.
- [38] M. N. Khaliq, T. Ouarda, J. C. Ondo, P. Gachon, and B. Bobee, "Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review," *Journal of Hydrology*, vol. 329, no. 3-4, pp. 534–552, 2006.
- [39] P. Xie, G. C. Chen, and J. Xia, "Hydrological frequency calculation principle of inconsistent annual runoff series under Changing environments (in Chinese with English abstract)," *Engineering Journal Of Wuhan University*, vol. 38, pp. 8–11+17, 2005.
- [40] G. Q. Wu, "Application of social spider optimization algorithm in parameter optimization of hydrological frequency curve (in Chinese with English abstract)," *Journal of Water Resources and Water Engineering*, vol. 26, pp. 123–126+131, 2015.
- [41] J. M. Zhang, X. H. Chen, and C. Q. Ye, "Calculation of negative-skewness hydrological series with Pearson-III frequency curve (in Chinese with English abstract)," *Journal of Hydraulic Engineering*, vol. 43, pp. 1296–1301, 2012.
- [42] M. I. Brunner, R. Furrer, and A. C. Favre, "Modeling the spatial dependence of floods using the Fisher copula," *Hydrology and Earth System Sciences*, vol. 23, no. 1, pp. 107–124, 2019.
- [43] R. C. Phillips, S. Z. Samadi, and M. E. Meadows, "How extreme was the October 2015 flood in the Carolinas? An assessment of flood frequency analysis and distribution tails," *Journal of Hydrology*, vol. 562, pp. 648–663, 2018.