

## Evaluating the Three Methods of Goodness of Fit Test for Frequency Analysis

Xiankui Zeng, Dong Wang\*, Jichun Wu

Key Laboratory of Surficial Geochemistry, Ministry of Education, School of Earth Sciences and Engineering,  
Nanjing University, Nanjing, China.

\*Email: wangdong@nju.edu.cn

Received 30 June 2015

Accepted 9 August 2015

### Abstract

In hydrological statistics, the traditional assessment of goodness of fit test is interested in the testing precision to the sample generated from supposed PDF. However, the ability to reject the hypotheses when the supposed PDF is different from real PDF should be also emphasized. In addition, the sensitivity of test method to series length is important in hydrological analysis. Three methods of goodness of fit test that include Chi-Square (C-S), Kolmogorov-Smirnov (K-S), and Anderson-Darling (A-D) tests are applied in this study. The results of power test indicate that the most powerful tests for normal, uniform, P3, and Weibull distribution are K-S, C-S, and A-D tests, respectively. The test method with the best comprehensive power is C-S test, followed by K-S and A-D test.

*Keywords:* Test method, Frequency analysis, Chi-Square, Kolmogorov-Smirnov, Anderson-Darling

### 1 Introduction

Frequency analysis is an important method in hydrological statistics, which has been applied in uncertainty assessment of hydrological process extensively, such as flood control, drought resistance, and safety design of water conservancy project (Raju and Kumar, 2011; Saf, 2009; Xu et al., 2010; Lee et al., 2012). This method serves many purposes in statistical data analysis by fitting a probability distribution function (PDF) to a series of observations. Then, the fitted PDF provides summary information about the data, and allows predictions beyond the range of data used for parameter estimation. This prediction is based on the information of past events in order to define the probabilities of future events. Therefore, the prediction is attached a risk of failure. The frequency analysis does not predict the future with certainty, but it provides good models for explaining and making efficient use of the events that had occurred in the past (Khaliq et al., 2006; Smakhtin, 2001; Yang et al., 2010).

There are three steps in frequency analysis (Khaliq et al., 2006): (i) selecting a suitable probability distribution function (PDF); (ii) estimating the parameters of PDF based on samples; (iii) assessing the uncertainty of objective of interest in prescribed confidence level. Hereinto, how to select a suitable

PDF is the most crucial step for frequency analysis, because the following parameter estimation is influenced by the supposed PDF. Therefore, the probability distribution characteristics of objective and its uncertainty assessment are determined by the selected PDF. Thus, plenty of studies focus on determining the PDF that fits best the observed hydrological series. Beard (1974) estimated the flood records at 300 stations in the USA by 8 different models, and the results showed that log-P3 and log-normal fitted the observations best. By applying the available methods to flood records which lengths varied between 44 and 97 year, the US Water Resources Council (1976) recommended the use of log-P3 distribution, with a regional skew coefficient in the case of short records. The Flood Studies Report of the NERC (1975) showed that the generalized extreme value (GEV) distribution was recommended for the use in regional flood frequency analysis. Vogel et al., (1993) analyzed the flood data at 61 stations in Australia with records of 20 years or more. They found the best PDF at one region was even different for winter and summer-dominated rainfall regimes. Recently, Shabri and Ariff (2009) found that the generalized logistic distribution was the most suitable PDF to fit the data of maximum daily rainfalls for stations in Selangor and Kuala Lumpur.

Song and Singh (2010) analyzed the joint probability distribution of different hydrological periods in the Wei River basin, China. The results showed that the drought duration and inter-arrival time followed the Weibull distribution, and the drought severity followed the gamma distribution. Similar to mentioned literatures, many others researchers made the goodness of fit tests for different types of hydrological series, and the best PDFs were not the same at different regions or periods (e.g., Gunasekara and Cunnane, 1991; Onoz and Bayazit, 1995; Onoz and Bayazit, 2001; Bi et al., 2010; Lansigan, 2009; Sandoval and Raynal-Villasenor, 2008; Shao et al., 2004). Therefore, there is not universal applicability to the selection of best PDF, and it should be selected based on the practical observations and effective comparison and testing.

The choice of the suitable probability distribution function (PDF) is almost arbitrary as no physical basis is available to rationalize the use of any particular function. A common way is selecting one from several often-used probability distribution functions (PDFs) by goodness of fit test. As a result, the result of frequency analysis is affected by the power of the goodness of fit test. In the context of probability statistics, there is a wealth of the literatures on the goodness of fit test, such as Pombo et al. (2015), Durot and Reboul (2010), Goegebeur and Guillou (2010), Lee (2010), Saldana-Zepeda et al. (2010), Stephens (1974a), Stephens (1974b), Woodruff et al. (1984). Nevertheless, for hydrological frequency analysis, except for the precision and efficiency of the goodness of fit test, some issues should also be given special consideration. Generally, there are two key points should be included. (i) The influence of series length on the goodness of fit test. It is hard to ensure the length of observations in hydrological process, the insufficiency of the sample size makes it difficult to choose the best PDF; (ii) the PDF for fitting observations is unknown in the ordinary course of events, it should be selected from several alternative PDFs. Therefore, not only is the precision important for goodness of fit test, but also the accuracy that selecting the suitable PDF and rejecting others from several alternative PDF is crucial.

The assessment to the goodness of fit test is common in hydrological literatures. According to the flood studies report of the NERC (1975), the conventional Chi-Square (C-S) and Kolmogorov-Smirnov (K-S) goodness of fit tests were insensitive to departures from the assumed distribution for the samples size of 30 to 60. Ahmad

et al. (1988) assessed the K-S, Anderson-Darling (A-D), and Cramer-von Mises (C-vM) goodness of fit tests by testing gamma distribution from several alternative distributions, which were gamma, weibull, normal, log-normal, uniform, and beta distribution. The sample size for each distribution was varied from 5 to 30, each test was repeated for 5000 times. The power of goodness of fit test was defined as the proportion of rejection to the hypothesis that the samples used for test were not from gamma distribution, but from another distribution. The testing result showed that C-vM test was most powerful, followed by A-D test, but the differences between these tests were very slight. Nevertheless, the sample sizes of that study (5-30) was not widely representative in hydrological series, and the definition of power of test was not reasonable by only considering the proportion of rejections.

Onoz and Bayazit (1995) studied the flood data of 19 stations with record lengths in the range 60 to 165, several test methods were applied to search for the best probability distribution of flood series. They found that C-S test was highly sensitive to the location of the data near the class limits, small errors in the parameters of the distribution may have significant effects on the test results, and the data inside the class intervals was insignificant. Thus, C-S was not suitable for application in flood flows. In addition, A-D test was only limit to test GEV distribution, as the critical values were not available for other distributions. Therefore, this study was not a comprehensive assessment to the power of goodness of fit tests. Shahabuddin et al. (2009) investigated the power of several goodness of fit test such as K-S, A-D, C-vM, etc., and the critical values of each test were obtained by Monte Carlo simulation. By comparing with the power, the modified K-S test was found to be the best. In that study, however, the power of test was defined as the proportion of hypotheses that were rejected. In addition, the parameters of PDF were assumed known in hypothesis, and the samples used for testing were generated from same distribution, only with small difference in location or scale parameter. For hydrological series, however, the parameters of supposed distribution are unknown, and should be estimated by observations, and the observation data may be from the supposed PDF, or from other distributions. Therefore, the conclusions from Shahabuddin et al. (2009) may be not applicable for frequency analysis in hydrological series.

As the special requirement in hydrologic frequency analysis, the supposed probability distribution

function (PDF) is arbitrary chosen from several common used functions. Moreover, the parameters of PDF should be estimated, and the length of series cannot be guaranteed. Therefore, the assessment of goodness of fit test should consider not only the testing precision to the sample from known distribution function, but also emphasis on the ability to reject the hypotheses when the supposed PDF is different from the real one. In this paper, three often used goodness of fit test methods, C-S, K-S, and A-D tests were assessed. The samples used for test were generated from 4 common functions: normal, P3, Weibull, and uniform.

The paper is organized as follows: An introduction is first provided for the frequency analysis and goodness of fit test in hydrologic frequency analysis. It follows the description to the methods, including C-S, K-S, and A-D tests, the probability distribution functions and parameters estimation, and samples generation. Next, the statistic results of test, power comparison of the three test methods, and a discussion on the application of the goodness of fit test in hydrologic frequency analysis are presented. Finally, the main conclusions drawn from the analysis are presented.

## 2 Methods

The methods used in this paper include goodness of fit tests, C-S, K-S, A-D tests, parameter estimation for probability distribution function (PDF), and samples generation.

### 2.1 Goodness of fit test

Assuming  $(x_1, x_2, \dots, x_n)$  is the samples from population  $X$ , making a hypothesis:  $H_0: F(x) = F_0(x)$ , where  $F_0(x)$  is the alternative PDF (supposed PDF) with the parameters estimated by samples.

#### Chi-Square test

Chi-Square (C-S) test is a simple and convenient method for hypothesis test, it is related to the overall fit, the process can be written as follows (Zhang, Luo, 2000):

(1) Choosing  $k-1$  numbers:  $-\infty < t_1 < t_2 < \dots < t_{k-1} < +\infty$ ,  $k \approx 1.87(n-1)^{0.4}$ , and the number axis is partitioned into  $k$  intervals,  $(-\infty, t_1]$ ,  $(t_1, t_2]$ ,  $\dots$ ,  $(t_{k-2}, t_{k-1}]$ ,  $(t_{k-1}, +\infty]$ .

(2) Collecting the number of samples dropped into the  $i$ -th interval  $n_i$ ,  $i=1, 2, \dots, k$ , and then calculating the probability of the population which obeys alternative PDF fallen into the  $i$ -th interval:

$$\begin{cases} p_1 = P(X \leq t_1) = F_0(t_1) \\ p_2 = P(t_1 < X \leq t_2) = F_0(t_2) - F_0(t_1) \\ \dots \\ p_{k-1} = P(t_{k-2} < X \leq t_{k-1}) = F_0(t_{k-1}) - F_0(t_{k-2}) \\ p_k = P(t_{k-1} < X) = 1 - F_0(t_{k-1}) \end{cases} \quad (1)$$

(3) Constructing a statistics:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (2)$$

which obeys Chi-square distribution with the degree of freedom  $m$ ,  $m=k-1$ , or  $m=k-1-r$  when there are  $r$  independent parameters of  $F_0(x)$  need to be estimated by samples. And specifying a significance level  $\alpha$ , if  $p(\chi^2 \geq \chi_{1-\alpha}^2) \geq \alpha$ , then accept the hypothesis, otherwise reject it.

#### Kolmogorov-Smirnov test

Kolmogorov-Smirnov (K-S) test measures the greatest discrepancy between the observed and hypothesized distribution. The process can be summarized as follows (Melo et al., 2009; Wang, Wang, 2010):

(1) Sorting the samples  $X(x_1, x_2, \dots, x_n)$  by ascending order, and storing it to a new vector  $X'(x'_1, x'_2, \dots, x'_n)$ , calculating the empirical distribution function:

$$F_n(x') = \begin{cases} 0, & x' < x'_1 \\ \frac{k}{n}, & x'_k \leq x' < x'_{k+1} \\ 1, & x' \geq x'_n \end{cases} \quad (3)$$

(2) Calculating the K-S statistics  $D^{(n)}$ :

$$\begin{aligned} D^{(n)} &= \max_{x'_i \leq x' \leq x'_{i+1}} |F_n(x') - F_0(x')| \\ &= \max_{1 \leq i \leq n} \left\| \frac{i}{n} - F_0(x'_i) \right\|, \left\| \frac{i-1}{n} - F_0(x'_i) \right\| \end{aligned} \quad (4)$$

Specifying a significance level  $\alpha$ , if  $p(D^{(n)} \geq D^{(n)}(1-\alpha)) \geq \alpha$ , then accept the hypothesis, otherwise reject it.

#### Anderson-Darling test

Anderson-Darling (A-D) test emphasizes discrepancies in both tails of the distribution, and that is often of prime importance in hydrologic frequency analysis. The process can be written as follows (Coronel-Brizio, Hernandez-Montoya, 2010; Zhang et al., 2009):

(1) Sorting the samples  $X(x_1, x_2, \dots, x_n)$  by ascending order, and storing it to a new vector  $X'(x'_1, x'_2, \dots, x'_n)$ , calculating the empirical distribution function (Eq. 3).

(2) Calculating the A-D statistics  $A_n^2$ :

$$\begin{aligned} A_n^2 &= -n - \frac{1}{n} \sum_{i=1}^n [(2i-1) \log F(x'_i) \\ &\quad + (2n+1-2i) \log(1-F(x'_i))] \quad i=1, 2, \dots, n \end{aligned} \quad (5)$$

And specifying a significance level  $\alpha$ , if  $A_n^2 < A_n^2(1-\alpha)$ , then accept the hypothesis, otherwise reject it.

However, the critical values of the distributions in this study are unavailable directly at the whole range of series length, and so, they are obtained by simulations, for which the following simulation steps are taken:

- (1) Generating a sample of length  $n$  from the selected distribution, and estimating the parameters based on sample.
- (2) Calculating the A-D statistics  $A_n^2$ .
- (3) Based on Monte Carlo simulation, step (1) – step (2) are repeated 10000 times, and the A-D statistics are sorted by ascending order, the critical values at the significance level  $\alpha = 0.01, 0.05, 0.1$  are the values at 9900<sup>th</sup>, 9500<sup>th</sup>, 9000<sup>th</sup> of the statistics series.

## 2.2 Probability distribution function and parameters estimation

Four probability distribution functions are chosen for goodness of fit test in this study. Two of this have two parameters (normal, uniform), and other two have three parameters (P3, Weibull). These distributions have been selected because P3 and Weibull distribution are currently applied in the hydrologic frequency analysis in various countries frequently. Moreover, normal and uniform

distributions have a significant different characteristics of probability distribution compared with other two distributions (Fig. 1), in addition, they are often used in other field, especially in hydrogeology field (Hassan et al., 2009; Kuczera, Parent, 1998). Therefore, the power of the goodness of fit test can be assessed comprehensively.

Parameters of the P3 distribution are estimated by the principle of maximum entropy (Chen et al., 2002; Singh, 1987; Singh, Singh, 1985), and maximum likelihood method is used to estimate the parameters of Weibull distribution (Qiao and Tsokos, 1994; Qiao and Tsokos, 1995).

## 2.3 Samples generation

The sample is fundamental to the power test of goodness of fit test in this paper. The assessment of the test methods is influenced by the characteristics of samples. Therefore, the characteristics of generated samples should be keeping identical with their probability density function as much as possible. The samples of 4 probability density functions (normal, P3, Weibull, and uniform distribution) are generated based on the procedure of random number generation of IMSL (IMSL, 1997).

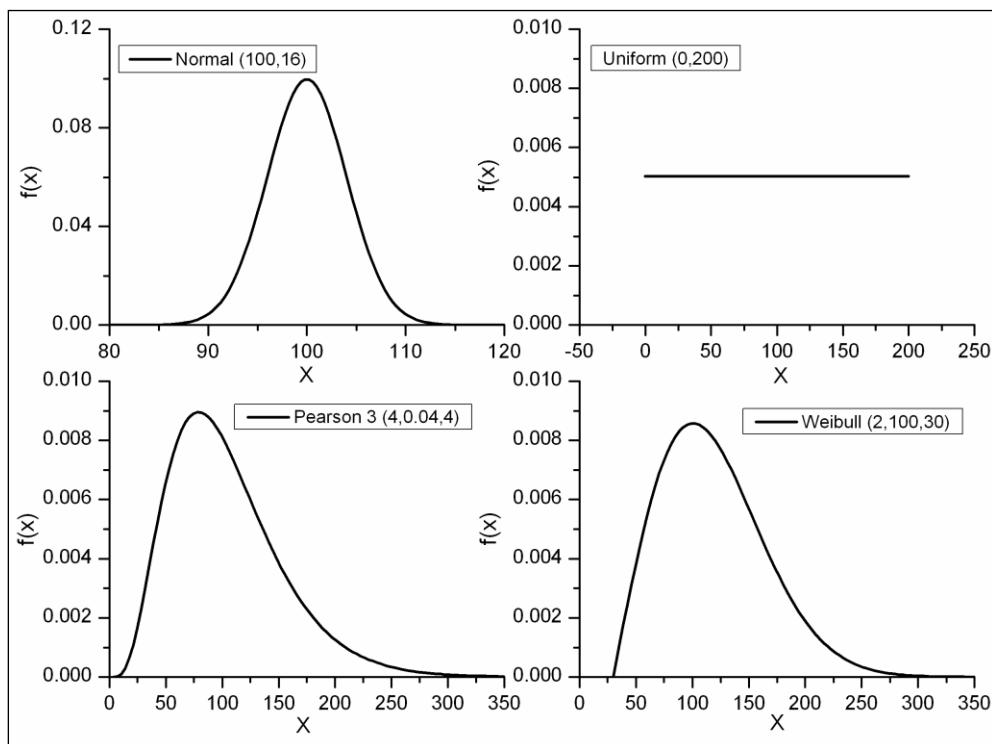


Fig. 1. The probability density curves of normal, uniform, P3, and Weibull distributions.

### 3 Results and discussion

The procedure of this power test is described as Fig. 2. The four samples with length  $n$  were generated from normal, P3, Weibull, and uniform distribution, and the supposed probability density function (PDF) in null hypothesis was set as one of four above functions respectively. After that, according to the hypothesis, estimating the parameters of PDF by sample, and testing all the hypotheses by C-S, K-S, and A-D tests in turn. The length of series  $n$  was varied from 20 to 300 with interval 10, each sample with length  $n$  was repeated for 2000 times based on Monte Carlo simulation, and the sample was generated randomly for each running.

The significance levels of Chi-Square (C-S), Kolmogorov-Smirnov (K-S), and Anderson-Darling (A-D) test are all set as 0.05, the acceptance proportion for each hypothesis is denoted as follows:

$$P_{M_S_H}(i) = M_{S_H}(i) / 2000 \quad (6)$$

Which  $M$  is the method used for goodness of fit test,  $S$  is the function used for generating samples,  $H$  is the supposed function in hypothesis,  $i = 1, 2, 3, \dots, 29$ , indicates the length of series varies from 20 to 300,  $M_{S_H}(i)$  is the accepting times for the specific hypothesis.

As showed by Fig. 3 – Fig. 5, for the three test methods, in most instances, the accepting proportion is increase with the length of series when the sample

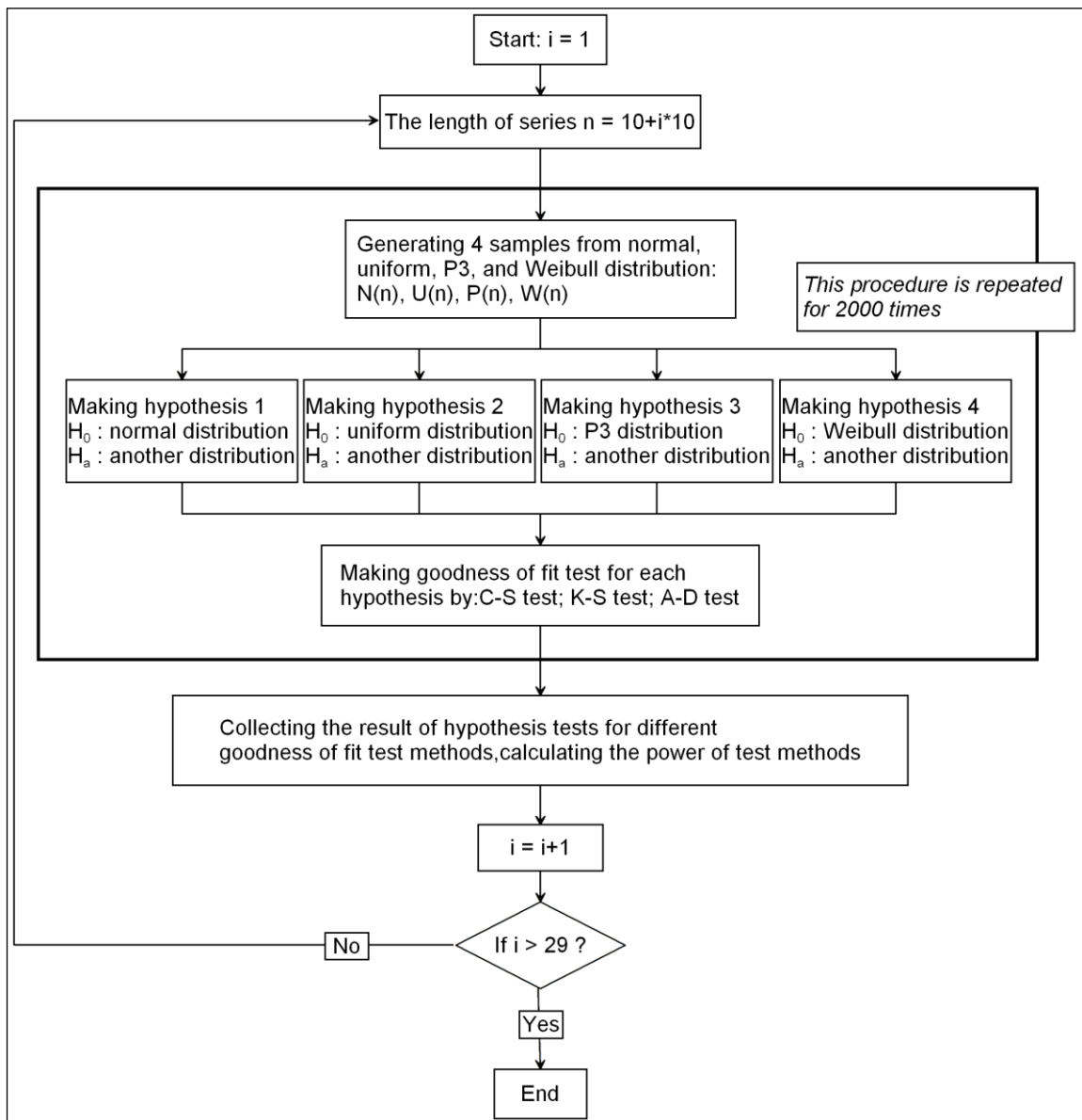


Fig. 2. The flow chart of this study.

is generated from supposed probability density function (PDF), and decrease when the sample is not generated from supposed function. In fact, it is attribute to the characteristics of sample which is closer to its PDF when the sample size is increased. However, when samples are generated from normal distribution, the hypotheses with Weibull as the assumed function have very high accepting proportion for all three goodness of fit tests. The accepting proportions of the hypotheses in which the actual and supposed PDF both are P3 distribution are low for C-S and K-S tests. K-S and A-D tests are easy to accept the hypotheses when the supposed function is Weibull distribution and the actual function is P3 distribution. Furthermore, the K-S test is ineffective for testing the normal function from the samples generated by Weibull distribution, and A-D test is very easy to accept the hypotheses when the supposed function is P3 and the samples are generated from Weibull or uniform distribution.

The ideal goodness of fit test should not only keep high testing precision for the samples generated from the supposed PDF, but also control the false accepting proportion when the samples are generated from other functions. Therefore, a formula is proposed for describing the power of goodness of fit test for a specific PDF:

$$Power\_M\_S(i) = P\_M\_S\_H(i) \quad (7)$$

$$= \frac{\sum_{S \neq H} (1 - P\_M\_S\_H(i))}{3}$$

The symbols in Eq. (7) are same as in Eq. (6).

Fig. 6 exhibits the powers of the three goodness of fit test methods for 4 specific PDFs. The power is increase with the length of series for each test method. Moreover, A-D test is faster to get a stable power value and the starting power is higher for each PDF (except uniform distribution) than C-S and K-S tests. For the 4 PDFs, the most powerful goodness of fit test for normal, uniform, P3, and Weibull distribution are K-S, C-S, A-D, and A-D test respectively. For the test methods, C-S test is good at testing uniform, normal, and Weibull distribution, but bad at P3 distribution. K-S test is powerful for testing normal and Weibull distribution, and weak at P3 distribution. Different from other two test methods, A-D test is good at P3, normal, and Weibull distribution, and ineffective to uniform distribution.

Considering the powers for different probability density functions, the comprehensive power for each goodness of fit test is described as follows:

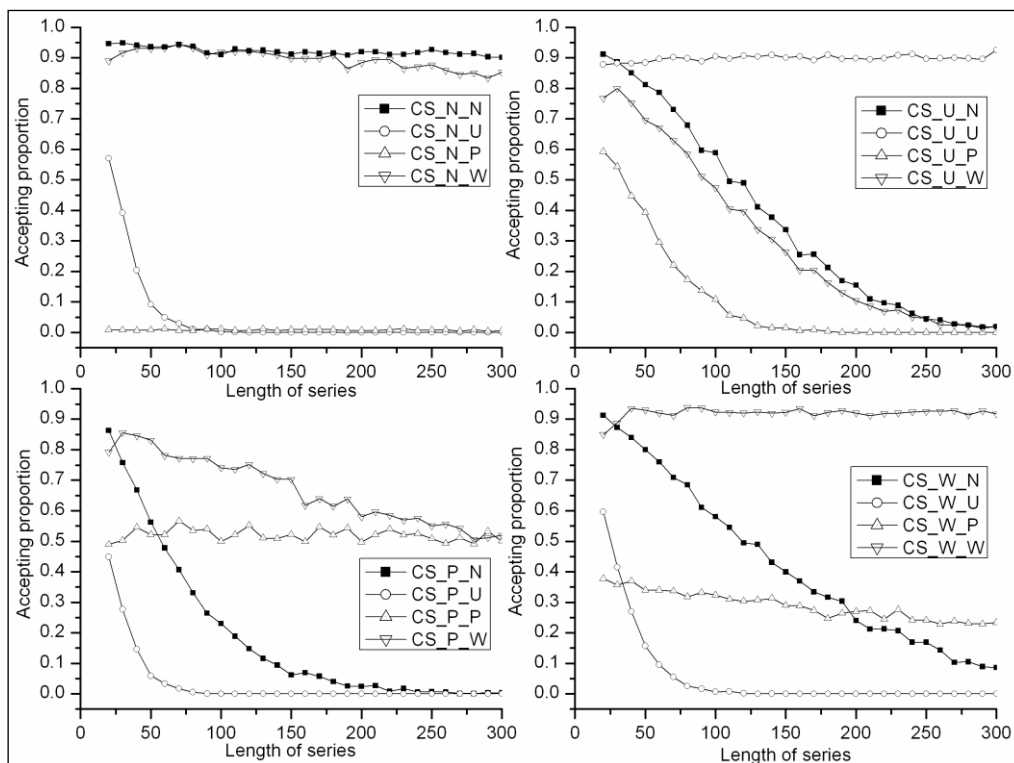


Fig. 3. The accepting proportion of hypotheses by Chi-Square test.

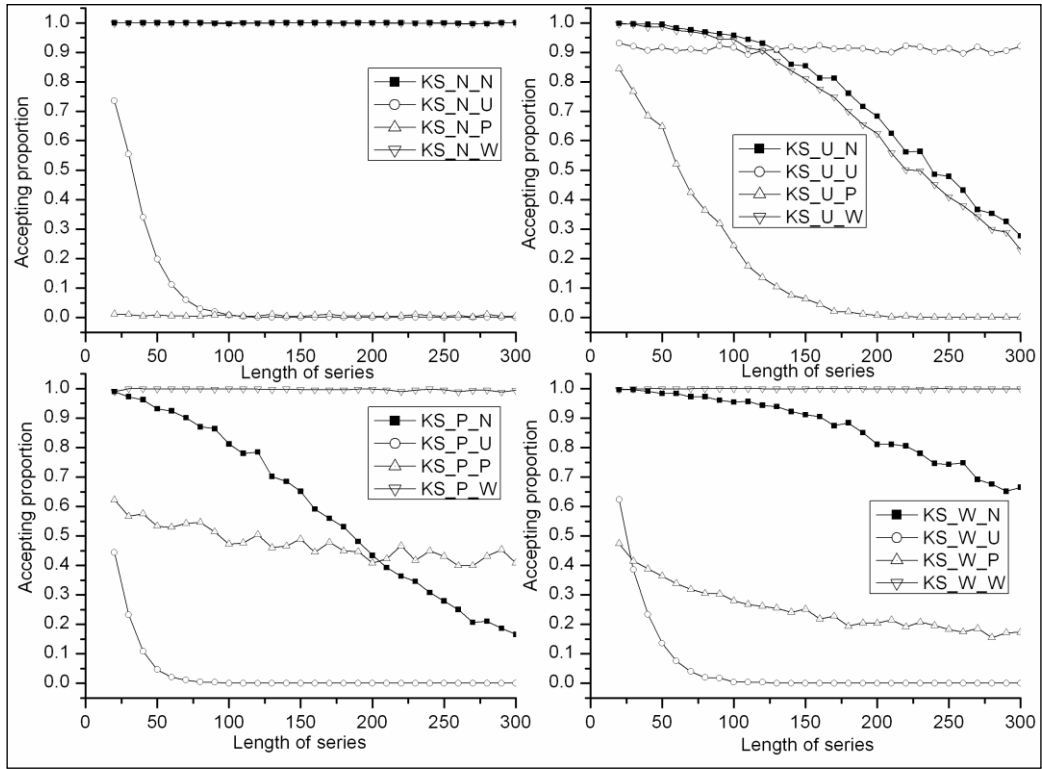


Fig. 4. The accepting proportion of hypotheses by Kolmogorov-Smirnov test.

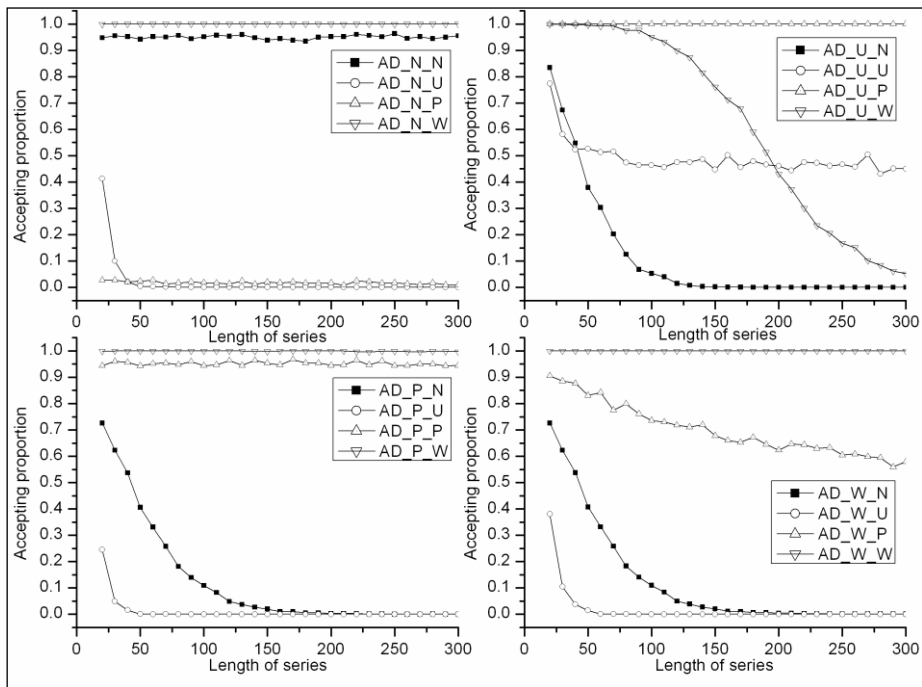


Fig. 5. The accepting proportion of hypotheses by Anderson-Darling test.



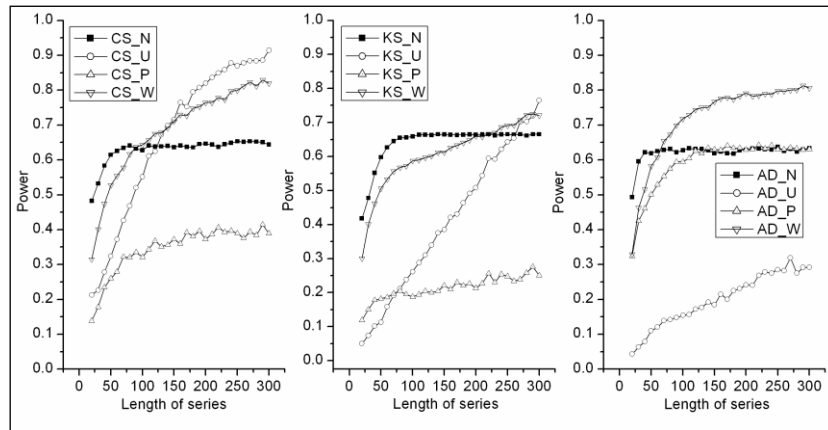


Fig. 6. The powers of Chi-square, Kolmogorov-Smirnov, and Anderson-darling tests for specific probability density function.

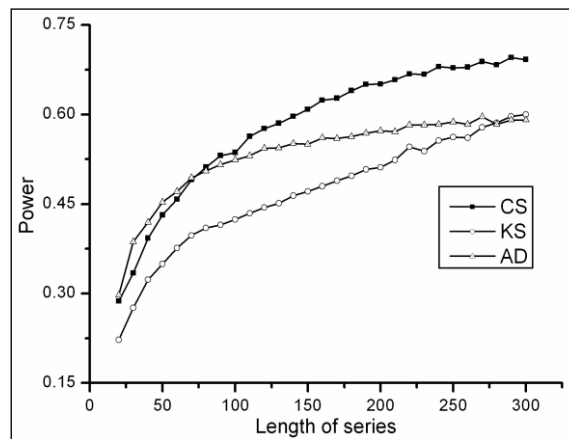


Fig. 7. The comprehensive powers of Chi-square, Kolmogorov-Smirnov, and Anderson-darling tests.

$$Power\_M = \frac{\sum Power\_M\_S}{4} \quad (8)$$

The symbols in Eq. (7) are same as in Eq. (6).

Fig. 7 shows that the comprehensive powers for the tests are all increase with the series length. The goodness of fit test with the best performance is C-S test, the second is K-S test, and the last is A-D test, nevertheless, the difference among the powers of 3 test methods is not so significant.

From the testing results above, it is concluded that the three test methods used in this study measure the different aspects of the goodness of fit. The C-S test is related to the overall fitting. Therefore, it is appropriate to select C-S as the goodness of fit test when nothing is known about the samples. In addition, C-S test is powerful for distinguishing uniform distribution from other distributions and distinguishing other distributions when the sample is generated from uniform distribution. The power of C-S test is not so sensitive to the length of series (Fig. 3). The K-S test measures the greatest discrepancy between the sample and supposed probability density

function (PDF). It has the highest precision for testing the samples which generated from supposed function, and there is no other advantage to others except this. Moreover, the K-S test is most insensitive to the length of series (Fig. 4). The A-D test is emphasis on the differences at the both tails. Thus, the power for specific PDF is most sensitive to the length of series, and it reach to a stable value quickly (Fig. 5). In addition, the A-D test is extremely inefficient for testing PDFs when the samples are generated from uniform distribution.

According to Fig. 3 – Fig. 5, it is obvious to found that the power for testing normal distribution from Weibull samples, and testing Weibull distribution from normal sample are fairly low for all the 3 test methods. Referring to the probability density functions of normal and Weibull distribution:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (9)$$

$$f(x; \alpha, \beta, \varepsilon) = \frac{\alpha}{\beta^\alpha} (x-\varepsilon)^{\alpha-1} \exp\left[-\frac{(x-\varepsilon)^\alpha}{\beta}\right], 0 < x < \infty \quad (10)$$



In fact, the above 2 distributions can be both transformed from exponential gamma distribution:

$$f(x; a, b, c, \delta) = \frac{b^a}{\delta \Gamma(a)} (x-c)^{\frac{a}{\delta}-1} \exp(-b(x-c)^{\frac{1}{\delta}}) \quad (11)$$

when the following conditions are satisfied respectively:

$$a = \frac{1}{2}, b = \frac{1}{2}, c = 0, \delta = \frac{1}{2} \quad (12)$$

$$a = 1, b = \beta^{-\alpha}, c = \varepsilon, \delta = \frac{1}{\alpha} \quad (13)$$

Therefore, the normal and Weibull distribution can be mutual transformed in some special condition, and it is hard to test them from each other with high precision.

To sum up the above arguments, each goodness of fit test has its own advantage and disadvantage. As a result of the power test in this paper adopted the different evaluation criterion from other studies, which considering the ability to distinguish the sample from both real and false distributions. Therefore, the A-D test is not always the best test method as reported in previous literatures. In order to acquire best testing effect in hydrological frequency analysis, it is beneficial to collect the prior information about the research variable as much as possible, and then select corresponding method. For example, from previous literatures, the observations always represent normal, Weibull, and P3 distribution, and then A-D test should be selected for testing, because the A-D test has greatest comprehensive power when uniform distribution is excluded. In addition, the A-D test should be adopted when the length of series is not so large. When nothing is known about the probability distribution characteristics of samples, the C-S test is preferred. In addition, parameters estimation of probability distribution functions is always included in the procedure of goodness of fit test. Thus, the methods used for estimating parameters may have influence on the results of hypothesis tests. Moreover, how to effectively distinguish normal distribution (or Weibull distribution) from Weibull distribution (or normal distribution) is the focus of future work.

#### 4 Conclusions

By choosing Chi-Square (C-S), Kolmogorov-Smirnov (K-S), and Anderson-Darling (A-D) test as the methods of goodness of fit test, the supposed PDF in hypothesis and the function used for generating sample are selected as normal, uniform, P3, and Weibull distributions. Based on the comparison of

goodness of fit test methods, following conclusions are demonstrated:

(1) Each goodness of fit test has its own advantage and disadvantage. The most powerful goodness of fit test method for normal, uniform, P3, and Weibull distribution are K-S, C-S, A-D, and A-D test respectively. C-S test is good at testing uniform, normal, and Weibull distribution, but bad at P3 distribution. K-S test is powerful for testing normal and Weibull distribution, and weak at P3 distribution. A-D test is good at P3, normal, and Weibull distribution, and ineffective to uniform distribution.

(2) The goodness of fit test with the best comprehensive power is C-S test, the second is K-S test, and the last is A-D test. It is beneficial to search for the prior information about the research variables as much as possible, and then select corresponding method.

#### Acknowledgements

This study was supported by the National Basic Research Program (2013CB956503), National Natural Science Fund of China (No. 41302181, 41172207, 51190091 and 51409161), and Program for New Century Excellent Talents in University (NCET-12-0262), Science and Technology program of Jiangsu Province (No. BK20140080).

#### References

- Ahmad, M.I., Sinclair, C.D., Spurr, B.D., 1988. Assessment of Flood Frequency Models Using Empirical Distribution Function Statistics. *Water Resour Res*, 24(8), 1323-1328.
- Beard, L., 1974) Flood flow frequency techniques. Center of Res. in Water Resour., University of Texas, Austin.
- Bi, T.A.G., Soro, G.E., Dao, A., Kouassi, F.W., Srohourou, B., 2010. Frequency Analysis and New Cartography of Extremes Daily Rainfall Events in Cote d'Ivoire. *Journal of Applied Sciences*, 1684-1694.
- Chen, Y.F., Hou, Y., Van Gelder, P., Zhigui, S., 2002. Study of parameter estimation methods for Pearson-III distribution in flood frequency analysis. *Iahs-Aish P*, 271, 263-269.
- Coronel-Brizio, H.F., Hernandez-Montoya, A.R., 2010. The Anderson-Darling test of fit for the power-law distribution from left-censored samples. *Physica a-Statistical Mechanics and Its Applications*, 389(17), 3508-3515.
- National Environment Research Council, 1975. Flood studies report, Vols. 1-5. NERC, London.
- US Water Resources Council, 1976. Guidelines for determining flood flow frequency, Bull. 17, Hydrol. Comm., Washington, D.C.
- Durot, C., Reboul, L., 2010. Goodness-of-Fit Test for Monotone Functions. *Scand J Stat*, 37(3), 422-441.
- Goegebeur, Y., Guillou, A., 2010. Goodness-of-fit testing for Weibull-type behavior. *J Stat Plan Infer*, 140(6),

- 1417-1436.
- Gunasekara, T.A.G., Cunnane, C., 1991. Expected Probabilities of Exceedance for Nonnormal Flood Distributions. *J Hydrol*, 128(1-4), 101-113.
- Hassan, A.E., Bekhit, H.M., Chapman, J.B., 2009. Using Markov Chain Monte Carlo to quantify parameter uncertainty and its effect on predictions of a groundwater flow model. *Environ Modell Softw*, 24(6), 749-763.
- IMSL, 1997. FORTRAN Subroutines for Mathematical Applications, IMSL inc., Houston, Texas, America. Vol 2, 1165-1217.
- Khalik, M.N., Ouarda, T.B.M.J., Ondo, J.C., Gachon, P., Bobee, B., 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations, A review. *J Hydrol*, 329(3-4): 534-552.
- Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models, the Metropolis algorithm. *J Hydrol*, 211(1-4), 69-85.
- Lansigan, F.P., 2009. Frequency analysis of extreme hydrologic events and assessment of water stress in a changing climate in the Philippines. From Headwaters to the Ocean: Hydrological Changes and Watershed Management, 497-501.
- Lee, J.J., Kwon, H.H., Kim, T. W., 2012. Spatio-temporal analysis of extreme precipitation regimes across South Korea and its application to regionalization. *Journal of Hydro-Environment Research* 6(2), 101-110.
- Lee, S., 2010. On the Goodness of Fit Test for Discretely Observed Sample from Diffusion Processes, Divergence Measure Approach. *J Korean Math Soc*, 47(6), 1137-1146.
- Melo, I., Tomasik, B., Torrieri, G., Vogel, S., Bleicher, M., Korony, S., Gintner, M., 2009. Kolmogorov-Smirnov test and its use for the identification of fireball fragmentation. *Phys Rev C*, 80(2), Artn 024904.
- Onoz, B., Bayazit, M., 1995. Best-Fit Distributions of Largest Available Flood Samples. *J Hydrol*, 167(1-4), 195-208.
- Onoz, B., Bayazit, M., 2001. Power distribution for low streamflows. *J Hydrol Eng*, 6(5), 429-435.
- Pombo, S., de Oliveira R. P., 2015. Evaluation of extreme precipitation estimates from TRMM in Angola. *J Hydrol*, 523: 663 - 679.
- Qiao, H.Z., Tsokos, C.P., 1994. Parameter-Estimation of the Weibull Probability-Distribution. *Mathematics and Computers in Simulation*, 37(1), 47-55.
- Qiao, H.Z., Tsokos, C.P., 1995. Estimation of the 3-Parameter Weibull Probability-Distribution. *Mathematics and Computers in Simulation*, 39(1-2), 173-185.
- Raju, K.S., and Kumar, D. N., 2011. Classification of microwatersheds based on morphological characteristics. *Journal of Hydro-Environment Research* 5(2), 101-109.
- Saf, B., 2009. Regional Flood Frequency Analysis Using L Moments for the Buyuk and Kucuk Menderes River Basins of Turkey. *J Hydrol Eng*, 14(8), 783-794.
- Saldana-Zepeda, D.P., Vaquera-Huerta, H., Arnold, B.C., 2010. A goodness of fit test for the Pareto distribution in the presence of Type II censoring, based on the cumulative hazard function. *Computational Statistics, Data Analysis*, 54(4), 833-842.
- Sandoval, C.E., Raynal-Villasenor, J., 2008. Trivariate generalized extreme value distribution in flood frequency analysis. *Hydrolog Sci J*, 53(3), 550-567.
- Shabri, A., Ariff, N.M., 2009. Frequency Analysis of Maximum Daily Rainfalls via L-Moment Approach. *Sains Malaysiana*, 38(2), 149-158.
- Shahabuddin, F.A.A., Ibrahim, K., Jemain, A.A., 2009. On The Comparison of Several Goodness of Fit Tests, With Application to Wind Speed Data. *Energy Problems and Environmental Engineering*, 394-398.
- Shao, Q.X., Wong, H., Xia, J., Ip, W.C., 2004. Models for extremes using the extended three-parameter Burr XII system with application to flood frequency analysis. *Hydrolog Sci J*, 49(4), 685-702.
- Singh, V.P., 1987. Derivation of the Pearson Type (PT) III distribution by using the principle of maximum-entropy (POME)-reply. *J Hydrol*, 90(3-4), 355-357.
- Singh, V.P., Singh, K., 1985. Derivation of the Pearson Type (PT) III distribution by using the principle of maximum-entropy (POME). *J Hydrol*, 80(3-4), 197-214.
- Smakhtin, V.U., 2001. Low flow hydrology: a review. *J Hydrol*, 240(3-4), 147-186.
- Song, S.B., Singh, V.P., 2010. Frequency analysis of droughts using the Plackett copula and parameter estimation by genetic algorithm. *Stoch Env Res Risk A*, 24(5), 783-805.
- Stephens, M.A., 1974a. Components of Goodness-of-Fit Statistics. *Ann IH Poincare B*, 10(1), 37-54.
- Stephens, M.A., 1974b. Edf Statistics for Goodness of Fit and Some Comparisons. *J Am Stat Assoc*, 69(347), 730-737.
- Vogel, R.M., McMahon, T.A., Chiew, F.H.S., 1993. Floodflow Frequency Model Selection in Australia. *J Hydrol*, 146(1-4), 421-449.
- Wang, F.G., Wang, X.D., 2010. Fast and Robust Modulation Classification via Kolmogorov-Smirnov Test. *Ieee T Commun*, 58(8), 2324-2332.
- Woodruff, B.W., Viviano, P.J., Moore, A.H., Dunne, E.J., 1984. Modified Goodness-of-Fit Tests for Gamma-Distributions with Unknown Location and Scale-Parameters. *Ieee T Reliab*, 33(3), 241-245.
- Xu, Y.P., Booiij, M.J., Tong, Y.B., 2010. Uncertainty analysis in statistical modeling of extreme hydrological events. *Stoch Env Res Risk A*, 24(5), 567-578.
- Yang, T., Xu, C.Y., Shao, Q.X., Chen, X., 2010. Regional flood frequency and spatial patterns analysis in the Pearl River Delta region using L-moments approach. *Stoch Env Res Risk A*, 24(2), 165-182.
- Yoo, C., Jung, K.S., Kim, T.W., 2005. Rainfall frequency analysis using a mixed Gamma distribution, evaluation of the global warming effect on daily rainfall. *Hydrol Process*, 19(19), 3851-3861.
- Zhang, D.P., Luo, Y.L., 2000. Applied Probability and Statistic. Higher Education Press, BeiJing. in Chinese
- Zhang, W., Yu, S., Zhang, G., 2009. Anderson-Darling test for clutter distribution identification. *Chinese Journal of Scientific Instrument*, 30(3), 631-635. in Chinese with English abstract.