

Research Article

A Robust Skewed Boxplot for Detecting Outliers in Rainfall Observations in Real-Time Flood Forecasting

Chao Zhao ¹ and Jinyan Yang ²

¹School of Environmental Science and Engineering, Xiamen University of Technology, 361024 Xiamen, China

²Suzhou Branch of Hydrology and Water Resources Investigation Bureau of Jiangsu Province, 215000 Suzhou, China

Correspondence should be addressed to Chao Zhao; 157096652@qq.com

Received 17 September 2018; Accepted 30 January 2019; Published 28 February 2019

Academic Editor: Harry D. Kambezidis

Copyright © 2019 Chao Zhao and Jinyan Yang. 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 standard boxplot is one of the most popular nonparametric tools for detecting outliers in univariate datasets. For Gaussian or symmetric distributions, the chance of data occurring outside of the standard boxplot fence is only 0.7%. However, for skewed data, such as telemetric rain observations in a real-time flood forecasting system, the probability is significantly higher. To overcome this problem, a medcouple (MC) that is robust to resisting outliers and sensitive to detecting skewness was introduced to construct a new robust skewed boxplot fence. Three types of boxplot fences related to MC were analyzed and compared, and the exponential function boxplot fence was selected. Operating on uncontaminated as well as simulated contaminated data, the results showed that the proposed method could produce a lower swamping rate and higher accuracy than the standard boxplot and semi-interquartile range boxplot. The outcomes of this study demonstrated that it is reasonable to use the new robust skewed boxplot method to detect outliers in skewed rain distributions.

1. Introduction

More real-time flood forecasting systems in China, particularly in the large basins where there are many remote gauge stations, now use telemetry systems to transmit the rainfall signals of rainfall stations because telemetry systems can provide timely, dense, and labor-saving hydrological information for remote rainfall stations [1]. However, it has been shown that telemetric rainfall information includes inevitable outliers caused by instrument malfunction, human-related errors, and/or signal acquisition errors resulting from signal leaks, and collisions or disturbances in the process of signal transmission, in addition to random errors normally distributed with zero mean and a small variance [1–3]. The outliers have an unknown distribution with a much greater variance and appear to be inconsistent with the remainder of the dataset and are relatively large in magnitude [4, 5]. Therefore, outliers should be treated differently [6], and in this paper, observations containing outliers were called abnormal data.

In real-time flood forecasting systems, rainfall observations represent the main input and determine the accuracy of the forecasting results. The presence of abnormal data can lead to unreliable forecast conclusions. The need to increase the accuracy and reliability of telemetric rainfall data has prompted researches on the construction of a robust method to efficiently detect these abnormal data before they have been entered into the hydrologic models [7].

One of the most frequently used nonparametric tools to detect outliers for a univariate dataset is based on the concept of the boxplot. The method, suggested by Tukey [8], has come into common use and has been studied extensively (see for example [9–16]). An observation is considered as “potential” abnormal data when its value does not belong to the interval (the fence): $(q_1 - 1.5 * IQR, q_3 + 1.5 * IQR)$, where q_1 and q_3 are the first and third quartiles, respectively, and IQR is the interquartile range, i.e., $IQR = q_3 - q_1$. The standard boxplot is fitted to normal or symmetric distributions in particular. For Gaussian data, the probability of

lying below $(q_1 - 1.5 * IQR)$ or above $(q_3 + 1.5 * IQR)$ is 0.0035 (0.35%) each.

However, the real hourly rainfall observations are often not normal or symmetric. When we apply the standard boxplot for real hourly rainfall observations, the percentage of data outside the standard boxplot fence becomes excessively high. As an example, we selected the real hourly rainfall data of flood events from 1988 to 2008 measured at the Wuyigong rain gauge in the Qilijie basin in Southeastern China. The standard boxplot of the datasets is shown in Figure 1. The figures for other rainfall stations were not included because they were nearly identical to that for the Wuyigong Station.

It is clear that the underlying distribution of the rainfall dataset was skewed to the right. Up to 13.04% of the observations were above $q_3 + 1.5 * IQR$. Clearly, it would not be correct to classify them all as real abnormal data.

To cope with this, several adjusted boxplot methods have been proposed in case of skewed data. Kimber [17] suggested the use of the semi-interquartile range (SIQR) rather than IQR, i.e., the fence of the SIQR boxplot is defined as $q_1 - 3 * (q_2 - q_1)$, $q_3 + 3 * (q_3 - q_2)$. The SIQR boxplot has also been applied to the real hourly rain observations from the Wuyigong rain gauge. The boxplots of the two methods are shown in Figure 2.

The SIQR boxplot adjusts itself to the right skewness, compared with the standard method (Figure 2). The SIQR method expands the upper boundary slightly, and consequently less data are detected as outliers. However, the adjustment is not enough [18]. The probability of lying outside of the SIQR fence was 9.29%. It was still highly risky to identify these values as real abnormal data. The poor performance of the SIQR was further analyzed in Section 3.

Carling [19], Schwertman et al. [14], and Schwertman and de Silva [20] suggested replacing q_1 and q_3 in the fences with the median q_2 . They also suggested replacing the constant 1.5 with the functions that combine sample size with skewness. These approaches can achieve a prespecified swamping, i.e., the potential of misclassifying an uncontaminated observation as real abnormal data [21]. The methods perform well when the distribution is a lambda distribution. It is not clear how they perform with other distributions. Finally, the functions in these methods depend on the sample size, and the procedures require some characteristics of the uncontaminated distribution, which is often difficult to estimate for the real-time hourly rainfall datasets.

The aim of the paper was to construct a new boxplot method that was robust to outliers and sensitive to skewness. The new method was independent of the sample size and performs well with the rainfall distribution. It can reduce swamping and rapidly detected abnormal telemetric rainfall data before they were entered into the real-time flood forecasting model. This paper was organized as follows. In Section 2, we detailed the proposed procedure that included a robust measure of skewness in the construction of the fences. Section 3 illustrated the differences between the standard, SIQR, and robust boxplots for uncontaminated as

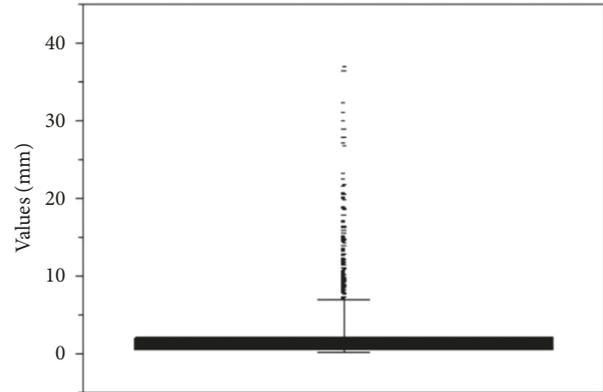


FIGURE 1: The standard boxplot of hourly rainfall datasets from Wuyigong Station.

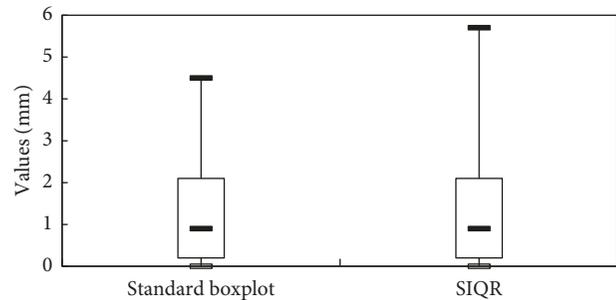


FIGURE 2: The fences of the two boxplot methods for Wuyigong Station.

well as abnormal datasets. Finally, we provided conclusions in Section 4.

2. Materials and Methods

2.1. Study Catchments. Three catchments located in southeastern and southern China were selected. They were the Qilijie catchment in Fujian Province; the Lushui reservoir catchment in Hubei Province; and the Yitang reservoir catchment in Guangdong Province. The three catchments showed some similar hydrological characteristics, such as excessive precipitation and a humid climate.

The areas of the three selected catchments varied, collecting together the basin characteristics in different sizes. The Qilijie catchment, as a representative of a large basin, covered 14,787 km² with 43 telemetric rain gauges; the Lushui catchment was representative of a middle basin covering 3,960 km² with 13 telemetric rain gauges; and the Yitang reservoir catchment represented a small watershed, covering 251 km² with 6 telemetric rain gauges.

2.2. Data. Historic hourly rainfall observations from 1988 to 1998 from all the rain stations were compiled and used. The datasets were considered as nonoutliers or normal data. The total number of all the rainfall records was greater than 200,000.

2.3. Methods

2.3.1. Robust Skewness. The asymmetry of a distribution can be described by the skewness coefficient. The classical skewness coefficient depends on the second and third empirical moments of the datasets. However, the moments are sensitive to outliers. Therefore, the classical skewness coefficient could be strongly affected by outliers. Even a single outlier can make it easy to distort and make it difficult to interpret [22]. To overcome this problem, the medcouple (MC), introduced in Brys et al. [22], was chosen to estimate the skewness of rainfall observations.

The datasets were sorted in ascending order, i.e., $p_1 \leq p_2 \leq \dots \leq p_s$, where s is the number of the datasets.

The MC of the datasets is defined as

$$MC = \text{med}_{p_i \leq m \leq p_j} h(p_i, p_j), \quad (1)$$

where m is the median of the observation samples, and for all $p_i \neq p_j$, the kernel function h is given by

$$h(p_i, p_j) = \frac{(p_j - m) - (m - p_i)}{p_j - p_i}. \quad (2)$$

For the special case $p_i = p_j = m$, the kernel function can be estimated as follows. Let $n_1 < n_2 < \dots < n_k$ denote the indices of the observations that are tied to the median m , i.e., $p_{n_l} = m$ for all $l = 1, 2, \dots, k$. Then

$$h(p_{n_i}, p_{n_j}) = \begin{cases} -1, & \text{if } i + j - 1 < k, \\ 0, & \text{if } i + j - 1 = k, \\ +1, & \text{if } i + j - 1 > k. \end{cases} \quad (3)$$

The MC equals the median of all $h(p_i, p_j)$ values for which $p_i \leq m \leq p_j$. According to the definitions in equations (1) and (2), MC is based on the quantiles and is therefore not as vulnerable to outliers as the classical skewness. It is clear that MC always lies between -1 and 1 . A distribution that is skewed to the right has a positive value of MC, whereas it becomes negative for a left skewed distribution. Finally, a symmetric distribution has a zero MC. As shown in Brys et al. [22], MC is robust to resisting outliers and sensitive to detecting skewness. It has a bounded influence function and a breakdown value of 25%, which means that MC can resist up to 25% of outliers in the data.

2.3.2. Combination the Robust Skewness with the Boxplot Fence. To construct the boxplot fences for the skewed data, we propose to insert medcouple (MC) into the boxplot method. The constant 1.5 in the standard boxplot fence is replaced by some functions related to MC, such as $f_l(\text{MC})$ and $f_u(\text{MC})$. The new robust skewed fence is defined by

$$[q_1 - f_l(\text{MC}) \cdot \text{IQR}, q_3 + f_u(\text{MC}) \cdot \text{IQR}]. \quad (4)$$

Let $f_l(0) = f_u(0) = 1.5$ to equal the standard boxplot at symmetric distributions. When distributions are asymmetric, $f_l(\text{MC})$ and $f_u(\text{MC})$ can be used to adjust the fence to fit the skewness. Using different functions, $f_l(\text{MC})$ and

$f_u(\text{MC})$ cause the effects of the adjustment to be different. In Section 2.3.3, we compare the differences between the different functions.

2.3.3. Comparison of Different Functions. Three types of simple functions including only a few parameters were selected, which are important for operational real-time flood forecasting systems.

The following three functions were considered:

(1) Linear function:

$$\begin{aligned} f_l(\text{MC}) &= 1.5 + a\text{MC}, \\ f_u(\text{MC}) &= 1.5 + b\text{MC}. \end{aligned} \quad (5)$$

(2) Quadratic function:

$$\begin{aligned} f_l(\text{MC}) &= 1.5 + c_1\text{MC} + c_2\text{MC}^2, \\ f_u(\text{MC}) &= 1.5 + d_1\text{MC} + d_2\text{MC}^2. \end{aligned} \quad (6)$$

(3) Exponential function:

$$\begin{aligned} f_l(\text{MC}) &= 1.5e^{g\text{MC}}, \\ f_u(\text{MC}) &= 1.5e^{f\text{MC}}, \end{aligned} \quad (7)$$

with $a, b, c_1, c_2, d_1, d_2, g, f \in \mathbb{R}$.

To determine the values of $a, b, c_1, c_2, d_1, d_2, g, f$, the expected percentage of observations beyond the robust skewed fence (equation (4)) was set to 0.7%, which was similar to the rule of the standard boxplot of the Gaussian distribution. According to the rule, the fence boundaries must reach

$$\begin{aligned} q_1 - f_l(\text{MC}) \cdot \text{IQR} &\approx q_\alpha, \\ q_3 + f_u(\text{MC}) \cdot \text{IQR} &\approx q_\beta, \end{aligned} \quad (8)$$

where q_α and q_β are the α th and β th quantile of the distribution, respectively, and $\alpha = 0.0035$ and $\beta = 0.9965$.

Combining equations (5)–(7) into equation (8), we can obtain

$$\begin{cases} \frac{(q_1 - q_\alpha)}{\text{IQR}} - 1.5 \approx a\text{MC}, \\ \frac{(q_\beta - q_3)}{\text{IQR}} - 1.5 \approx b\text{MC}, \end{cases} \quad (9)$$

$$\begin{cases} \frac{(q_1 - q_\alpha)}{\text{IQR}} - 1.5 \approx c_1\text{MC} + c_2\text{MC}^2, \\ \frac{(q_\beta - q_3)}{\text{IQR}} - 1.5 \approx d_1\text{MC} + d_2\text{MC}^2, \end{cases} \quad (10)$$

$$\begin{cases} \ln\left(\frac{2}{3} \frac{q_1 - q_\alpha}{IQR}\right) \approx gMC, \\ \ln\left(\frac{2}{3} \frac{q_\beta - q_3}{IQR}\right) \approx fMC. \end{cases} \quad (11)$$

Based on the historic hourly rainfall records, equations (9)–(11) were constructed for each rain gauge. The parameter values of the three functions could be derived using linear least squares estimation.

3. Results and Discussion

3.1. Data Preprocessing. In humid catchments, the gap between the maximum and minimum of the hourly rainfall records was large. Analyzing the datasets together will make it difficult to correctly detect outliers. To overcome this issue, the data records must be preprocessed. First, the hourly areal mean rainfall (HAMP) was calculated using the Tyson polygon. Then, the hourly areal mean rainfall (HAMP) was divided into three groups: (0, 1 mm] (Group 1), (1 mm, 2 mm] (Group 2), and (2 mm, +∞) (Group 3). Finally, the rainfall records for each station are divided into the three groups based on the HAMP. The analysis was repeated for each group.

3.2. MC Results. The average results for MC for the three groups are listed in Table 1.

It was clear that the average MCs are greater than zero. The rainfall datasets of the three groups were right skewed. The average MCs were less than 0.5, which demonstrated that the distributions were not extremely skewed. It was risky to use the standard and SIQR boxplot to detect outliers for skewed rainfall distributions.

3.3. Results of the Function Comparisons. To compare the behavior of the three different functions, they were applied to fitting the lower and upper boundaries of the fence. The fit of the linear, quadratic, and exponential functions is displayed in Figures 3 and 4. $\ln((2/3)((q_1 - q_\alpha)/IQR))$ is set as the vertical axis in Figure 3, and $\ln((2/3)((q_\beta - q_3)/IQR))$ is set as the vertical axis in Figure 4. The special defined vertical axis is the same as in equation (11) that allows exponential functions to form a straight line.

When MC was greater than 0.43, the linear function decreases abruptly and fully separated from the observation samples (Figure 3). At the same time, the quadratic and exponential functions fit the samples better. The fit of the exponential function was slightly better than that of the quadratic function (Figure 4). The exponential function was simply needed to determine fewer parameters than the quadratic function. As a result, the exponential function was selected to conduct the robust boxplot fence. The new robust skewed boxplot fence is defined by

$$[q_1 - 1.5e^{gMC} \cdot IQR, q_3 + 1.5e^{fMC} \cdot IQR]. \quad (12)$$

TABLE 1: Average MC values.

Groups	Average MC
Group 1	0.19
Group 2	0.47
Group 3	0.39

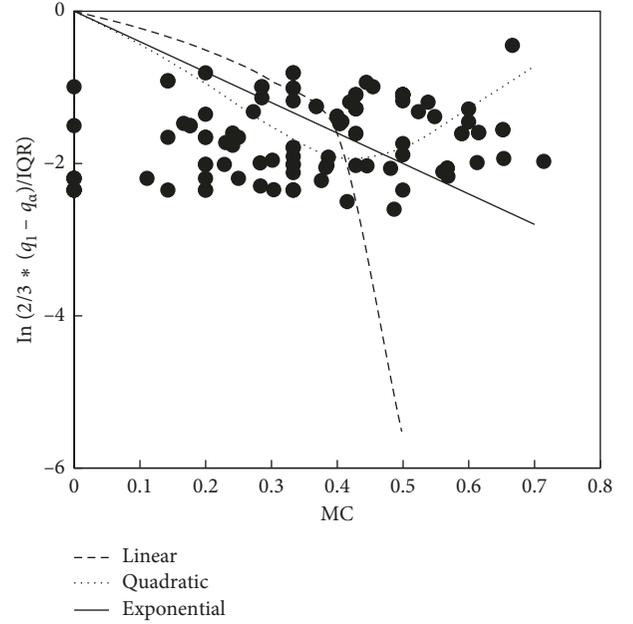


FIGURE 3: The fit of the lower boundary.

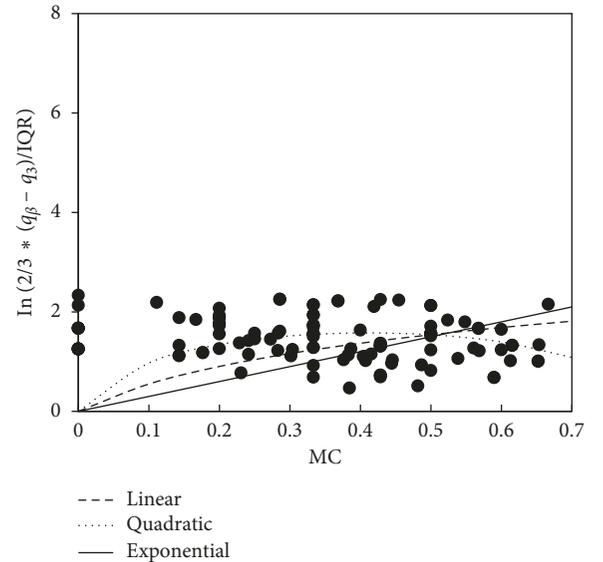


FIGURE 4: The fit of the upper boundary.

The parameters in the new fence are estimated using the rainfall records from the three basins mentioned above. To simplify the practical application of the new method, the estimated values are taken by rounding up $g = -3.96$ and $f = 3.35$ to $g = -4$ and $f = 3$. Note that, rounding up the values to the nearest *smaller* integer yields a smaller fence and

TABLE 2: Performance of the three boxplots without outliers.

Method	SR	Accuracy
Standard boxplot	6.10	93.90
SIQR	5.71	94.29
Robust skewed boxplot	1.32	98.68

consequently a more robust model. The new robust skewed fence is

$$[q_1 - 1.5e^{-4MC} \cdot IQR, q_3 + 1.5e^{3MC} \cdot IQR]. \quad (13)$$

Note that, although MC is a robust estimator, it can be affected by outliers, particularly at high percentages of outliers. To reduce the effects of a high percentage of outliers on MC and new boxplot fences, a low percentage of outliers ($\leq 5\%$) is considered. This low percentage of outliers coincides with the characteristics of the telemetric rainfall observations.

3.4. Performance of Noncontaminated Data. In this section, we compare the performance of the standard boxplot, SIQR [17], and the proposed robust skewed boxplot using real data without outliers, the real rain observations from the three

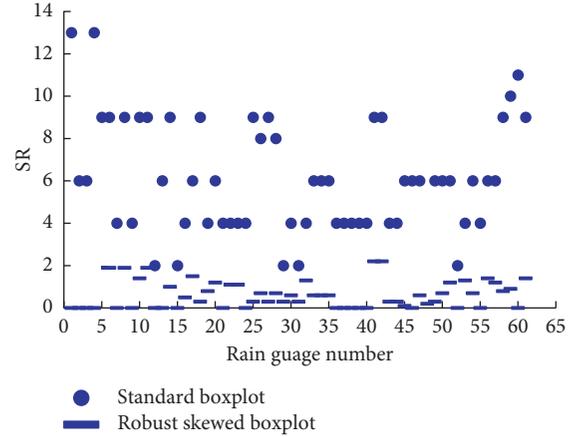


FIGURE 5: The average SR of each rain gauge.

basins. By calculating the swamping rate (SR) (the proportion of “good” data identified as outliers) [23] and accuracy (the proportion of outliers and “good” data identified correctly) [23], the differences in the three boxplots were analyzed. The SR and accuracy are defined as

$$SR = \frac{\text{the number of good data identified as outliers}}{\text{total number of good data}} * 100, \quad (14)$$

$$\text{accuracy} = \frac{\text{number of outliers correctly identified} + \text{number of good data correctly identified}}{\text{total number of data sets}} * 100. \quad (15)$$

When there is no outlier, we know that the total number of the good data was equal to the total number of datasets. In this case, SR plus accuracy equaled 1.0. Table 2 lists the average SR and accuracy for the three methods. Figure 5 displays the average SR of every rain gauge for the standard boxplot and robust skewed boxplot for the purpose of clarity.

Based on equation (14), SR is an estimator of method risk. From Table 2 and Figure 5, the results clearly showed that the robust skewed boxplot had a lower SR than the standard boxplot and SIQR without outliers. The SR of the proposed boxplot was 1.3% and was close to the portion outside of the standard boxplot fence for Gaussian data (0.7%). SIQR was slightly superior to the standard boxplot; however, the SR was still far greater than 0.7%. The accuracy of the robust boxplot was the highest in the three methods. The standard boxplot performed much worse than the SIQR.

We also observed that the proposed robust skewed boxplots for each rain gauge yielded much better SR values on the skewed uncontaminated observations than the other boxplots. This was due to the fact that the robust skewed boxplot used MC to adjust the fences to the skewed data. The detailed information on the boxplots is shown in Figure 6.

The rain observations ($HAMP \in (1, 2]$) at Wuyigong Station were selected to demonstrate that the proposed method was able to adjust the fence to the skewed data. The results are presented in Figure 6. MC equaled 0.52, and the distribution was skewed to the right.

It is clear that the robust skewed boxplot yielded the larger upper boundary than standard boxplot and SIQR, and it was adjusted to better reflect the right skewed data. The proposed boxplot identified fewer large good data points as the upper outliers. At the same time, the new method had less adjustment of the lower fence than the other methods that may result in the smallest good data being marked as the lower outliers (Figure 6). However, for the flood forecasting system, the influence effects of the larger good data identified as outliers on the forecasting results were far greater than those of the smallest good data identified as outliers. The robust skewed boxplot had practical value.

3.5. Performance under Contamination. We now compared the robustness of the three boxplots using the contaminated data. To understand the detailed information of the outliers, synthetic datasets were generated by superimposing the following upper outliers on the real rain observations in Section 2.2:

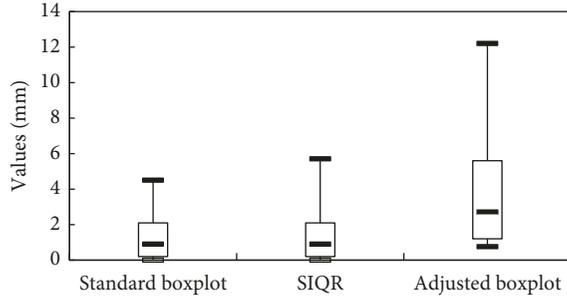


FIGURE 6: Boxplots of the three methods for Wuyigong Station.

$$e_i = \begin{cases} r \cdot T, & i = \text{int}\left(\frac{i}{L}\right) * L, \\ 0, & i \neq \text{int}\left(\frac{i}{L}\right) * L, \end{cases} \quad (16)$$

where r is a random number and T is a constant that controls the maximum of e . L is the frequency of outliers, for example, $L = 20$ means that the outlier percentage is 5%. By adjusting T and L , outliers of different magnitude and frequency could be generated.

We generated outlier samples and ran the experiment 1000 times for every T ($T = 5$ mm, 10 mm, 20 mm, 30 mm, 40 mm, 50 mm, and 100 mm) and $L = 20$.

The MC of different T values at Wuyigong Station in the Qilijie basin is listed in Table 3. We obtained comparable results for the other rain gauge stations.

The results in Table 3 showed that MC changes little when T increased. This demonstrated that MC was not influenced evidently by the outliers. The average performance of the three boxplots with outliers is shown in Table 4 when $T = 40$ mm.

Under contamination, the total number of good data did not equal the total number of the datasets, and SR plus accuracy did not equal 1.0.

Because of based on quantiles, the three boxplots all had the ability to resist outliers and they maintained robust results for noncontaminated and contaminated data (Tables 2 and 4). Compared with the standard boxplot and SIQR, besides the quantiles, the new boxplots used the MC that was robust to outliers and sensitive to skewness to construct the new fences. By moving the upper boundary up, the proposed method had a much lower SR and higher accuracy. This illustrated again that the proposed boxplot accounted sufficiently for skewness.

To further analyze the effects of different values of L on performance, we ran the experiment 1000 times for $L = 10$. The average performance of $L = 10$ is listed in Table 5 ($T = 40$ mm).

The outlier frequency changed from 5% to 10%, and the performance of the three boxplots varied a little. It was clear that the size of the outliers and the frequency had only a small effect on SR and accuracy. However, the results required some restrictive conditions, for example, the proportions of the outliers could not be too high.

TABLE 3: MC of different size outliers at Wuyigong Station.

T (mm)	MC		
	$0 < \text{HAMP} \leq 1$	$1 < \text{HAMP} \leq 2$	$2 < \text{HAMP}$
0	0.13	0.50	0.28
5	0.14	0.50	0.29
10	0.14	0.50	0.32
20	0.14	0.50	0.35
30	0.14	0.50	0.32
40	0.14	0.50	0.32
50	0.14	0.50	0.32
100	0.14	0.52	0.33
Standard error	0	0.007	0.018

TABLE 4: Performance of the three boxplots with $T = 40$ mm.

Method	SR	Accuracy
Standard boxplot	6.20	93.84
SIQR	5.85	94.00
Robust skewed boxplot	1.27	98.80

TABLE 5: Performance of the three boxplots with $L = 10$.

Method	SR	Accuracy
Standard boxplot	6.35	93.69
SIQR	5.90	93.92
Robust skewed boxplot	1.28	98.84

4. Conclusions

The standard boxplot is a popular nonparametric method to detect outliers in data series. Unfortunately, when it is used on skewed data, such as hourly rainfall series, the probability of identifying good data as outliers was high. Therefore, a MC that was not only robust to outliers but also sensitive to skewness, and different simple function styles were produced to adjust the standard boxplot fence to fit the skewed hourly rainfall distributions. The exponential function was then selected based on comparisons.

The comparison of the results using uncontaminated and abnormal data showed that the proposed method had robust performance and a lower risk of identifying good data as outliers, compared with the standard boxplot and SIQR.

In flood forecasting systems, the decision to eliminate data as outliers is a serious matter and should not be taken lightly. The unusual good observations often provide valuable information. Therefore, a more conservative approach (less risk of identifying good data as outliers), such as the robust skewed boxplot method, is reliable for practical applications.

Data Availability

The rainfall observation data used to support the findings of this study were supplied by the branch of hydrology and water resources investigation bureau of Fujian Province under license and so cannot be made freely available. Requests for access to these data should be made to the branch

of hydrology and water resources investigation bureau of Fujian Province, <http://www.fjsw.gov.cn/>.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

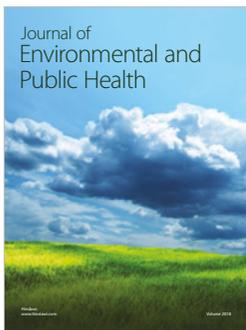
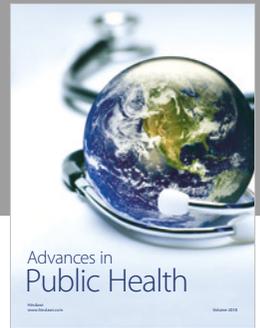
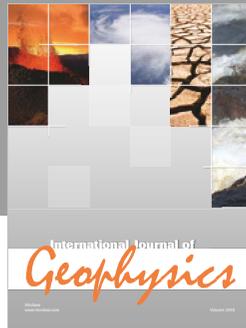
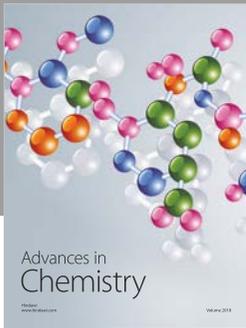
C. Z. proposed the research ideas and methods in the manuscript and was responsible for data collection and writing. H. L. and J. Y. suggested revisions for the paper.

Acknowledgments

This study was funded by the National Nature Science foundation of China (grant no. 50909084) and Nature Science Foundation of Fujian Province (grant nos. 2017J01721 and 2018J01525).

References

- [1] D. Han, I. D. Cluckie, D. Karbassioun, J. Lawry, and B. Krauskopf, "River flow modeling using fuzzy decision," *Water Resources Management*, vol. 16, no. 6, pp. 431–445, 2002.
- [2] C. Zhao, H. S. Hong, W. M. Bao, and L. P. Zhang, "Robust recursive estimation of auto-regressive updating model parameters for real-time flood forecasting," *Journal of Hydrology*, vol. 349, no. 3–4, pp. 376–382, 2008.
- [3] H. Liu, S. Shah, and W. Jiang, "On-line outlier detection and data cleaning," *Computers & Chemical Engineering*, vol. 28, no. 9, pp. 1635–1647, 2004.
- [4] V. Barnett and T. Lewis, *Outlier in Statistical Data*, John Wiley and Sons, New York, NY, USA, 3rd edition, 1995.
- [5] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 73–79, 2011.
- [6] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, NY, USA, 1986.
- [7] M. J. P. Mens, F. Klijn, K. M. de Bruijn, and E. van Beek, "The meaning of system robustness for flood risk management," *Environmental Science & Policy*, vol. 14, no. 8, pp. 1121–1131, 2011.
- [8] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, Boston, MA, USA, 1st edition, 1977.
- [9] D. C. Hoaglin, F. Mosteller, and J. W. Tukey, *Understanding Robust and Exploratory Data Analysis*, John Wiley and Sons, New York, NY, USA, 1983.
- [10] D. C. Hoaglin, B. Iglewicz, and J. W. Tukey, "Performance of some resistant rules for outlier labeling," *Journal of the American Statistical Association*, vol. 81, no. 396, pp. 991–999, 1986.
- [11] M. Frigge, D. C. Hoaglin, and B. Iglewicz, "Some implementations of the boxplot," *American Statistician*, vol. 43, no. 1, pp. 50–54, 1989.
- [12] W. S. Cleveland, *The Elements of Graphing Data*, Hobart Press, Summit, NJ, USA, 1985.
- [13] R. J. Hyndman and Y. Fan, "Sample quantiles in statistical packages," *American Statistician*, vol. 50, no. 4, pp. 361–365, 1996.
- [14] N. C. Schwertman, M. A. Owens, and R. Adnan, "A simple more general boxplot method for identifying outliers," *Computational Statistics & Data Analysis*, vol. 47, no. 1, pp. 165–174, 2004.
- [15] M. Bounessah and B. P. Atkin, "An application of exploratory data analysis (EDA) as a robust non-parametric technique for geochemical mapping in a semi-arid climate," *Applied Geochemistry*, vol. 18, no. 8, pp. 1185–1195, 2003.
- [16] Q. Zhou, S. N. Li, X. P. Li, W. Wang, and Z. G. Wang, "Detection of outliers and establishment of targets in external quality assessment programs," *Clinica Chimica Acta*, vol. 372, no. 1–2, pp. 94–97, 2006.
- [17] A. C. Kimber, "Exploratory data analysis for possibly censored data from skewed distributions," *Applied Statistics*, vol. 39, no. 1, pp. 21–30, 1990.
- [18] M. Huber and E. Vandervieren, "An adjusted boxplot for skewed distributions," *Computational Statistics & Data Analysis*, vol. 52, no. 12, pp. 5186–5201, 2008.
- [19] K. Carling, "Resistant outlier rules and the non-Gaussian case," *Computational Statistics & Data Analysis*, vol. 33, no. 3, pp. 249–258, 2000.
- [20] N. C. Schwertman and R. de Silva, "Identifying outliers with sequential fences," *Computational Statistics & Data Analysis*, vol. 51, no. 8, pp. 3800–3810, 2006.
- [21] N. J. Carter, N. C. Schwertman, and T. L. Kiser, "A comparison of two boxplot methods for detecting univariate outliers which adjust for sample size and asymmetry," *Statistical Methodology*, vol. 6, no. 6, pp. 604–621, 2009.
- [22] G. Brys, M. Hubert, and A. Struyf, "A robust measure of skewness," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 996–1017, 2004.
- [23] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.



Hindawi

Submit your manuscripts at
www.hindawi.com

