

Research Article

Uncertainty Assessment of Hydrological Frequency Analysis Using Bootstrap Method

Yi-Ming Hu,^{1,2} Zhong-Min Liang,^{1,2} Bin-Quan Li,^{1,2} and Zhong-Bo Yu^{1,2,3}

¹ State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

² College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China

³ Department of Geoscience, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Correspondence should be addressed to Zhong-Min Liang; zmliang@hhu.edu.cn

Received 29 December 2012; Accepted 8 May 2013

Academic Editor: Yongping Li

Copyright © 2013 Yi-Ming Hu 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.

The hydrological frequency analysis (HFA) is the foundation for the hydraulic engineering design and water resources management. Hydrological extreme observations or samples are the basis for HFA; the representativeness of a sample series to the population distribution is extremely important for the estimation reliability of the hydrological design value or quantile. However, for most of hydrological extreme data obtained in practical application, the size of the samples is usually small, for example, in China about 40~50 years. Generally, samples with small size cannot completely display the statistical properties of the population distribution, thus leading to uncertainties in the estimation of hydrological design values. In this paper, a new method based on bootstrap is put forward to analyze the impact of sampling uncertainty on the design value. By bootstrap resampling technique, a large number of bootstrap samples are constructed from the original flood extreme observations; the corresponding design value or quantile is estimated for each bootstrap sample, so that the sampling distribution of design value is constructed; based on the sampling distribution, the uncertainty of quantile estimation can be quantified. Compared with the conventional approach, this method provides not only the point estimation of a design value but also quantitative evaluation on uncertainties of the estimation.

1. Introduction

Estimations of flood frequency and magnitude are required for hydraulic engineering design, water resources planning, and management, for example, [1]. They involve the sampling of a sample series, the selection of a population distribution represented usually by a probability density function (pdf) where the samples are coming from, and the estimation of parameters of the pdf. The occurrences of various sources of errors in the above-mentioned procedure result in the existence of uncertainties in the final estimation of the hydrological design value. Among these errors, the sampling error, that is, an error due to the fact that the short of sample length is always taken as the main contribution to the significance of estimation uncertainties.

During the past several decades, studies on uncertainty analysis of the estimation for hydrological design value

mainly have focused on the selection of the type of a pdf and the estimation of parameters of the pdf. The conventional method of moment (MOM) was once a widely used approach for parameter estimation, but Wallis et al. [2] and Kirby [3] showed that MOM bears high bias and is subjected to algebraic bounds, which indicated that MOM may bring extra uncertainties to quantile estimations. In order to improve the accuracy of parameter estimation and reduce uncertainties on quantile estimates, various parameter estimation approaches have been developed such as maximum Likelihood estimators (ML), probability-weighted moment [4], and linear-moment method (L-M) [5]; the regional frequency analysis method, which combines the information of single site and hydrological similar neighboring sites to decrease the uncertainty of hydrological design values at single site, also have been applied [6, 7]. In the view of assessing the uncertainty of the hydrological frequency analysis, Wood and

Rodriguez-Iturbe [8, 9] analyzed the impact of parameter uncertainties on the uncertainty of the design value based on Bayes method. Later, they proposed a theoretical framework for the uncertainty assessment of design values by coupling uncertainties from both the selection of a population distribution and the estimation of distribution parameters. However, due to the limitation of computation technique, they only analyzed the uncertainties of the design values based on a few simple probability distributions, such as normal distribution and lognormal distribution. Merz and Thielen [10] analyzed natural and epistemic uncertainties in flood frequency analysis, and they found that the former cannot be reduced, but the latter can be reduced by more knowledge. Reis and Stedinger [11] applied a fully Bayesian approach to provide an accurate description of flood risk and parameter uncertainties for flood quantiles. Ribatet et al. [12] developed a regional Bayesian POT model to reduce the estimation uncertainty considering the fact that the size of the sample is small. Lee and Kim [13] employed the Bayesian Markov Chain Monte Carlo (MCMC) method and the maximum likelihood estimation (MLE) to assess uncertainties of low-flow frequency analysis. Zhongmin et al. [14, 15] applied the Bayesian theory to separately analyze the impact of parameters uncertainty and both frequency model and parameters uncertainties on the uncertainties of estimation of design value or quantile.

However, studies on the impact of the sample representativeness on the uncertainty of design value estimation have been seldom reported. In this paper, considering the fact that the small size sample cannot perfectly reflect the statistical properties of the population distribution and it may cause uncertainties on design value estimation, we proposed a new approach based on bootstrap technique to analyze the impact of sampling uncertainty on the design value. The bootstrap method has been widely used for uncertainty analysis in hydrology. For example, Zucchini and Adamson [16] applied the bootstrap method to estimate the confidence intervals of the design storms from exceedance series. Overeem et al. [17] applied the bootstrap method to analyze the uncertainty of rainfall depth-duration-frequency (DDF) curves and obtained 95% confidence bands of the DDF curves. They utilized the bootstrap-based artificial neural networks to investigate the uncertainty of flood forecasting.

In this paper, by bootstrap technique, resampling from the original flood extreme observations, a large number of bootstrap samples are constructed, on which the corresponding design value or quantile is estimated for each bootstrap sample, and then the distribution of estimation of the design value with some probabilities can be obtained. This method obtains the sampling distributions of design values; therefore, it can afford not only the point estimation as a conventional method does but also the quantitative evaluation on uncertainty of hydrological frequency analysis. As an example, annual precipitation series in the Kunming municipal area, China was applied to validate the proposed approach in this paper.

Throughout the following sections of this paper, the bootstrap method will be described first, and then the method considering the impact of the sample representativeness on

the uncertainty of design value estimation will be outlined in Section 3. Then, Section 4 presents an example and result analysis. Finally, Section 5 presents concluding remarks.

2. Description of Bootstrap Method

The bootstrap method, a resampling technique, was introduced and named by Efron [18, 19]. The idea behind bootstrap is that sample values are the best guide to the underlying true distribution even when the information about the true distribution is lacking. It does not need the assumption of the true distribution and only depends on the obtained sample values. N groups of bootstrap samples can be generated by resampling from the original sample repeatedly. Then, based on these bootstrap samples, N estimations of the statistical parameter (mean, variance, etc.) can be derived. Finally, the distribution of the parameter is then approximated by N estimated values.

Resampling from the original sample is the essence of the bootstrap method. Assuming that $X = (x_1, x_2, \dots, x_n)$ is an original sample; based on the sample, the empirical distribution F_n can be described as follows:

$$F_n = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_{(k)} \leq x < x_{(k+1)}, \\ 1, & x \geq x_{(k)}. \end{cases} \quad (1)$$

By resampling from the distribution F_n , the same size sample $X^* = (x_1^*, x_2^*, \dots, x_n^*)$ can be obtained. Based on the bootstrap sample X^* , an estimation θ^* of parameter θ of the distribution function can be calculated.

Repeating the bootstrap sampling for N times, N groups of bootstrap samples can be obtained, $X^{*(j)} = (x_1^{*(j)}, x_2^{*(j)}, \dots, x_n^{*(j)})$, $j = 1, 2, \dots, N$, and N estimations of parameter θ can be derived, $\theta^{*(j)}$, $j = 1, 2, \dots, N$.

Taking the $\theta^{*(j)}$, $j = 1, 2, \dots, N$ as the sample of the unknown parameter θ , we can obtain the distribution of parameter θ . Based on the distribution, point estimation or interval estimation of parameter θ can be derived.

3. Considering the Impact of Sample Representativeness on Design Value Uncertainty

In the procedure of hydrological frequency analysis, numbers of extreme events or sample sizes are usually not large enough, so that estimations of hydrological design values based on the limited sample information inevitably come into being uncertainties. How to assess the uncertainties is a key issue because they are relevant to the safety of hydraulic engineering works. In this section, bootstrap method is applied to resample from the original sample to derive the distribution of estimation of design value corresponding to some nonexceedance probabilities, and based on the distribution, the impact of sampling uncertainty on hydrological frequency analysis results can be assessed.

In China, Person-type three probability distributions (P -III) are widely used for hydrological frequency analysis. A P -III function contains three parameters, that is, mean value E_x , coefficient of deviation C_v , and coefficient of skew C_s . Assuming that $X = (x_1, x_2, \dots, x_n)$ is a hydrological extreme sample series, E_x, C_v, C_s represent the three population parameters and x_p the design value or quantile of the population distribution corresponding to a nonexceedance probability p ; $\hat{E}_x^{(j)}, \hat{C}_v^{(j)}, \hat{C}_s^{(j)}$ denote the estimations of E_x, C_v, C_s , respectively, by the j th resampled series, and \hat{x}_p is the estimation of x_p . The bootstrap procedure for assessing the impact of a sample on the uncertainty of quantile estimation can be described as follows.

- (1) Resampling from the original sample $X = (x_1, x_2, \dots, x_n)$ for N time, N groups of bootstrap samples $X^{*(j)} = (x_1^{*(j)}, x_2^{*(j)}, \dots, x_n^{*(j)}), j = 1, 2, \dots, N$ can be obtained.
- (2) Using the linear-moment method (L-M) to estimate parameters E_x, C_v, C_s , gaining N estimations $\hat{E}_x^{(j)}, \hat{C}_v^{(j)}, \hat{C}_s^{(j)}, j = 1, 2, \dots, N$.
- (3) Based on the N groups of $\hat{E}_x^{(j)}, \hat{C}_v^{(j)}, \hat{C}_s^{(j)}, j = 1, 2, \dots, N$, the N estimations of design value or quantile $\hat{x}_p^{(j)}, j = 1, 2, \dots, N$ can be acquired.
- (4) Taking the $\hat{x}_p^{(j)}, j = 1, 2, \dots, N$ as the sampling series of x_p , we can obtain the sampling distribution of x_p , on which point estimation or interval estimation of x_p can be derived.

4. Case Study

4.1. Study Region and Data. The annual precipitation series from 1951 to 2010 in Kunming municipal area was applied to validate the proposed method in this paper. Kunming, capital of Yunnan province, is located in the middle of the Yunnan-Guizhou Plateau in southwest China, with a latitude of 24.23° ~ 26.22° N and a longitude of 102.10° ~ 103.40° E. The city area is about 21473 km^2 . It belongs to the subtropical monsoon climate, mild and humid, with four distinct seasons. In the history, its highest and lowest extreme temperatures are 31.2°C and -7.8°C , with annual average temperature about 15°C . Annual precipitation distribution is uneven; that is, precipitation in the rainy season from May to October accounts for about 85% of the 1000 mm average annual precipitation, while precipitation in the dry season from November to April accounts for only about 15%.

4.2. Precipitation Series Analysis. Considering the fact that frequency analysis is based on the assumption that the hydrological time series is stationary, statistically independent and identically distributed (iid), trend test and change-point detection for the annual precipitation series of Kunming city are necessary. The annual precipitation time series was presented in Figure 1. From Figure 1, the linear regression trend of precipitation time series has not been detected, which shows that the series is stationary.

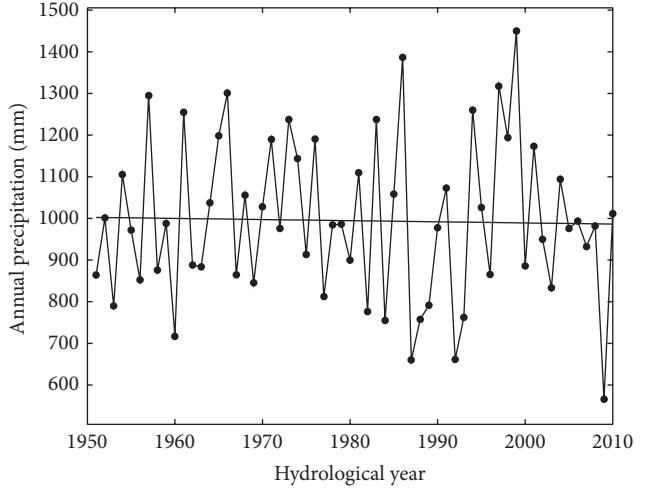


FIGURE 1: Temporal trend of annual precipitation in Kunming municipal area from 1951 to 2010: the straight line is the trend line of the observation series.

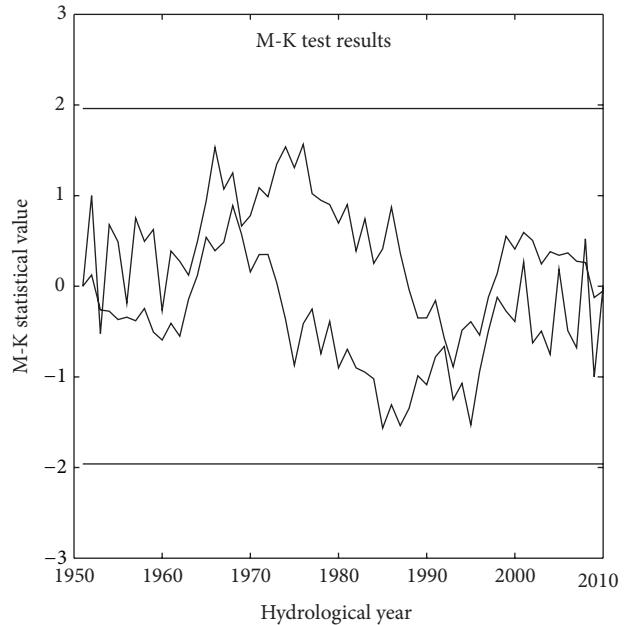


FIGURE 2: Trend detection by Mann-Kendall test for the annual precipitation time series of Kunming, China: the straight lines represent the boundary condition of the M-K test with the significance level $\alpha = 0.05$, and the related threshold values are ± 1.96 . The sample M-K statistical value is between -1.96 and 1.96 , which implies that the annual precipitation series does not have a mutation trend.

In order to further analyze the mutation trend, the Mann-Kendall (M-K) nonparametric test method, widely used for trend test and change-point detection, was applied [20, 21]. The M-K test results were shown in Figure 2. In the M-K test, the significance level $\alpha = 0.05$, and the threshold values of the statistical test of Mann-Kendall are ± 1.96 . Figure 2 shows that the sample M-K statistical value is between -1.96 and 1.96 . It implies that the annual precipitation series does not

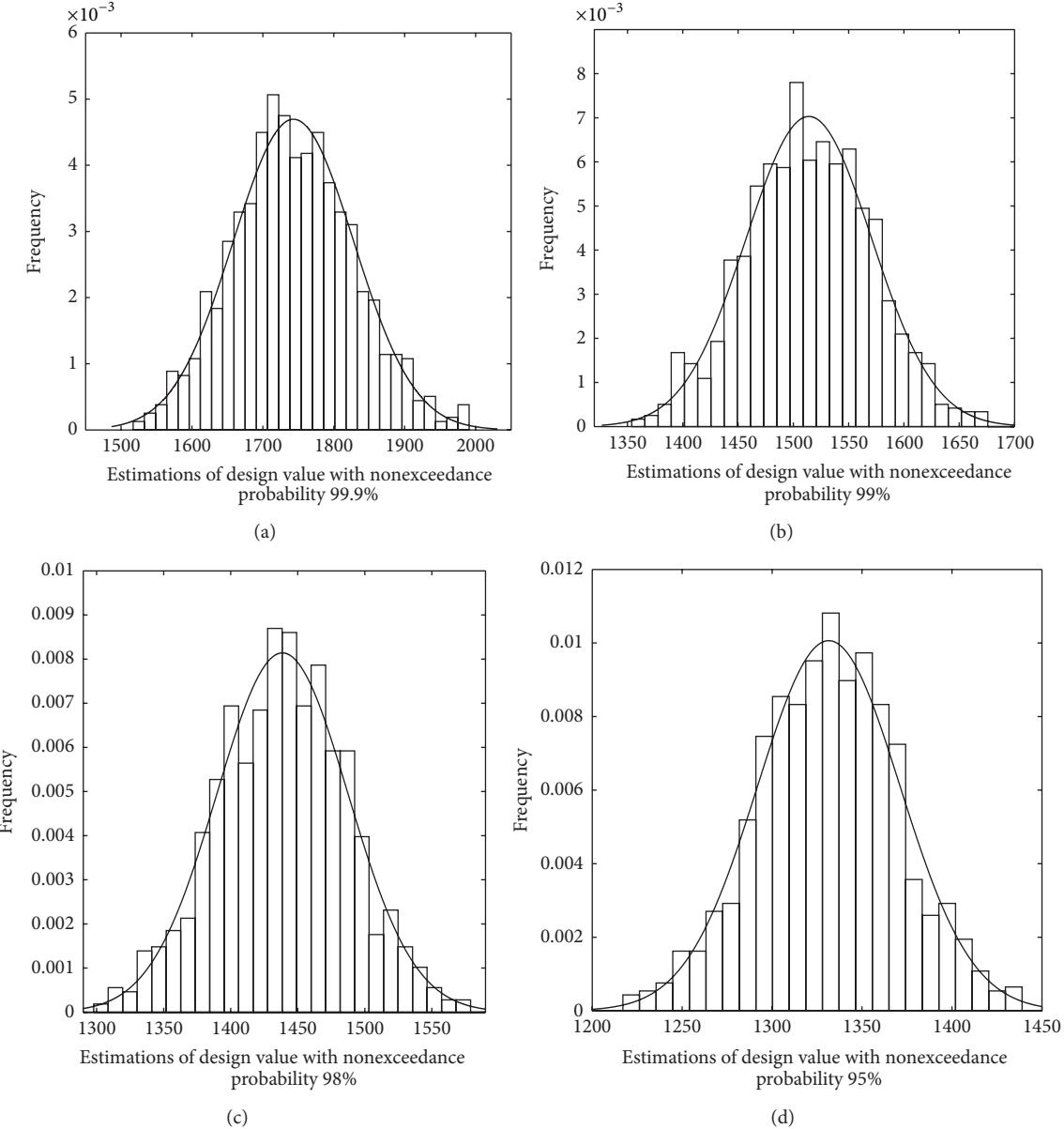


FIGURE 3: Frequency histogram of estimations of the design value x_p with nonexceedance probabilities 99.9%, 99%, 98%, and 95%. The curve is normal density function.

have a mutation trend. The analyses above indicate that the precipitation time series can be taken as a stationary and iid sample.

4.3. Design Value Estimation Considering Sample Uncertainty. According to Section 3, by resampling from the annual precipitation series, N groups of bootstrap samples can be obtained. Considering the fact that the lower limit of a Pearson-III distribution is a negative number when the value C_s/C_v is less than 2.0; in the process of resampling, those bootstrap samples whose value C_s/C_v is equal or greater than 2.0 were selected, and the total sets are $N = 1000$. L-M method was applied for estimations of the parameters E_x , C_s , C_v , and N estimations $\hat{E}_x^{(j)}$, $\hat{C}_v^{(j)}$, $\hat{C}_s^{(j)}$, $j = 1, 2, \dots, N$

can be obtained. In combination with the Pearson-III distribution, N estimations $\hat{x}_p^{(j)}$, $j = 1, 2, \dots, N$ of the design value x_p corresponding to nonexceedance probability P can be calculated. By the N estimations $\hat{x}_p^{(j)}$, $j = 1, 2, \dots, N$, the distribution of x_p can be approximated. Figure 3 presented the frequency histograms of the design values x_p with $P = 99.9\%$, 99%, 98%, and 95%. Intuitively from these figures, a normal distribution may match well with the frequency histograms.

Normal probability plot was used to test whether the sampling distribution of x_p matches with a normal distribution. Figures 3 and 4 demonstrate, respectively, the fitting of a normal frequency curve to the empirical frequency data of $\hat{x}_p^{(j)}$, $j = 1, 2, \dots, N$, it is shown that plots uniformly distribute

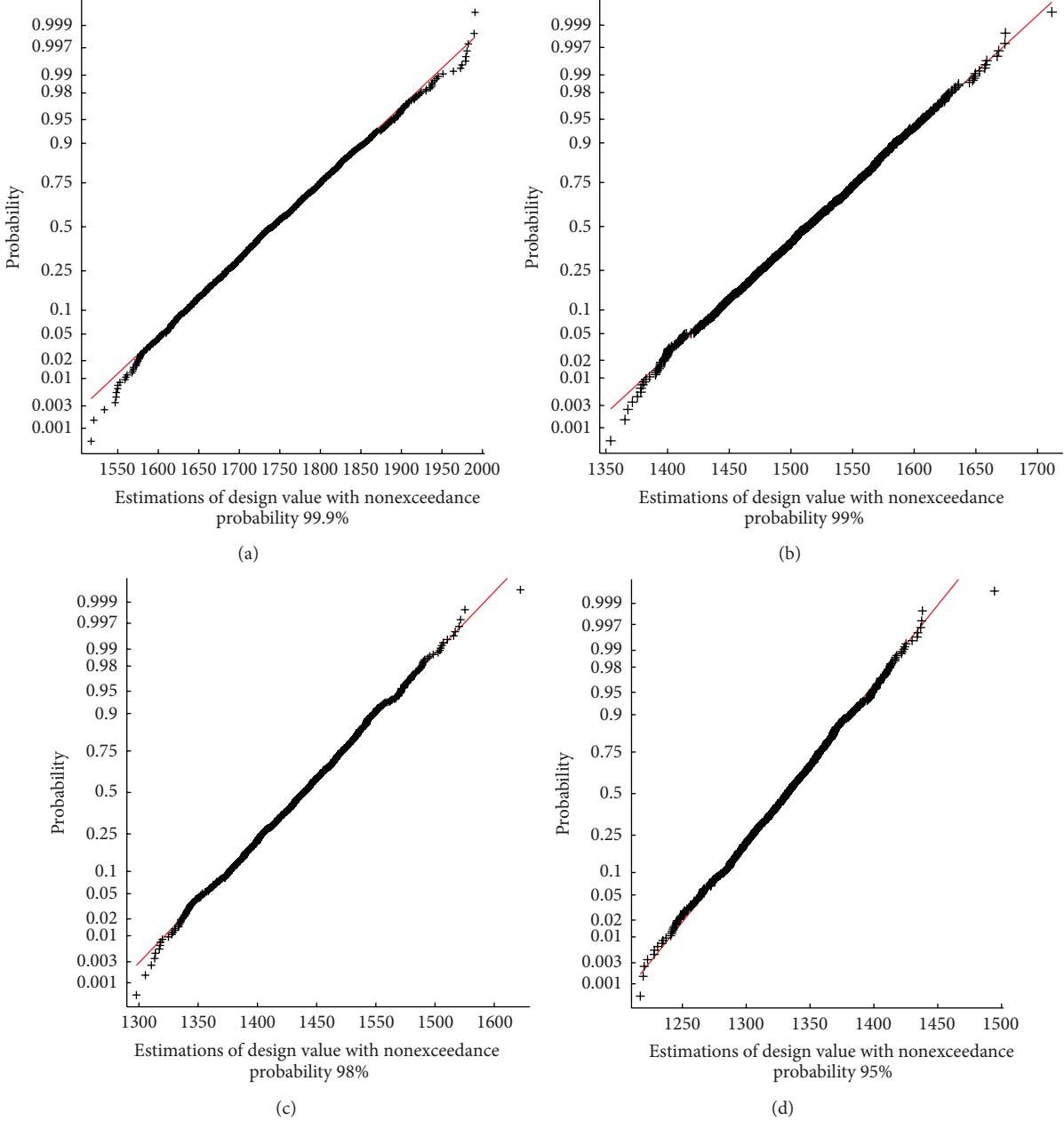


FIGURE 4: Normal probability plot of estimations of the design value x_p with nonexceedance probabilities 99.9%, 99%, 98%, and 95%: the symbol “+” is normal probability plot reflecting the relationship of the estimation of design value and the related normal quantile; the straight line is the fitted line of the normal probability plots; from Figure 4, it can be seen that plots uniformly distribute in the vicinity of the straight line, which means that the sampling distribution of design value x_p can be approximated by a normal probability distribution.

in the vicinity of the normal frequency line. Therefore, the sampling distribution of design value x_p can be approximated by a normal probability distribution. Similarly, sampling distributions of x_p with any nonexceedance probability p can be estimated.

Upon attaining the sampling distribution of x_p , hydrological frequency analysis based on the sampling distributions can be conducted as illustrated by Figure 5. Different from the conventional hydrological frequency analysis that provides

only a point or a unique estimation of x_p , the method proposed in this study provides the sampling distribution of x_p . That is to say, for a specific unknown population x_p , there exists its infinite estimates, \hat{x}_p , and these estimates form the sampling distribution of x_p . It is this distribution that provided the entire information about x_p ; accordingly, the estimation or inference on x_p should rely on this distribution. For example, the expectation value of the sampling distribution Ex_p can replace the point estimation of any conventional

TABLE 1: Estimation of design value x_p .

Nonexceedance Probability p (%)	Sampling distribution of x_p	Bootstrap method			Curve-fitting method		
		Ex_p (mm)	$CI_{90\%}$	CIL	SD	Point estimation (mm)	
99.9	$X_{99.9\%} \sim N(1748, 89)$	1748	[1601, 1895]	294	89	1704	
99	$X_{99\%} \sim N(1517, 59)$	1517	[1420, 1615]	195	59	1497	
98	$X_{98\%} \sim N(1441, 51)$	1441	[1357, 1525]	168	51	1427	
95	$X_{95\%} \sim N(1333, 41)$	1333	[1265, 1401]	136	41	1328	

Notation: Ex_p is the expectation of the sampling distribution of x_p , and SD is its standard deviation; $CI_{90\%}$ is the 90% confidence intervals, and CIL_{90%} is its length.

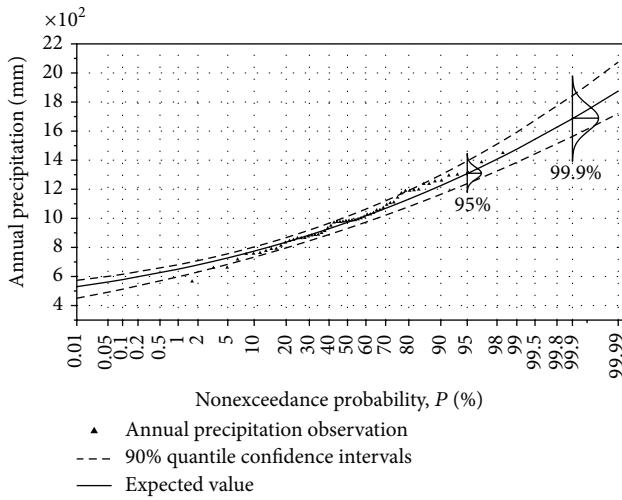


FIGURE 5: Estimation of x_p by bootstrap method for annual precipitation in Kunming municipal area: the first small figure denoted that 95% is the distribution of estimations of the design value with the nonexceedance 95%; the second small figure denoted that 99.9% is the distribution of estimations of the design value with the nonexceedance 99.9%.

method, while the confidence interval can be used to evaluate the uncertainties of hydrological frequency analysis with the given original samples.

The characteristics of the sampling distribution of the design value x_p for the 4 specific nonexceedance probabilities are summarized in Table 1, including the expectation Ex_p , standard deviation SD, and 90% confidence intervals $CI_{90\%}$ and its length $CIL_{90\%}$. For comparison, the corresponding point estimations of x_p from a conventional approach, namely, curve-fitting method, are also included. The curve-fitting method is a widely applied procedure in China for hydrological frequency analysis. From Table 1, it seems that there is no big gap between the expectation value Ex_p of sampling distribution and the point estimation of curve-fitting method for the four cases of nonexceedance probabilities. However, compared to the curve-fitting approach, the bootstrap method affords more information about the estimation. For example, with the increase of design standard or nonexceedance probability, the value of SD and $CIL_{90\%}$ is getting larger, ranging from 41 to 89 and from 136 to 294, respectively, which implies that the uncertainty of the

estimation of design value with the rare probability will be much larger, so that it is necessary to provide quantitative uncertainty assessment for the estimation of the design value.

5. Summary and Conclusion

The sample size of hydrological extreme events in practical application is usually not large enough to guarantee that a reliable estimation is always obtained. Hence, for a given sample series, how to improve the estimation and evaluate quantitatively the uncertainty of the estimation is of use to the design of hydraulic engineering works. In this study, a new method based on bootstrap is put forward to analyze the impact of sampling uncertainty on the design value, which is not restricted to the probability models. Through resampling from the original sample series, the sampling distribution of the design value x_p or quantile is available, by which x_p could be estimated in a new manner. Compared to the conventional hydrological frequency analysis methods, such as the curve-fitting method, the proposed bootstrap-based method can provide not only the point estimation of x_p as the curve-fitting does but also afford more abundant information about the estimation; the uncertainties of the estimation on x_p could also be assessed by either the standard deviation SD of the sampling distribution or the confidence interval CI derived from the sampling distribution.

As an example, annual precipitation data in Kunming municipal area were used to validate the proposed method in this paper. As illustrated by Figure 5, the sampling distribution of x_p for any nonexceedance probability p was estimated by resampling, on which the expectation values of the sampling distribution can be used as the conventional point estimation, while the confidence interval can be the indicator to evaluate uncertainties of the estimation. Comparisons between the bootstrap and curve-fitting approaches (see Table 1) for 4 specific cases of $P = 99.9\%, 99\%, 98\%$, and 95% indicate that, notwithstanding, there is almost no gap between point estimations of curve fitting and expectation values of bootstrap-based approach; the other features of the sampling distribution, such as the SD and $CI_{90\%}$ and its length $CIL_{90\%}$, provide the designers with extra information to assess the reliability of the design values to be adopted.

As we all know, uncertainties of estimation on design value x_p originate unreliabilities from the sample, the statistical model, and the parameter; this study addressed only the impact of sampling on the estimation uncertainties of x_p .

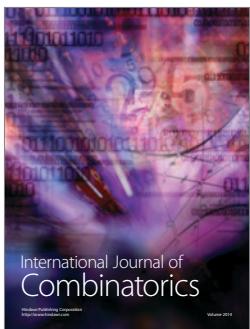
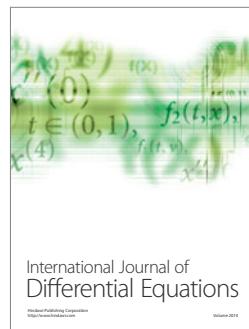
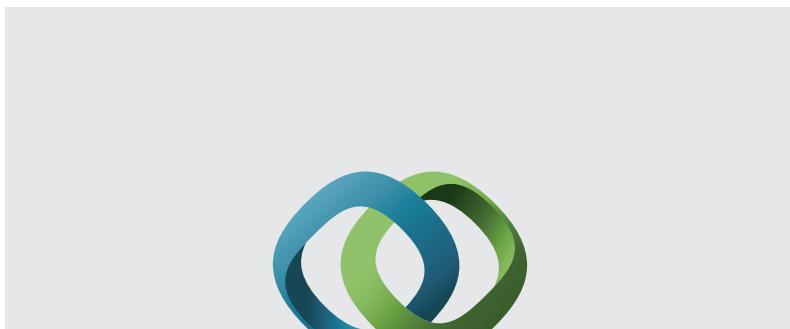
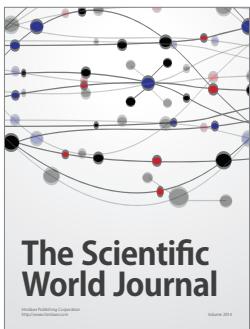
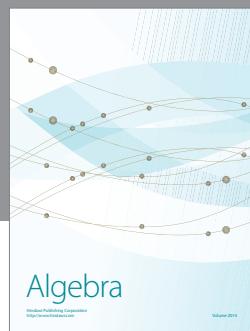
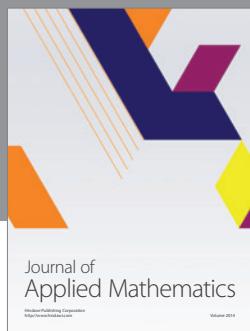
Although the bootstrap-based method was performed to P -III probability distribution, the procedure is capable of being applied to any other probability models. In this paper, due to the restriction of resampling condition, that is, considering the fact that the lower limit of a Pearson-III distribution is a negative number when the value C_s/C_v is less than 2.0, in the process of resampling, those bootstrap samples whose value C_s/C_v is equal or greater than 2.0 were selected, and the conclusions may not be applicable for other probability models because they do not have this restriction.

Acknowledgments

This study was supported by the Major Program of the National Natural Science Foundation of China (Grant no. 51190095), the National Basic Research Program of China (also called 973 Program, Grant no. 2010CB951102), and the National Natural Science Foundation of China (Grant no. 51079039 and 51179046). The authors wish to thank the referee, whose suggestions help them to improve the paper.

References

- [1] J. R. Stedinger, R. M. Vogel, and E. Foufoula-Georgiou, "Frequency analysis of extreme events," in *Handbook of Hydrology*, R. Maidment, Ed., chapter 18, pp. 18. 1–18. 66, McGraw-Hill, New York, NY, USA, 1992.
- [2] J. R. Wallis, N. C. Matalas, N. C. Matas, and J. R. Slack, "Just a moment!," *Water Resources Research*, vol. 10, pp. 211–219, 1974.
- [3] W. Kirby, "Algebraic boundedness of sample statistics," *Water Resources Research*, vol. 10, no. 2, pp. 220–222, 1974.
- [4] J. A. Greenwood, J. M. Landwehr, N. C. Matalas, and J. R. Wallis, "Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form," *Water Resources Research*, vol. 15, no. 5, pp. 1049–1054, 1979.
- [5] J. R. M. Hosking, "L-moments: analysis and estimation of distributions using linear combinations of order statistics," *Journal of the Royal Statistical Society B*, vol. 52, no. 1, pp. 105–124, 1990.
- [6] J. R. M. Hosking and J. R. Wallis, *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, New York, NY, USA, 1997.
- [7] G. Kuczera, "Combining site-specific and regional information: an empirical Bayes approach," *Water Resources Research*, vol. 18, no. 2, pp. 306–314, 1982.
- [8] E. F. Wood and I. Rodriguez-Iturbe, "Bayesian inference and decision making for extreme hydrological events," *Water Resources Research*, vol. 11, no. 4, pp. 533–554, 1975.
- [9] E. F. Wood and I. Rodriguez-Iturbe, "A Bayesian approach to analyzing uncertainty among flood frequency models," *Water Resources Research*, vol. 11, no. 6, pp. 839–843, 1975.
- [10] B. Merz and A. H. Thielen, "Separating natural and epistemic uncertainty in flood frequency analysis," *Journal of Hydrology*, vol. 309, no. 1–4, pp. 114–132, 2005.
- [11] D. S. Reis Jr. and J. R. Stedinger, "Bayesian MCMC flood frequency analysis with historical information," *Journal of Hydrology*, vol. 313, no. 1–2, pp. 97–116, 2005.
- [12] M. Ribatet, E. Sauquet, J.-M. Grésillon, and T. B. M. J. Ouarda, "A regional Bayesian POT model for flood frequency analysis," *Stochastic Environmental Research and Risk Assessment (SERRA)*, vol. 21, no. 4, pp. 327–339, 2007.
- [13] K. S. Lee and S. U. Kim, "Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method," *Hydrological Processes*, vol. 22, no. 12, pp. 1949–1964, 2008.
- [14] L. Zhongmin, L. Binquan, Y. Zhongbo et al., "Application of Bayesian approach to hydrological frequency analysis," *Science China Technological Sciences*, vol. 54, no. 5, pp. 1183–1192, 2011.
- [15] L. Zhongmin, C. Wenjun, and L. Binquan, "Bayesian flood frequency analysis in the light of model and parameter uncertainties," *Stochastic Environmental Research and Risk Assessment*, vol. 26, pp. 721–730, 2012.
- [16] W. Zucchini and P. T. Adamson, "Bootstrap confidence intervals for design storms from exceedance series," *Hydrological Sciences*, vol. 34, no. 6, pp. 719–721, 1989.
- [17] A. Overeem, A. Buishand, and I. Holleman, "Rainfall depth-duration-frequency curves and their uncertainties," *Journal of Hydrology*, vol. 348, no. 1–2, pp. 124–134, 2008.
- [18] B. Efron, "Bootstrap methods: another look at the jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.
- [19] B. Efron, "Computers and the theory of statistics: thinking the unthinkable," *SIAM Review*, vol. 21, no. 4, pp. 460–480, 1979.
- [20] K. H. Hamed, "Trend detection in hydrologic data: the Mann-Kendall trend test under the scaling hypothesis," *Journal of Hydrology*, vol. 349, no. 3–4, pp. 350–363, 2008.
- [21] M. Gocic and S. Trajkovic, "Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia," *Global and Planetary Change*, vol. 100, pp. 172–182, 2013.



Submit your manuscripts at
<http://www.hindawi.com>

