

Article

Comparative Study on the Selection Criteria for Fitting Flood Frequency Distribution Models with Emphasis on Upper-Tail Behavior

Xiaohong Chen ^{1,2}, Quanxi Shao ³, Chong-Yu Xu ^{4,5}, Jiaming Zhang ^{1,2}, Lijuan Zhang ^{1,2} and Changqing Ye ^{6,*}

¹ Center for Water Resources and Environment, Sun Yat-sen University, Guangzhou 510275, China; eescxh@mail.sysu.edu.cn (X.C.); zhjiam3@gmail.com (J.Z.); purple336@163.com (L.Z.)

² Key Laboratory of Water Cycle and Water Security, Southern China of Guangdong High Education Institute, Sun Yat-sen University, Guangzhou 510275, China

³ CSIRO Mathematics, Informatics and Statistics, Private Bag No 5, Wembley 6913, Australia; Quanxi.shao@csiro.au

⁴ State Key Laboratory of Hydrology–Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China; c.y.xu@geo.uio.no

⁵ Department of Geosciences, University of Oslo, Oslo N-0316, Norway

⁶ Institute of Tropical Agriculture and Forestry, Hainan University, Haikou 570228, China

* Correspondence: yechangqing2001@hotmail.com; Tel.: +86-898-6619-2915

Academic Editor: Marco Franchini

Received: 21 February 2017; Accepted: 25 April 2017; Published: 2 May 2017

Abstract: The upper tail of a flood frequency distribution is always specifically concerned with flood control. However, different model selection criteria often give different optimal distributions when the focus is on the upper tail of distribution. With emphasis on the upper-tail behavior, five distribution selection criteria including two hypothesis tests and three information-based criteria are evaluated in selecting the best fitted distribution from eight widely used distributions by using datasets from Thames River, Wabash River, Beiji River and Huai River. The performance of the five selection criteria is verified by using a composite criterion with focus on upper tail events. This paper demonstrated an approach for optimally selecting suitable flood frequency distributions. Results illustrate that (1) there are different selections of frequency distributions in the four rivers by using hypothesis tests and information-based criteria approaches. Hypothesis tests are more likely to choose complex, parametric models, and information-based criteria prefer to choose simple, effective models. Different selection criteria have no particular tendency toward the tail of the distribution; (2) The information-based criteria perform better than hypothesis tests in most cases when the focus is on the goodness of predictions of the extreme upper tail events. The distributions selected by information-based criteria are more likely to be close to true values than the distributions selected by hypothesis test methods in the upper tail of the frequency curve; (3) The proposed composite criterion not only can select the optimal distribution, but also can evaluate the error of estimated value, which often plays an important role in the risk assessment and engineering design. In order to decide on a particular distribution to fit the high flow, it would be better to use the composite criterion.

Keywords: flood frequency analysis; probability distributions; hypothesis testing; information-based criteria; upper-tail behavior

1. Introduction

Flood frequency analysis plays a key role and is a constant topic in hydrology and water resources, especially for hydraulic design and flood hazard mitigation and management (e.g., [1,2]). Adequate

estimations of extreme annual maximum daily flow are very important for flood control in which the upper-tail behavior of the flood frequency distribution is the key [3,4]. The frequency analysis of hydrological extremes requires a fit of a probability distribution to the observed data in order to suitably represent the frequency of occurrence of rare events [5]. More than 20 statistical distributions have been used as the flood frequency distributions [3]. Statistical criteria must be used to determine the suitable distribution for flood frequency analysis [6]. However, for a given region, different model selection methods often result in different optimal distributions, especially when the focus is on the upper tail of flood frequency distribution [7]. The flood estimation vary widely for different distributions. Therefore, the most suitable distribution must be chosen.

There are mainly two kinds of model selection techniques: hypothesis tests based on goodness-of-fit and information-based criteria [5]. The commonly used hypothesis tests are the Kolmogorov–Smirnov (KS) test, Anderson–Darling (AD) test, probability plot correlation coefficient (PPCC), chi-squared test and log-likelihood ratio tests (t-test and F-test). Information-based criteria include the Akaike Information Criterion [8], Akaike Information Criterion–second order variant (AICc) and Bayesian Information Criterion (BIC).

There have been some studies in the past on the comparison of various model selection methods. The choice of a distribution for flood frequency should be based on features reflecting the upper tail shape [9]. However, there are rare studies about the comparison of model selection criteria with emphasis on the upper tail of flood frequency distribution. Cicioni et al. (1973) considered the two-parameter lognormal (LN2), three-parameter log-normal (LN3), Pearson type III distribution (P3) and Generalized Extreme Value (GEV) distributions for the flood data from 108 stations in Italy with record length of more than 27 years, and used Chi-squared, KS, Cramer–Von Mises and AD tests for distribution selection, giving the result that the Chi-squared test selected LN2 but other tests selected GEV [7]. Haktanir and Horlacher (1993) applied a statistical model comprising nine different probability distributions for flood frequency analysis of annual flood peak series for 11 unregulated streams [10]. The distributions were compared by classical goodness-of-fit tests (GOFT) on the observed series. However, different classical goodness-of-fit tests often result in different distributions for a specific region. Haddad et al. (2012) presented a case study with flood data from Tasmania in Australia in order to select the best fit flood frequency distribution by examining four model selection criteria: AIC, AICc, BIC and a modified Anderson–Darling (AD) Criterion [11]. It was found from the Monte Carlo simulation that AD is more successful in correctly recognizing the parent distribution than AIC and BIC when the parent is a three-parameter distribution. On the other hand, AIC and BIC are better at correctly recognizing the parent distribution when the parent is a two-parameter distribution. Baldassarre (2009) demonstrated that model selection criteria such as AIC, BIC and AD which are seldom used in hydrological applications, can help to identify the best probability model [12]. These three methods were compared through an extensive numerical analysis by using synthetic data samples. The model selection criteria based on AIC, BIC and AD were also adopted by Laio et al. (2009) and Calenda et al. (2009) [5,13], with further investigation to verify which of the selection criteria is more efficient, especially in the case of small samples and heavy tailed distributions, as these are commonly encountered in flood frequency analysis. The studies were carried out by a Monte Carlo simulation to investigate the robustness of the model selection criteria in recognizing the real parent distributions. Overall, none of the classical hypothesis tests and information-based criteria can be used as a universal indicator to select the suitable distributions for different stations around the world. Burnham and Anderson (2002) indicated that the hypothesis test and information-based approaches have different selection frequencies [14]. Even if the same parameter estimation method is used, different model selection criteria result in different optimal distributions. This is perhaps because each type of model selection criteria has its own characteristics and applicable scope [15]. Therefore, it is not surprising that the results of these tests are not always in agreement.

Estimating the magnitude and frequency of large floods is difficult and involves a large degree of uncertainty, especially when the flow record is of limited length. The Monte Carlo method and Paleohydrologic techniques offer a way to lengthen a short-term data record and, to reduce the uncertainty in hydrologic analysis [16–18].

The basic assumption of traditional frequency analysis methods is that the hydrological data used are stationary, independent and identically distributed over time. However, in the past decades this stationarity assumption has been severely challenged because global climate change [19] and/or large-scale human activities [20] have altered the statistical characteristics of hydrological processes [21]. Some hydrologists have declared that “stationarity is dead” [22], and suggest that nonstationary probabilistic models need to be identified and possibly used in some practical cases when the characteristics of hydrological processes have been significantly changed [23–25].

Selection of a flood frequency distribution is a necessary step in flood frequency analysis. However, selection of the best fit distribution from a large number of candidate distributions available in the literature is a difficult task. There are two reasons behind having no unique probability distributions for a given region. (1) Flood characteristics are different in different rivers; (2) there is a lack of an effective model selection criterion to be used to determine the suitable distribution for flood frequency analysis.

Flood frequency curves of different distributions show differences mainly at the tails of the distributions, especially at the high flow part which generally shows big differences for different distributions [10]. Hosking and Wallis (1986) argued that the choice of a distribution for flood frequency should be based on features reflecting the upper tail shape [9]. The observed flow data at the high flow part play an important role in the flood frequency analysis and should be addressed in the goodness-of-fit. The question is which model selection criterion can be a good indicator of the goodness of prediction for the extreme upper tail quantiles such as return periods of 100 years or more. In order to determine the more efficient model selection criterion which focuses on the upper-tail behaviour and reduces the influence of the lower tail end, a new composite criterion method to identify the optimal distribution is proposed in this study. The composite criterion can evaluate the goodness of predictions of the extreme upper-tail events carried out using synthetic samples of data by Monte Carlo simulation with Kappa distribution as the parent distribution. Stochastic simulation is widely applied for estimating the design flood of various hydrological systems.

In order to reveal the best fitted distribution for different regions in the flood frequency analysis with emphasis on the upper-tail behavior, the study aims at clarifying how the model selection methods work in different situations in the flood frequency analysis by (1) verifying whether hypothesis tests or information-based criteria methods are more efficient at the high flow part by clarifying the characteristic of model selection methods, and (2) trying to establish a composite of model selection criteria methods which can meet the demand of the engineering design. The findings from this study will benefit hazard mitigation and water resources management.

2. Methodology

2.1. Typical Probability Distributions

Many probability distributions (PDs) have been considered, in different situations, for the probabilistic model of extreme events, including P3, LP3, LN2, LN3, Gumbel (Extreme value type I, EV1), Weibull (Extreme value type III), GEV and Generalized logistic distribution (GLO). Rao and Hamed (2000) and Reiss and Thomas (2001) provided details of their probability density functions [26,27]. Eight well-known flood frequency probability distributions were used in this study. Two of them have two parameters (LN2 and Gumbel) and six have three parameters (LN3, Weibull, GEV, GLO, P3 and LP3). Two of them are heavy tail distributions (GLO and LP3), i.e., distribution tends to have large values with outliers (very high values); an often used definition of heavy tailed distributions is based on the fourth central moment [28]; four of them are mixed tail distributions (GEV, Gumbel, LN3 and LN2) and the other two are light tail distributions (P3, Weibull which can

also be subexponential). More details regarding the tail of the PDs can be found in, for example, Adlouni et al. (2008) [28].

2.2. Model Selection Methods

There are mainly two kinds of model selection techniques: hypothesis tests based on goodness-of-fit and information-based criteria [5]. The traditional hypothesis testing methods are KS and AD [7]. KS and AD methods involve the confidence level and threshold (p values). If the p value is greater than the confidence level (typically 0.05), the original hypothesis is accepted as the data obeys the distribution, otherwise the original hypothesis is rejected. It was found from related researches that information-based criteria (AIC, BIC and AICc) can help to identify the best probability model in certain situations [11,12]. With respect to the distribution selection, two hypothesis tests (KS and AD) and three information-based criteria (AIC, BIC and AICc) are used in this paper (Table 1). The distributions are ranked according to their performances against each test or criterion. The best fitted distributions are the ones which perform in the top three of all the tests and criteria. Specific steps of computing the information-based criteria for each probability model are as follows.

- (1) The log-likelihood function value for each probability model was computed according to Table 1. Where parameters P (scale, location, shape) are the parameter values that maximize the log-likelihood function. The estimation method for parameter P of flood frequency probability models is the maximum likelihood, which was used to compute the log-likelihood function for each probability model.
- (2) The values of AIC, BIC, AICc can be computed according to Table 1 on the basis of the value of log-likelihood function and the number of parameters.

2.3. Parameter Estimation

The most common parameter estimation methods in flood frequency analysis are moments and the maximum likelihood [29]. Because the maximum likelihood estimation (MLE) generally shows less bias than other methods and provides a more consistent result to parameter estimation, it is recommended by Federal Emergency Management Agency of the United States (FEMA)'s guideline (2004) [30]. Therefore, in this paper, the MLE method was used for parameter estimation. More details regarding methods on parameter estimation can be found in, for example, Martins and Stedinger (2000), Hirose (1996), and Otten and Montfort (1980) [31–33].

Table 1. Model selection criteria methods for hydrological frequency analysis.

Goodness-of-Fit Test (GOFT)	Statistic Value	Description	Characteristic
KS	$D_n = \max_{1 \leq i \leq n} \left[\frac{i}{N} - F(x_{(i)}), F(x_{(i)}) - \frac{i-1}{N} \right]$ [34]	$x_{(i)}$ is a plot on the Empirical frequency curve and $F^{-1}(p)$ is the Inverse function of cumulative distribution function $F(x)$ for probability $P_{(i)}$. N is the size of samples.	KS test measures the greatest discrepancy between the observed and hypothesized distributions.
AD	$A^2 = -N - \frac{1}{N} \sum_{i=1}^N (2i-1) \left[\ln F(x_{(i)}) + \ln \left\{ 1 - F(x_{(N-i+1)}) \right\} \right]$ [35]		AD uses the sum of the squared differences between the empirical and theoretical distributions with weights to emphasize discrepancies in the tails. AD Statistic has shown good capabilities for a small sample size and heavy tailed distributions [15,36].
AIC	$AIC = -2 \ln \left[L(D \hat{\theta}) \right] + 2m$ [8]	$L(D \hat{\theta})$ is the likelihood function of a certain distribution with parameter set $\hat{\theta}$ and data array D . m is the number of parameters P and n is the size of the sample.	The log-likelihood maximised function value is used to select the model and penalize heavier for the number of estimated parameters P . In some situations where the sample size n is small with respect to the number of estimated parameters P , the AIC may perform inadequately [11]; a second-order variant of AIC, called AICc, should be used.
BIC	$BIC = -2 \ln \left[L(D \hat{\theta}) \right] + \ln(n)m$ [37]		Similar to the AIC, but developed in a Bayesian framework. BIC penalizes heavier than AIC for number of estimated parameters P and small sample sizes [11].
AICc	$AICc = -2 \ln \left[L(D \hat{\theta}) \right] + 2m \left(\frac{n}{n-m-1} \right)$ [14]		The AICc penalizes heavier than AIC for number of estimated parameters P and can be adopted when $n/P < 40$ to reduce bias [13].

2.4. Rigorous Program to Select the Optimal Distribution by Hypothesis Tests and Information-Based Criteria

In order to perform more rigorous and systematical analysis, we only present the first two optimal distributions for the hypothesis test and the information-based criteria. This is achieved through a rigorous program in finding the two optimal distributions from the candidate distributions. The procedure is demonstrated here by taking the information-based criteria as an example.

- (1) The candidate distributions are ordered from most to least favourite with AIC, BIC, AICc criteria. If the first distribution with the highest number of occurrences was selected respectively by AIC, BIC, AICc, then it is selected as the first optimal distribution of the information criteria.
- (2) After selecting the first optimal distribution, it is removed from the candidate distributions. Repeat step (1) to find the best distribution from the remaining distributions as the second optimal distribution.
- (3) In step (1), if two or more distributions have the same number of times appearing at the first position, then they will be sorted by the total number of occurrences in the preferred distribution (two or more distributions) selected respectively by AIC, BIC, AICc; the distribution with more occurrences is preferred.

2.5. Composite Criterion for Model Selection with Focus on the High Flow Part

An additional composite model selection criterion, based on an extensive numerical analysis by using synthetic data samples, is proposed here. Because the choice of a distribution for flood frequency should be based on features reflecting the upper tail shape [9], the composite criterion will be considered as a standard to make the final decision in this paper. The performances of the five model selection methods (Table 1) are compared in the “Results and Discussions” section. The upper tail of the frequency curve of this paper refers to the part of probability of exceedance <50%, which is greater than the 2-year flood. The observed flow data at the high flow part play a key role in flood frequency analysis. However, most classical model selection methods cannot evaluate the high flow part well [38]. The purpose of a composite criterion is to test and verify the performance at the upper tail of flood frequency distribution (return period more than 5-year), including the verification of the epitaxy capability (return period more than 100,200-year) of the model. Due to the limited length of observations (Table 2), the significance of perturbation at the upper tail of observed flood flow was assessed by generating synthetic samples using Monte Carlo simulation. In order to avoid overlooking the ‘true’ distribution caused by randomly multiple sampling the observed data, the representative of observed data samples was intensively analyzed before the flood frequency calculation (Table 3). Specific steps to verify the performance at the upper tail of flood frequency distribution are as follows.

Table 2. Background information of the four study basins.

Basin and Station Name	Country	Area (Km ²)	Terrain	Climate Zone	Data Length	Cs of the Flood Series
Kingston at Thames	UK	9948	Plain	Temperate	1883–2009	1.181
Lafayette at Wabash	USA	18,821	Alluvial Plain	Humid Continental Climate	1907–1991	0.280
Shijiao at Beijiang	China	38,363	Hill	Subtropical Monsoon	1956–2008	0.230
Lutaizi at Huai	China	91,620	Hill	Warm Temperate and Half Wet Monsoon Climate	1951–1998	1.198

Table 3. Randomness test for annual maximum daily flows in the four rivers.

Study Area	Significance	Persistency	Trend	Jump
		<i>t</i>	Spearman	Hurst Coefficient
Kingston at Thames	Stats	0.057	1.446	0.568
	Critical Value (5%)	1.979	1.96	0.628
	Accept or Not	yes	yes	yes
Lafayette at Wabash	Stats	1.286	−0.453	0.491
	Critical Value (5%)	1.989	1.96	0.323
	Accept or Not	yes	yes	yes
Shijiao at Beijiang	Stats	−0.953	0.927	0.500
	Critical Value (5%)	2.008	1.96	0.674
	Accept or Not	yes	yes	yes
Lutaizi at Huai	Stats	−0.534	−0.925	0.435
	Critical Value (5%)	2.014	1.96	0.255
	Accept or Not	yes	yes	yes

- (1) Choose a distribution from which the simulated data are generated. The Kappa and Wakeby distributions are widely recommendable choices [12]. Hosking (1997) used the four-parameter Kappa distribution as the overall simulation in regional flood frequency analysis and obtained reliable simulation results. The same distribution was used for the simulations in this study [39].
- (2) The four-parameter Kappa distribution, as the parent distribution, was estimated by L-moments of samples for the observed flood flow to determine parameter values. The synthetic samples, with the same length of the observations, were randomly simulated from the fitted four-parameter Kappa distribution. The detailed steps are described below:

First, the first four order linear moments are obtained based on the observed sequence. Then, based on the linear moment of the observed data, the L-moments method is used to estimate the parameters of the Kappa distribution. Finally, a random sample is generated using the Kappa distribution with the estimated parameter values. The length of the random sample is the same as the length of the observed sequence.

- (3) The simulated samples were fitted by eight distributions as recommended before. All eight probability distributions were then used to estimate the design floods with return periods $T=5, 10, 20, 30, 50, 70, 90, 100$ and 200 years.
- (4) Repeat steps (2) and (3) for a given number of times (denoted by N_{sim}), and save the calculated results. $N_{sim} = 500$ in this study.
- (5) The relative error of the design value (RE) for each simulation was calculated by

$$RE = \frac{\hat{X}_{i,T} - X_T}{X_T} \quad (1)$$

where T is the return period, X_T is the quantile of Kappa distribution with the parameter values obtained through L-moments for the observations, and $\hat{X}_{i,T}$ is the quantile of the fitted distribution by using one of the designed distributions. The Box plots were drawn according to 500 relative errors (REs), which reflect the overall situation of REs , as well as the deviation of the design value. The criteria of goodness were both the smallness in magnitude of the median of 500 REs and, equally important, the narrowness of the Box plots and of the max–min ranges of all the REs .

- (6) The root-mean-square error ($RMSE$) was calculated as the quantile corresponding to the assigned return periods, $T = 5, 10, 20, 30, 50, 70, 90, 100$ and 200 years.

$$RMSE(T) = \sqrt{\frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} \left(\frac{\hat{X}_{i,T} - X_T}{X_T} \right)^2} \quad (2)$$

where N_{sim} is the number of Monte Carlo simulations; other notations are the same as in Equation (1).

- (7) The arithmetic mean \overline{RMSE} of the $RMSE$ was calculated for the return period T for a given distribution.
- (8) The \overline{RMSE} and Box plots of REs are the composite criteria used for assessing the degree of the goodness-of-fit at the high flow part. The smaller \overline{RMSE} value means a better fitting.

2.6. Verify the Performance of the Five Selection Criteria by Using a Composite Criterion

The performance of the five selection criteria was verified by using a composite criterion with focus on upper tail events. The procedure is as follows.

- (1) The optimal (ranked as the top two) distributions selected by hypothesis tests and information-based criteria are listed first.
- (2) Test the performance of distribution selected by hypothesis tests and information-based criteria on the large floods with a long return period by a composite criterion.
- (3) Based on the test results by the composite criterion, compare the estimation error of distribution selected by hypothesis tests and information-based criteria for large floods. If the estimation error is small, this criterion which selected the distribution is better for high flow part (Shown as Box plots of $RMSE$ and RE).

2.7. Change Point of Flood Series Detection

The Rescaled Range (R/S) analysis method and Hurst Coefficient method are used to identify the change point and test the variation degree of time series. The variability and variation degree of time series are determined by the value of the Hurst Coefficient, which can be obtained by R/S analysis [40]. The Hurst coefficient value is equal to 0.5 when a time series does not have long persistence and increases/decreases from 0.5 when a series has long persistence/anti-persistence. More details regarding the method introduction can be found in, for example, Xie et al. (2008) [40], Wallis and Matalas (1970) [41]. R/S is defined as,

$$R/S = (c\tau)^h \quad (3)$$

where R is the range of cumulative departures from the mean, S is the standard deviation, and τ is the sample length, $\tau \geq 1$. According to the observed data, the least squares method can be used to obtain the parameters c and Hurst coefficients h .

3. Study Area and Data

In order to verify the applicability of the methodology in different regions around the world, four hydrological stations with long historical data are used as case studies, including Kingston at Thames River, Lafayette at Wabash River, Shijiao at Beiji River and Lutaizi at Huai River. These four stations are located in different areas in 23.5–66.5 degrees north of latitude in China, the UK and the US respectively (see Figure 1 and Table 2) with long-term data ranging from 48 to 127 years. Figure 1 gives their geographical locations and Table 2 summarizes the geographical and data information. The stations cover a wide range of climate conditions. Annual maximum daily flows are used in the analysis.

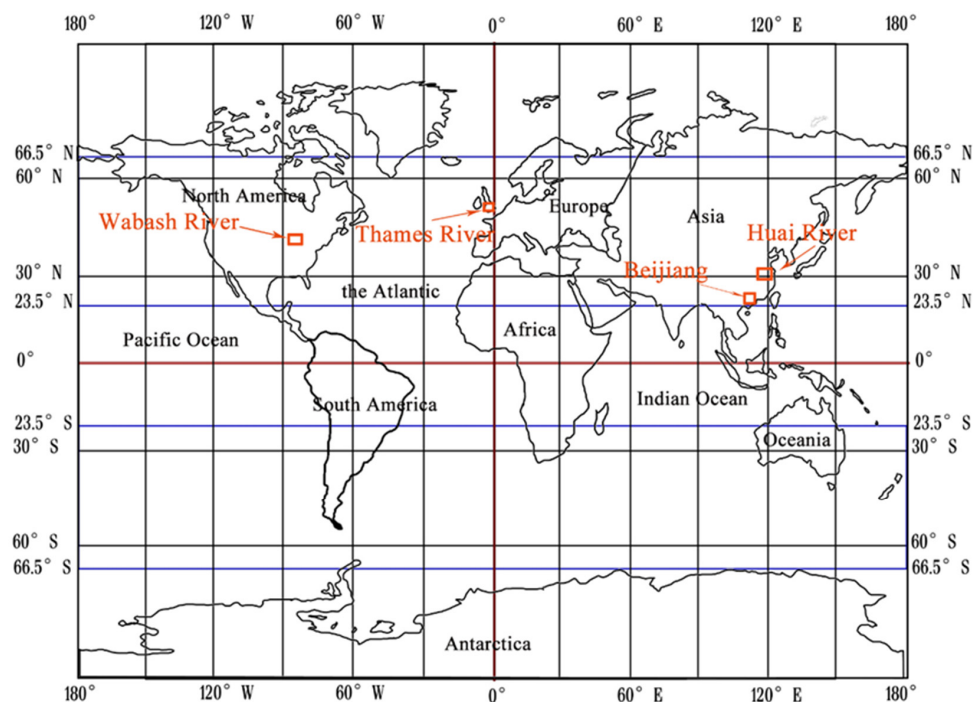


Figure 1. The locations of the studied stations.

Thames River is the biggest river in the UK with the length of 338 km and drainage area of 9948 km². It is located at a temperate climate zone with high humidity and relatively stable temperature. Kingston station, located at the lower reach of Thames River, is used in the study. The skewness coefficient C_s of the flood series at Kingston station is large with the value of 1.181, which implies a steep upper tail of the optimal frequency distribution.

With a length of 810 km, Wabash River is the largest and most important river in Indiana, USA. Wabash basin, mostly in Indiana, is dominated by a humid continental climate with cold winters, and warm and wet summers. Lafayette station, which is located at the middle reach of Wabash River and controls a drainage area of 18,821 km², is used in the study. The small C_s value of 0.280 for the flood series at Lafayette station indicates that the upper tail of the optimal frequency distributions is gentle at this station.

Beijiing River, located at the subtropical monsoon climate zone of China, has an annual average temperature between 14 and 22 °C, and an annual mean rainfall of 1700 mm. Shijiao station, the main controlling station (controlling a drainage area of 38,363 km²) located at the lower reach of the Beijiing River, is used in this study. The small C_s value of 0.230 for the flood series at Shijiao station shows a gentle upper tail frequency distribution at this station.

Huai River, located between Changjiang River (Yangtze River) and Huanghe River (Yellow River), covers a large area. Its north part is in a warm temperate zone, while the south part is in a monsoon climate zone with an annual average temperature between 11 and 16 °C. Lutaizi station, the control station in the middle river reach with a drainage area of 91,620 km², is selected as a case study in this paper. For Lutaizi Station, the large C_s value of 1.198 infers a steep upper tail frequency distribution.

The record lengths of the data are given in Table 2 in descending order. The observed flood discharge series at each station is visually investigated to see if there are apparent trends or jumps. Statistical tests including the Spearman test for trend and the R/S analysis method for change point are conducted formally and summarized in Table 3, from which it can be seen that there are no statistically significant trends and change point for annual maximum daily discharges. The fluctuation change of annual maximum flow is the biggest at Lafayette station and is the lowest at Kingston station. The autocorrelation coefficient and randomness test indicate that hydrological sequences satisfy the

independent assumption (Figure 2). Therefore, the flood series data of the studied rivers fulfil the basic assumptions of traditional frequency analysis methods, i.e., stationary, independent and identically distributed over time.

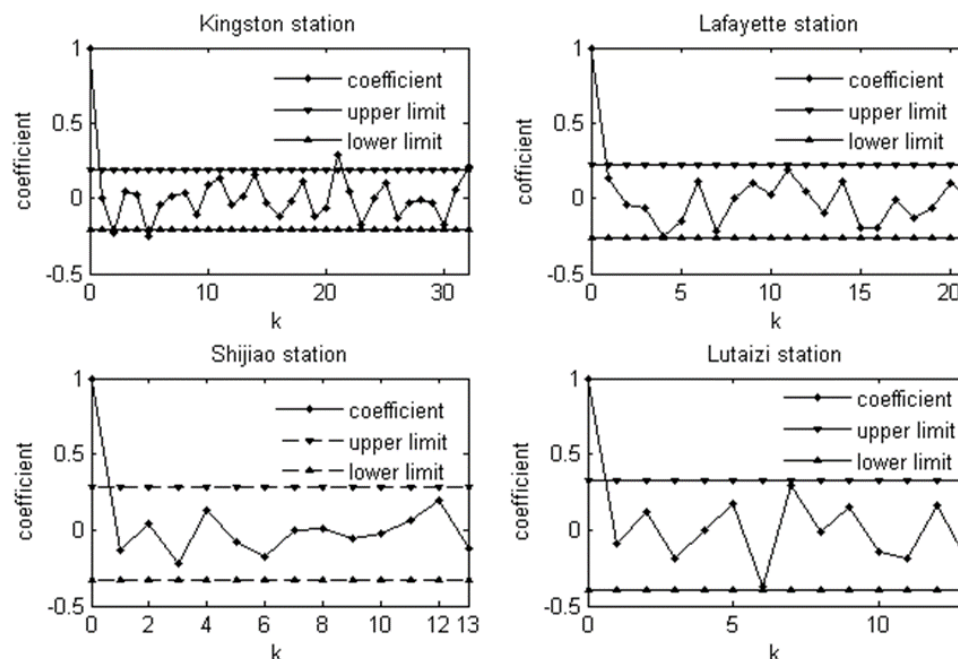


Figure 2. Autocorrelation coefficient for annual maximum daily flows in the four rivers.

4. Results and Discussions

The MLE method is conducted for parameter estimation of all eight distributions (P3, GLO, GEV, Weibull, Gumbel, LN3, LN2 and LP3), and the results are given in Table 4 with the associated return levels being plotted in Figure 3. The values of hypothesis tests and information-based criteria are summarized in Table 5, in which the smaller value for the test statistics means a better fitting by that test.

Table 4. Parameter estimation for annual maximum daily flows in the four rivers.

Study Area	PDs	Parameters (MLE)		
		Scale	Shape	Location
Kingston at Thames	P3	0.027	8.51	15.31
	GLO	56.67	−0.16	310.86
	GEV	89.059	0.036	278.97
	Weibull	282.29	2.36	76.12
	Gumbel	88.47	—	277.22
	LN3	0.26	5.94	−69.57
	LN2	0.33	5.73	—
	LP3	39.104	171.62	1.34
Lafayette at Wabash	P3	0.011	32.36	−1644.73
	GLO	299.93	−0.065	1365.28
	GEV	506.81	0.203	1185.48
	Weibull	1433.11	2.56	117.46
	Gumbel	490.76	—	1133.93
	LN3	0.12	8.41	−3140.96
	LN2	0.45	7.15	—
	LP3	28.22	171.62	1.082

Table 4. Cont.

Study Area	PDs	Parameters (MLE)		
		Scale	Shape	Location
Shijiao at Beiji	P3	0.0019	36.96	−9811.59
	GLO	1819.32	−0.068	9436.47
	GEV	3054.67	0.22	8325.95
	Weibull	8802.03	2.64	1782.32
	Gumbel	2912.39	— —	8000.08
	LN3	0.11	10.26	−19384.8
	LN2	0.37	9.11	— —
	LP3	34.13	171.36	4.087
Lutaizi at Huai	P3	0.00055	1.89	566.84
	GLO	1264.17	−0.36	3480.44
	GEV	1671.23	−0.12	2861.17
	Weibull	3676.23	1.39	672.028
	Gumbel	1750.16	— —	2942.83
	LN3	0.43	8.47	−1281.45
	LN2	0.62	8.109	— —
	LP3	20.55	166.68	0.00001

Table 5. A comparison of the test statistic values of the eight typical frequency distributions for hypothesis tests and information-based criteria.

Study Area	Frequency Distributions	KS	AD	AIC	BIC	AICc
Kingston at Thames	P3	0.064	0.502	1542.419	1550.951	1542.614
	GLO	0.053	0.292	1538.728	1547.260	1538.923
	GEV	0.054	0.355	1540.212	1548.745	1540.407
	Weibull	0.089	1.366	1552.091	1560.623	1552.286
	Gumbel	0.055	0.384	1538.708	1544.396	1538.804
	LN3	0.057	0.389	1540.801	1549.334	1540.996
	LN2	0.056	0.396	1540.227	1545.916	1540.324
	LP3	0.072	0.532	1544.700	1553.300	1544.900
Lafayette at Wabash	P3	0.060	0.443	1313.154	1320.482	1313.450
	GLO	0.070	0.397	1314.564	1321.892	1314.860
	GEV	0.063	0.455	1312.927	1320.255	1313.224
	Weibull	0.073	0.563	1312.474	1319.802	1312.770
	Gumbel	0.086	1.040	1317.343	1322.228	1317.489
	LN3	0.060	0.437	1313.190	1320.518	1313.487
	LN2	0.110	1.849	1324.446	1329.332	1324.593
	LP3	0.114	2.174	1331.104	1338.432	1331.401
Shijiao at Beiji	P3	0.106	0.436	1013.854	1019.764	1014.343
	GLO	0.122	0.573	1012.876	1018.787	1013.366
	GEV	0.098	0.424	1010.321	1016.232	1010.811
	Weibull	0.096	0.416	1010.571	1016.481	1011.060
	Gumbel	0.109	0.644	1012.438	1016.379	1012.678
	LN3	0.106	0.439	1010.753	1016.664	1011.243
	LN2	0.114	0.709	1014.045	1017.986	1014.285
	LP3	0.119	0.867	1018.547	1024.458	1019.037
Lutaizi at Huai	P3	0.077	0.181	873.463	879.077	874.008
	GLO	0.069	0.290	876.908	882.521	877.453
	GEV	0.085	0.255	875.763	881.377	876.308
	Weibull	0.070	0.188	872.889	878.503	873.435
	Gumbel	0.096	0.370	874.770	878.513	875.037
	LN3	0.088	0.237	875.352	880.966	875.898
	LN2	0.080	0.233	873.109	876.852	873.376
	LP3	0.077	0.289	875.939	881.553	876.485

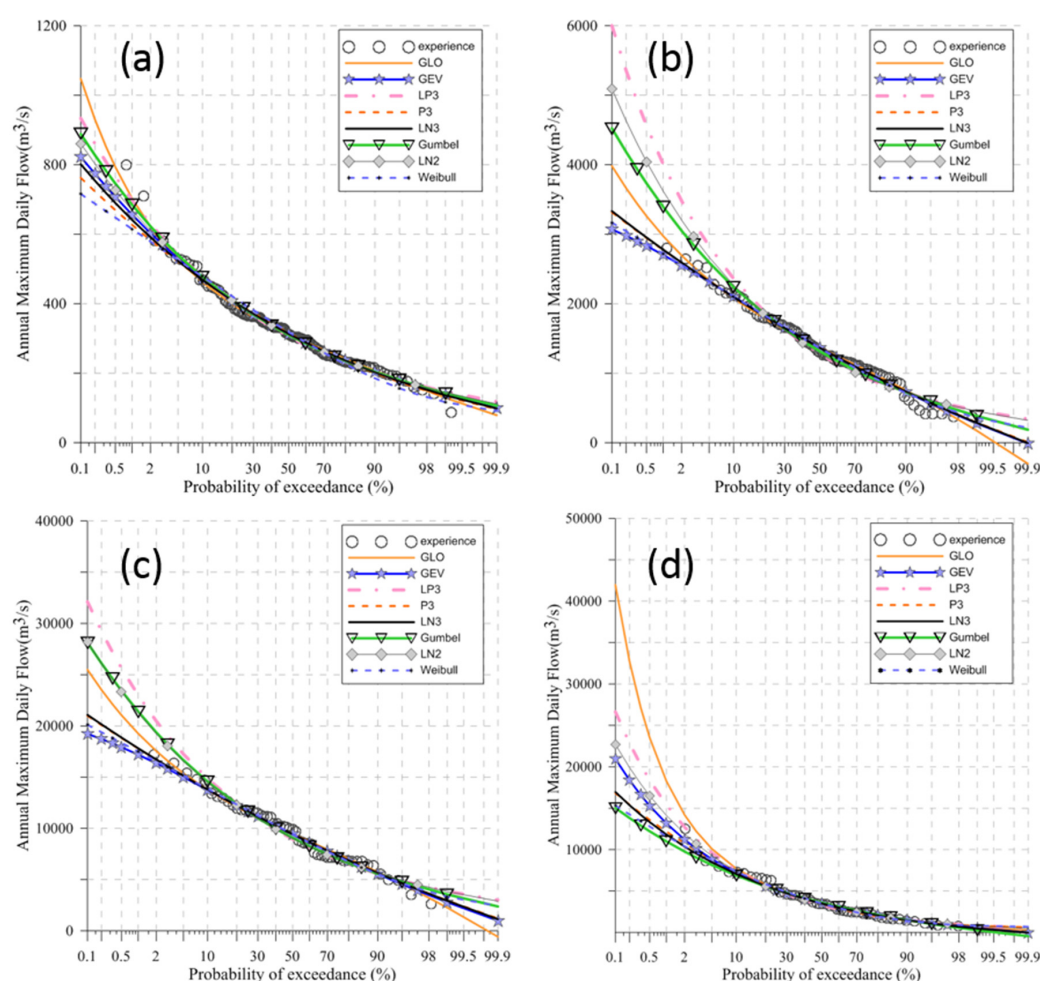


Figure 3. A comparison of the eight typical frequency distributions for four rivers with parameters estimated by MLE. (a) Thames River; (b) Wabash River; (c) Beijiing River and (d) Huai River.

4.1. Optimal Frequency Distribution for Different Model Selection Methods

There are different selections of frequency distributions by using hypothesis tests and information-based criteria approaches for each river. Taking Thames River as an example, for the hypothesis tests KS and AD, the comparison results indicate that the data are best fitted by GLO distribution, followed by GEV and Gumbel distributions (Table 5). When information-based criteria methods (including AIC, AICc and BIC) are used in the comparison, results show that Gumbel fits the observed floods best, followed by GLO distribution (see Table 5; Figure 3). Some different results can be found between hypothesis tests and information-based criteria methods. Heavy tailed GLO distribution is the best fitted frequency distribution by the hypothesis tests, while mixed tailed Gumbel distribution is the best by the information-based criteria in Thames River.

As is the case for Thames River, the best fitted flood frequency distributions in Wabash River vary slightly between two types of model selection methods. Mixed tailed LN3 distribution is the best fitted frequency distribution for hypothesis tests, while light tailed Weibull distribution is the best for information-based criteria (Table 5).

There is always a difference between the two types of selection methods in the other two river basins. In Huai River, light tailed (P3, Weibull) distributions are suitable frequency distributions for hypothesis tests, while mixed tailed (LN2) or light tailed (Weibull) distributions are the best for information-based criteria (Table 5). In Beijiing River, light tailed (Weibull, P3) and mixed tailed (GEV) distributions are suitable frequency distributions for hypothesis tests, while mixed tailed (GEV) and

light tailed (Weibull) distributions are the best for information-based criteria. The results show that the optimal flood frequency distributions are basically the same in both rivers although slightly different orders exist in Beijiing River. The study points out that in Beijiing River there is a slight tendency towards the selection of light tailed distributions, while heavy tailed distributions are inappropriate (Table 5).

4.2. Composite Criterion for Model Selection

For Thames River, the composite criterion of \overline{RMSE} and Box plots of REs can correctly recognize, in most of the cases, that the optimal distribution belongs to the Gumbel. Information-based criteria turn out to be the best methods in this case, even with varying return periods (Table 6 and Figure 4). The Cs values have a close relationship with the optimal frequency distributions (Figure 3), the large Cs value of 1.181 for the flood series at Kingston station agrees with the selection of mixed tail distribution Gumbel as the optimal distribution.

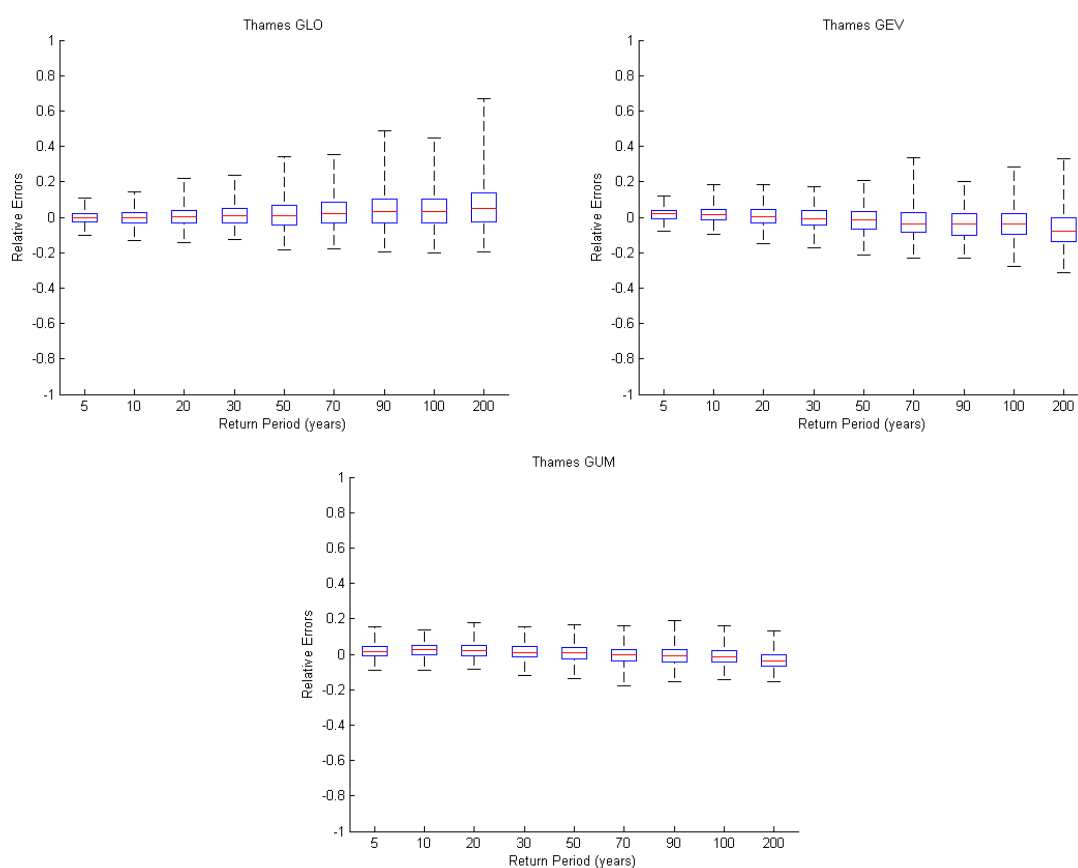


Figure 4. Box plots of the relative errors (REs) of the Kingston at Thames River for sample series length 127, with Kappa as the parent probability distribution (PD).

As is the case for Thames River, information-based criteria are shown to be the best methods in Wabash River, even with varying return periods (Table 6 and Figure 5). It is found that Weibull can be judged as a suitable flood frequency distribution, which fits high flows well and is insensitive to low flows. For Lafayette station, the smaller Cs value of 0.280 is reflected by the selection of light tail Weibull distribution. There is a slight tendency towards the selection of light tailed distributions in Wabash River.

However, hypothesis tests appear to be the best methods in Beijiing River, even with varying return periods (Table 6 and Figure 6). In this river basin, Weibull is inferred as the suitable flood

frequency distribution based on the composite criterion of \overline{RMSE} and Box plots of REs . The smallest C_s value of 0.230 for Shijiao station is consistent with the selection of light tail Weibull distribution.

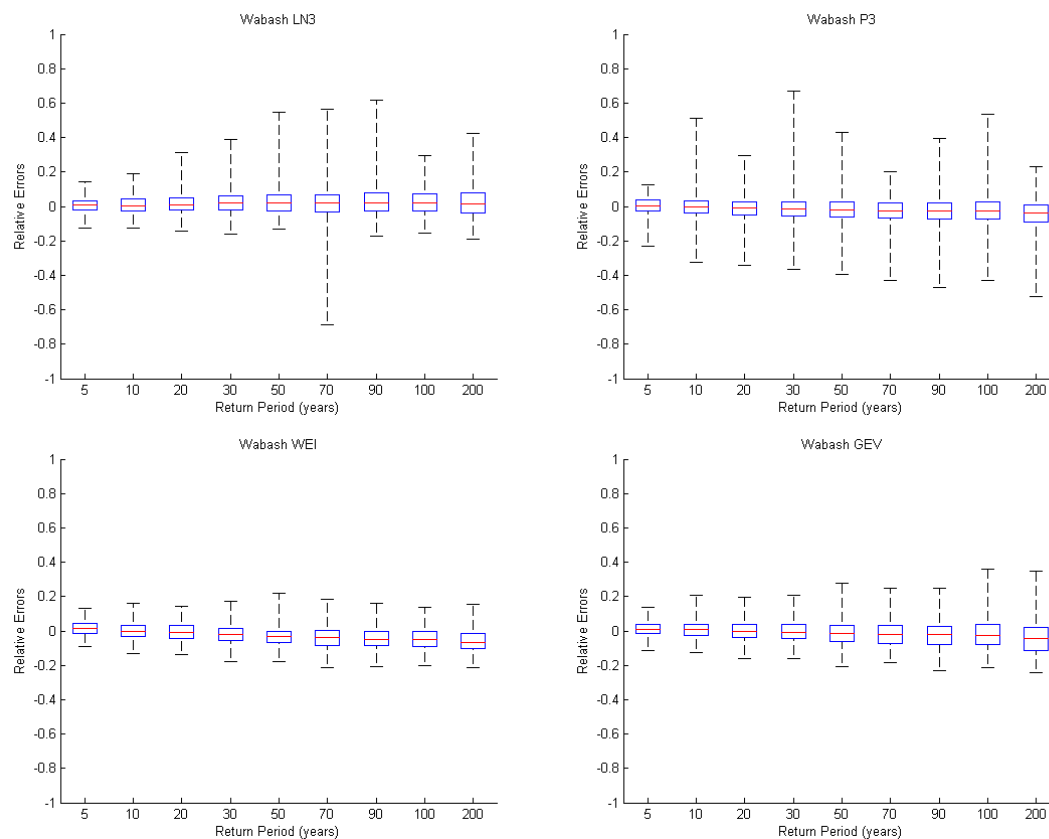


Figure 5. Box plots of the relative errors (REs) of the Lafayette at Wabash River for sample series length 85, with Kappa as the parent PD.

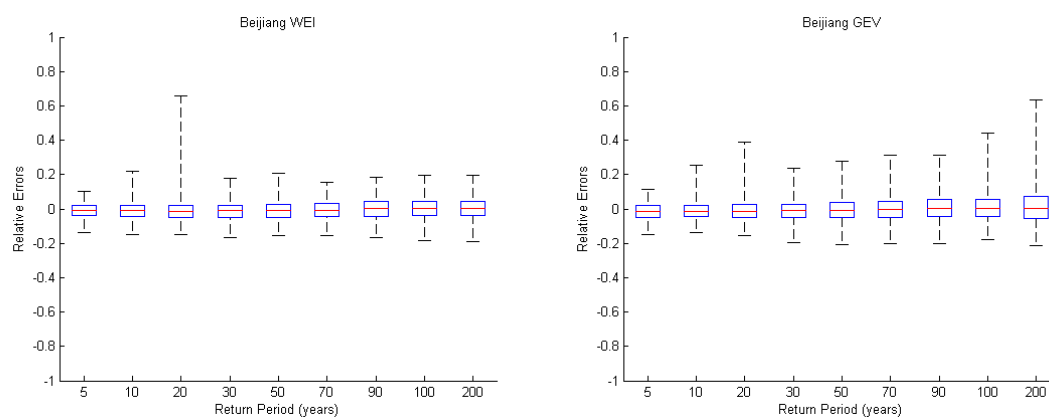


Figure 6. Box plots of the relative errors (REs) of the Shijiao at Beijiag River for sample series length 53, with Kappa as the parent PD.

It should be noted that hypothesis tests and information-based criteria methods all give unsatisfactory performance in Huai River (Table 6 and Figure 7); Weibull can be viewed as the preferable flood frequency distribution in Huai River by the results of composite criterion. Its large C_s value of 1.198 is not consistent with the selection of light tail Weibull distribution, mainly because the influence of the extremely large flood in 1954.

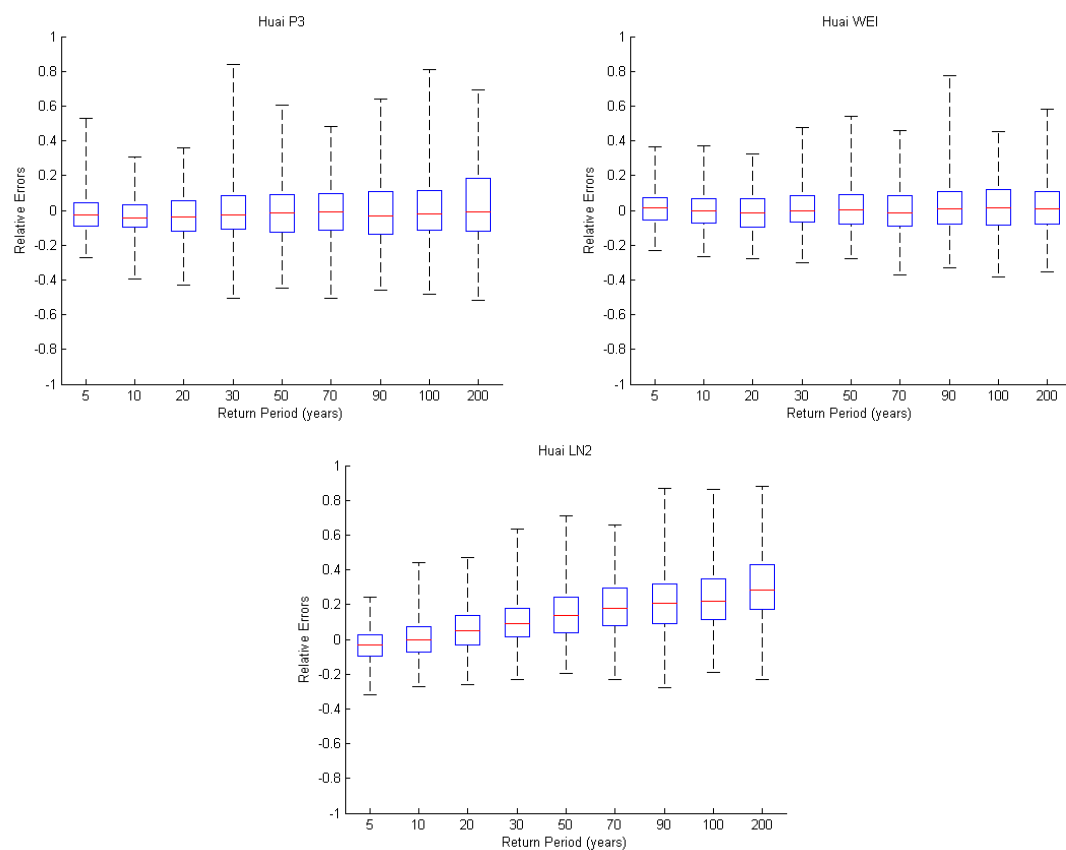


Figure 7. Box plots of the relative errors (REs) of the Lutaizi at Huai River for sample series length 48, with Kappa as the parent PD.

Table 6. $RMSE$ and \overline{RMSE} calculated for different return periods T in the four rivers.

River	PD	T = 5	T = 10	T = 20	T = 30	T = 50	T = 70	T = 90	T = 100	T = 200	\overline{RMSE}
Thames	GUM	0.039	0.044	0.049	0.047	0.050	0.048	0.052	0.052	0.059	0.049
	GLO	0.032	0.042	0.056	0.066	0.083	0.094	0.111	0.114	0.145	0.083
	GEV	0.038	0.045	0.055	0.063	0.076	0.088	0.089	0.093	0.126	0.075
Wabash	LN3	0.043	0.050	0.062	0.067	0.079	0.092	0.086	0.079	0.095	0.073
	P3	0.049	0.066	0.070	0.081	0.081	0.088	0.086	0.095	0.101	0.080
	WEI	0.043	0.046	0.052	0.058	0.064	0.075	0.075	0.076	0.088	0.064
	GEV	0.043	0.048	0.058	0.064	0.077	0.077	0.085	0.090	0.107	0.072
	GLO	0.041	0.045	0.059	0.068	0.092	0.110	0.129	0.136	0.176	0.095
Beijiang	WEI	0.046	0.051	0.061	0.055	0.057	0.060	0.063	0.063	0.064	0.058
	GEV	0.048	0.052	0.059	0.058	0.069	0.071	0.078	0.085	0.096	0.068
	P3	0.055	0.061	0.085	0.078	0.080	0.089	0.089	0.093	0.114	0.083
Huai	WEI	0.099	0.103	0.114	0.118	0.126	0.133	0.145	0.146	0.146	0.126
	P3	0.103	0.109	0.136	0.161	0.169	0.171	0.190	0.188	0.216	0.160
	LN2	0.097	0.107	0.138	0.166	0.214	0.245	0.273	0.294	0.369	0.211
	GLO	0.101	0.132	0.234	0.331	0.443	0.468	0.508	0.500	0.551	0.363

4.3. Comparison on Hypothesis Tests and Information-Based Criteria for Upper Tail

The objective of this section is to verify whether the hypothesis tests and information-based criteria work correctly for the upper tail of flood frequency distributions and to analyse the cause and the mechanism when they are applied to identify the PDs of hydrological extremes.

4.3.1. Characteristics of Statistical Hypothesis Test

(1) Kolmogorov–Smirnov (KS)

The KS test measures the greatest discrepancy between the observed and hypothesized distributions which locate at the upper tail or lower tail of the distribution. So the optimal PDs selected by KS are different from the ones selected by a composite criterion when the greatest discrepancy locates at the lower tail. The optimal PD selected by KS is not suitable for fitting high flow. For example, although the values of the KS test for GLO PD in Thames River, LN3 PD in Wabash River, and GLO PD in Huai River are considerably smaller than that of all the other PDs, these particular models overestimate or underestimate the upper tail events a great number of times. Furthermore, these particular distributions always have a rather wide spread of REs, with \overline{RMSE} value appreciably large (see Figures 4, 5, 7 and Table 6).

(2) Anderson–Darling Criterion (AD)

AD uses the sum of the squared differences between the empirical and theoretical distributions with weights to emphasize discrepancies in the tails. AD not only focuses on high flow end, but also addresses low flow end. Similar to KS, the optimal PD selected by AD is different from the one selected by a composite criterion when the emphasis is on the discrepancies located at the lower tail. For example, although the values of the AD test for GLO PD in Thames River and GLO PD in Wabash River are considerably smaller than that of all the other PDs, these models overestimate the upper tail events a greater number of times. Furthermore, these distributions always have a rather wide spread of REs, with \overline{RMSE} value appreciably large (Figures 4 and 5; Table 6). The optimal PDs selected by AD are never suitable for fitting high flows. In contrast, although GLO and LN3 do not perform so well at high flows, they fit the data well at the lower tail of the distribution, and these PDs are selected by AD in Wabash River as a final selection.

(3) Characteristics Summary

The statistical hypothesis tests (KS and AD) do not show rigorous results when focusing on the goodness of predictions of the extreme upper tail events. Although the values of the composite criterion for Gumbel PD in Thames River and Weibull PD in Wabash River show the best fitted distributions, the fitted order of Gumbel PD by KS and AD tests in Thames River is in the third place, and Weibull PD in Wabash River ranks fifth. Weibull PD selected by the composite criterion in Huai river ranks second by KS and AD. The results confirm some findings recently presented in the scientific literature. Laio et al. (2009) indicated that the statistical hypothesis testing methods have some evident limitations, because the obtained results are subjective, depending, for example, on the significance level chosen, and ambiguous, as often more than one distribution passes the goodness-of-fit tests [5].

4.3.2. Characteristics of Information-Based Criteria

(1) AIC, AICc Criteria

The optimal distributions selected by AIC and AICc are basically the same, and perform consistently with the distributions selected by the composite criterion. Although there are some differences in the values of AIC and AICc criteria for GEV PD in Beijiing River, they are considerably smaller than that of all the other PDs. However, these models overestimate the upper tail events when the return periods are greater than 70 years and underestimate the upper tail events for other return periods occasionally. The LN2 PD is selected by the AICc criterion in Huai River, however, LN2 PD sometimes overestimates the upper tail events and always has a rather wide range of REs, and with large \overline{RMSE} values (Figure 6 and Table 6).

(2) BIC Criterion

BIC is a Bayesian version of the AIC which incorporates some information about the prior distribution of the parameters of the model. BIC penalizes heavier than AIC and AICc for the number of estimated parameters P and small sample sizes [11]. So it is easier to select a distribution with fewer parameters, such as LN2 and Gumbel for the same sequence length. This is why the optimal distribution (LN2) selected by BIC does not perform consistently with the Weibull PD selected by the composite criterion. LN2 PD often overestimates the upper tail events (Figure 7). In addition, the BIC criterion often prefers the LN2 PD to AIC and AICc in Huai River, Thames River and Beijiing River, and prefers the Gumbel PD to AIC and AICc in Beijiing River and Huai River.

(3) Characteristics Summary

The optimal frequency distributions selected by AIC, BIC and AICc are basically the same as the distributions selected by the composite criterion. The information-based criteria are more sensitive to the high flow than hypothesis tests. BIC and AICc have a slight tendency towards the selection of two-parameter distributions. These results are due to the characteristics in penalizing for the number of estimated parameters P , by which BIC and AICc penalize heavier than AIC for small sample sizes. This is the reason that the optimal distribution (LN2) selected by information-based criteria does not perform consistently with the distribution selected by composite criterion (Weibull) in Huai River. This result confirms some findings recently presented in the scientific literature such as Baldassarre (2009) [12]. The capability of the information-based criteria to recognize the correct parent distribution from available data samples varies from case to case; it is rather good in some cases, in particular when the parent is a two-parameter distribution [5].

In general, the information-based criteria perform better than hypothesis tests when the focus is on the goodness of predictions of the extreme upper tail events. Although the order is not always ranked first for the best fitted distributions selected by the composite criterion, these distributions all can be identified correctly by AIC, BIC and AICc in all the four rivers. Furthermore, these particular distributions selected by information-based criteria always have a rather narrow spread of REs, with small \overline{RMSE} value. In contrast, the optimal frequency distributions for KS and AD are basically not the same as the distribution selected by the composite criterion. The reasons that information-based criteria are more sensitive to the high flow than hypothesis tests are as follows. The KS and AD criteria compare the distance of the flood point between theoretical and empirical frequencies. The closer the distance between the two, the better the model fitting degree. For the measured flood samples, small- and medium-level floods occur more frequently than big floods; the data for big floods at the upper tail of flood frequency distribution are scarce. Therefore, KS and AD may choose the distributions which focus on small- and medium-level floods (especially for the three-parameter distributions, because the fitting multi-parameter model can theoretically achieve good effect). This is different from the principle of information-based criteria, which do not compare the data distance between theoretical and empirical flood frequencies (distributions were selected on the basis of maximum likelihood values). Besides, information-based criteria can avoid over fitting and ensure the selection of the distribution which has a good epitaxial predictability by penalizing the model complexity. Furthermore, the value of the log-likelihood function can also reflect the goodness-of-fit of the probability model to observed points. The optimal distributions selected respectively by the KS and AD are often different. This can be easily seen from the results in Table 7 for Wabash River and Huai River. In contrast, the optimal frequency distributions selected respectively by AIC, BIC and AICc are basically the same. It is generally believed that AIC, BIC, AICc are stable for high flow in different rivers. In order to decide whether a particular distribution fits the high flow, it would be better to use the composite criterion which has the strongest applicability, followed by information-based criteria. The applicability of hypothesis tests is poor.

Table 7. The best fitted frequency distributions in the four rivers.

River	Hypothesis Tests				Information-Based Criteria					Composite Criterion
	KS	AD	KS, AD	Average Number of Parameters	AIC	BIC	AICc	AIC, BIC, AICc	Average Number of Parameters	
Kingston at Thames	GLO, GEV, GUM	GLO, GEV, GUM	Glo, Gev (Heavy or Mixed)	3	GUM, GLO, GEV	GUM, LN2, GLO	GUM, GLO, LN2	Gum,Glo (Mixed or Heavy)	2.5	Gum (Mixed)
Lafayette at Wabash	LN3, P3, GEV	GLO, LN3, P3	Ln3, P3 (Mixed or Light)	3	WEI, GEV, P3	WEI, GEV, P3	WEI, GEV, P3	Wei, Gev (Light or Mixed)	3	Wei (Light)
Shijiao at Beijiang	WEI, GEV, P3	WEI, GEV, P3	Wei, Gev (Light or Mixed)	3	GEV, WEI, P3	GEV, GUM, WEI	GEV, WEI, LN3	Gev, Wei (Mixed or Light)	3	Wei (Light)
Lutaizi at Huai	GLO, WEI, P3	P3, WEI, LN2	P3, Wei (Light)	3	WEI, LN2, P3	LN2, WEI, GUM	LN2, WEI, P3	Ln2, Wei (Mixed or Light)	2.5	Wei (Light)

5. Conclusions

In this study, eight probability distributions have been used for flood frequency analysis in four selected rivers with different climatic conditions, and their goodness-of-fit has been examined by various statistical methods. By applying all the distributions with different selection criteria for comparison to a composite criterion, the following conclusions are drawn.

- (1) There are different selections of frequency distributions in the four rivers by using hypothesis tests and information-based criteria approaches. Hypothesis tests are more likely to choose complex, parametric models, and information-based criteria prefer to choose simple, effective models. Different selection criteria have no particular tendency toward the tail of the distribution.
- (2) The information-based criteria perform better than hypothesis test methods most of the time when focusing on the goodness of predictions of the extreme upper tails of PDs. The distributions selected by information-based criteria are more likely to be close to true values than the distributions selected by hypothesis test methods in the upper tail of the frequency curve.
- (3) The composite criterion not only can select the optimal distribution, but also can evaluate the error of the estimated value. In order to decide on a particular distribution to fit the high flow, it would be better to use the composite criterion.

Acknowledgments: The research is financially supported by the National Natural Science Foundation of China (Grant No.51210013, 51569009, 51479216, 51509127), the National Science and Technology Support Program (Grant No 2012BAC21B0103), the project from Natural science foundation of Hainan (414192, 20164157), the Public Welfare Project of Ministry of Water Resources (Grant No201401048). We also thank two anonymous referees for their valuable comments which helped the improvement of the manuscript.

Author Contributions: Xiaohong Chen designed the content and ideas of the research. Quanxi Shao and Chong-Yu Xu revised and improved the quality of the research, Jiaming Zhang, Lijuan Zhang and Changqing Ye calculated the data, Changqing Ye analysed the results. All authors reviewed the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hussain, Z.; Pasha, G.R. Regional flood frequency analysis of the seven sites of Punjab, Pakistan, using L-Moments. *Water Resour. Manag.* **2009**, *23*, 1917–1933. [[CrossRef](#)]
2. Chérif, R.; Bargaoui, Z. Regionalisation of Maximum Annual Runoff Using Hierarchical and Trellis Methods with Topographic Information. *Water Resour. Manag.* **2013**, *27*, 2947–2963. [[CrossRef](#)]
3. Faulkner, D.; Keef, C.; Martin, J. Setting design inflows to hydrodynamic flood models using a dependence model. *Hydrol. Res.* **2012**, *43*, 663–674. [[CrossRef](#)]
4. Xia, J. Identification of a constrained nonlinear hydrological system described by Volterra Functional Series. *Water Resour. Res.* **1991**, *27*, 2415–2420.

5. Laio, F.; Baldassarre, D.G.; Montanari, A. Model selection techniques for the frequency analysis of hydrological extremes. *Water Resour. Res.* **2009**, *45*. [[CrossRef](#)]
6. Gomes, O.; Combes, C.; Dussauchoy, A. Parameter estimation of the generalized gamma distribution. *Math. Comput. Simul.* **2008**, *79*, 955–963. [[CrossRef](#)]
7. Cicioni, G.; Guiliano, G.; Spaziani, F.M. Best fitting of probability functions to a set of data for flood studies. In Proceedings of the 2nd International Symposium on Hydrology of Floods and Droughts, Fort Collins, CO, USA, 11–13 September 1972; Water Resource Publication: Fort Collins, CO, USA, 1973; pp. 304–314.
8. Akaike, H. Information theory and an extension of the maximum likelihood principle. In Proceeding of the International Symposium on Information Theory, Budapest, Hungary, 2–8 September 1971; Peter, B.N., Csaki, F., Eds.; Akademiai Kiado: Budapest, Hungary, 1973; pp. 267–281.
9. Hosking, J.R.M.; Wallis, J.R. The Value of Historical Data in Flood Frequency Analysis. *Water Resour. Res.* **1986**, *22*, 1606–1612. [[CrossRef](#)]
10. Haktanir, T.; Horlacher, H.B. Evaluation of various distributions for flood frequency analysis. *Hydrol. Sci. J.* **1993**, *38*, 15–32. [[CrossRef](#)]
11. Haddad, K.; Rahman, A.; Stedinger, J.J.R. Regional flood frequency analysis using Bayesian generalized least squares: A comparison between quantile and parameter regression techniques. *Hydrol. Process.* **2012**, *26*, 1008–1021. [[CrossRef](#)]
12. Baldassarre, G.D.; Laio, F.; Montanari, A. Design flood estimation using model selection criteria. *Phys. Chem. Earth* **2009**, *34*, 606–611. [[CrossRef](#)]
13. Calenda, G.; Mancini, C.P.; Volpi, E. Selection of the probabilistic model of extreme floods: The case of the River Tiber in Rome. *J. Hydrol.* **2009**, *371*, 1–11. [[CrossRef](#)]
14. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodal Inference: A Practical Information-Theoretic Information-Theoretic Approach*, 2nd ed.; Springer-Verlag: New York, NY, USA, 2002.
15. Önöz, B.; Bayazit, M. Best-fit distributions of largest available flood samples. *J. Hydrol.* **1995**, *167*, 195–208. [[CrossRef](#)]
16. House, P.K.; Baker, V.R. Paleohydrology of flash floods in small desert watersheds in western Arizona. *Water Resour. Res.* **2001**, *37*, 1825–1839. [[CrossRef](#)]
17. Hosking, J.R.M.; Wallis, J.R. Paleoflood Hydrology and Flood Frequency Analysis. *Water Resour. Res.* **1986**, *22*, 543–550. [[CrossRef](#)]
18. Luo, P.P.; He, B.; Takara, K.; APIP; Nover, D.; Kobayashi, K.; Yamashiki, Y. Paleo-hydrology and Paleo-flow Reconstruction in the Yodo River Basin. *Annu. Disas. Prev. Res. Inst. Kyoto Univ.* **2011**, *54*, 119–128.
19. Begueria, S.; Angulo-Martinez, M.; Vicente-Serrano, S.M.; Lopez-Moreno, J.I.; El-Kenawy, A. Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: A case study in northeast Spain from 1930 to 2006. *Int. J. Climatol.* **2011**, *31*, 2102–2114. [[CrossRef](#)]
20. Magilligan, F.J.; Nislow, K.H. Changes in hydrologic regimes by dams. *Geomorphology* **2005**, *71*, 61–78. [[CrossRef](#)]
21. Salas, J.D.; Obeysekera, J. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *J. Hydrol. Eng.* **2014**, *19*, 554–568. [[CrossRef](#)]
22. Milly, P.C.D.; Betancourt, J.B.; Falkenmark, M.; Hirsch, R.M.; Kundzewicz, Z.W.; Lettenmaier, D.P.; Stouffer, R.J. Stationarity is dead: Whither water management? *Science* **2008**, *319*, 2. [[CrossRef](#)] [[PubMed](#)]
23. Khaliq, M.N.; Ouara, T.B.M.J.; Ondo, J.C.; Gachon, P.; Bobée, B. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *J. Hydrol.* **2006**, *329*, 534–552. [[CrossRef](#)]
24. Chen, X.H.; Zhang, L.J.; Xu, C.-Y.; Zhang, J.M.; Ye, C.Q. Hydrological design of non-stationary flood extremes and durations in Wujiang river, South China: Changing properties, causes and impacts. *Math. Probl. Eng.* **2013**, *2013*. [[CrossRef](#)]
25. Yan, L.; Xiong, L.H.; Liu, D.D.; Hu, T.S.; Xu, C.-Y. Frequency analysis of nonstationary annual maximum flood series using the time-varying two-component mixture distributions. *Hydrol. Process.* **2017**, *31*, 69–89. [[CrossRef](#)]
26. Rao, A.R.; Hamed, K.H. *Flood Frequency Analysis*; CRC Press: Boca Raton, FL, USA, 2000; pp. 10–12.
27. Reiss, R.-D.; Thomas, M. *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*; Birkhäuser: Basel, Switzerland, 2001; pp. 49–50.

28. El Adlouni, S.; Bobée, B.; Ouarda, T.B.M.J. On the tails of extreme event distributions in hydrology. *J. Hydrol.* **2008**, *355*, 16–33. [[CrossRef](#)]
29. Reis, D.D.S.; Stedinger, J.J.R. Bayesian MCMC flood frequency analysis with historical information. *J. Hydrol.* **2005**, *313*, 97–116. [[CrossRef](#)]
30. Huang, W.R.; Xu, S.D.; Nnaji, S. Evaluation of GEV model for frequency analysis of annual maximum water levels in the coast of United States. *Ocean Eng.* **2008**, *35*, 1132–1147. [[CrossRef](#)]
31. Martins, E.S.; Stedinger, J.R. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.* **2000**, *36*, 737–744. [[CrossRef](#)]
32. Hirose, H. Maximum likelihood estimation in the 3-parameter Weibull distribution: A look through the generalized extreme-value. *IEEE Trans. Dielectr. Electr. Insul.* **1996**, *3*, 43–55. [[CrossRef](#)]
33. Otten, A.; Montfort, V. Maximum-likelihood estimation of the general extreme-value distribution parameters. *J. Hydrol.* **1980**, *47*, 187–192. [[CrossRef](#)]
34. Yevjevich, V. *Probability and Statistics in Hydrology*; Water Resources Publications: Fort Collins, CO, USA, 1972; pp. 214–232.
35. Ben-Zvi, A. Rainfall intensity–duration–frequency relationships derived from large partial duration series. *J. Hydrol.* **2009**, *367*, 104–114. [[CrossRef](#)]
36. Laio, F. Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resour. Res.* **2004**, *40*, W09308. [[CrossRef](#)]
37. Burnham, K.P.; Anderson, D.R. Multimodel Inference: Understanding AIC and BIC in model selection. *Sociol. Methods Res.* **2004**, *33*, 261–304. [[CrossRef](#)]
38. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [[CrossRef](#)]
39. Hosking, J.R.M.; Wallis, J.R. *Regional Frequency Analysis: An Approach Based on L-moments*; Cambridge University Press: Cambridge, UK, 1997; pp. 19–20.
40. Xie, P.; Lei, H.F.; Chen, G.C.; Li, J. A Spatial and Temporal Variation Analysis Method of Watershed Rainfall Based on Hurst Coefficient. *J. China hydrol.* **2008**, *28*, 6–10.
41. Wallis, J.R.; Matalas, N.C. Small sample properties of H and K—Estimations of the Hurst coefficient h. *Water Resour. Res.* **1970**, *6*, 1583–1594. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).