Analytical and Numerical Solutions for Nonlinear Equations
©Available online at https://ansne.du.ac.ir/
Online ISSN: 3060–785X
2026, Volume 11, Issue 2, pp. 1–9

Review article

Bootstrap Highest Density Confidence Interval by Comparing Two Climatological Regions

Ali Shadrokh[1,*], Mehdi Pejman[1]

[1] Department of Statistics, Payame Noor University, P.O.Box 19395–4697, Tehran, Iran

* **Corresponding author(s):** a.shadrokh@pnu.ac.ir

Received: 06/05/2026 **Revised:** 20/06/2026 **Accepted:** 27/06/2026 **Published:** xx/xx/2026 10.22128/ansne.2026.3310.1214

Abstract

Bootstrap is a resampling method based on high calculations, which can help us a lot for statistical inference in cases where the amount of data that we have is limited. For example, in the design of hydraulic structures such as bridges or dams, etc. there is a need to estimate hydrological events such as, floods or precipitations by statistical inference of quantiles of a probability distribution. In this paper, we aim to estimate precipitation quantiles. For calculating this estimation, the confidence interval for quantiles has been introduced with percentile bootstrap, accelerated bias-corrected bootstrap, t-bootstrap methods; that in this article, we want to compare these methods with the confidence interval made by the highest density method based on bootstrap data and we obtain the average length of the confidence intervals as a criterion to evaluate the methods. To calculate the average length of confidence intervals using different methods, first, the best distribution among commonly used distributions is fitted to the original data, and its parameters are estimated using the maximum likelihood method, and quantiles are obtained from it. Then, we continue until the coverage probability of the real quantile reaches the nominal confidence level of 95% by repeating the simulated bootstrap samples. The results of the performed simulation show that the bootstrap highest density method has the smallest average length of confidence intervals among all methods. The data used in the article are 24-hour annual maximum precipitation records in five meteorological stations in Mexico, which are compared with the data of five stations in Gilan province of Iran.

Keywords: Resampling bootstrap method, Quantile, Coverage probability, Extreme-Values.

Mathematics Subject Classification (2020): 60E05, 62F40, 62G09, 62G32

1 Introduction

Frequency analysis of hydrological data requires determining the magnitude of hydrological variables corresponding to a return period. The return period is defined as the average time interval (over a long period) at which the magnitude of a hydrological variable is reached or exceeded [1, 8]. The study and design of hydraulic structures require estimating hydrological events (such as floods, precipitation, etc.) through statistical inference from a probability distribution. For example, a specified quantile for a probability distribution represents the magnitude of a hydrological variable (such as river flow or precipitation amount) corresponding to that return period. Usually, determining quantiles is based on multi-year hydrological observations and statistical analysis of maximum event values. In hydrology, obtaining the magnitude of an event corresponding to a return period is not sufficient because there is a level of uncertainty in estimating the parameter and quantile from a finite sample, which needs to be considered in decision-making. One solution to this problem is obtaining confidence interval that includes the real value of the desired quantity with a high confidence level. Typically, the confidence interval for the desired quantity with an estimator $\hat{\theta}$ is obtained using the standard error estimator (S) of estimator $\hat{\theta}$. For example, a 95% confidence interval is obtained as $\hat{\theta} \pm 1.96S$ (assuming the distribution of the estimator $\hat{\theta}$ is approximately normal), [3]. However, there are many cases where the assumption of normality of the distribution of the estimator of the desired quantity is not valid, or it is difficult to estimate the standard error of the estimator. One solution to this problem is to obtain nonparametric confidence intervals, and an important method in this category is the bootstrap resampling method [5]. The distribution of the assumed statistic (such as mean, quantile, etc.) in the resampling proposed by [4] is obtained based on the available data. For this purpose, many samples are generated from the original sample using Monte Carlo simulation. There are two main methods for bootstrap resampling: parametric and nonparametric. In parametric bootstrap, random samples are generated from the distribution fitted to the data. In nonparametric bootstrap, bootstrap samples are constructed by resampling with replacement from the original sample. The parametric bootstrap is similar to Monte Carlo simulation, with the difference that the parameters used in the parametric bootstrap are estimated from the original sample, while the parameters in Monte Carlo simulation are based on the purposes of the simulation. However, these two methods are similar in practice [8].

Multiple methods have been introduced for constructing confidence intervals from bootstrap samples, among which the percentile bootstrap (PB), bias-corrected and accelerated bootstrap (BCA), t-bootstrap and bootstrap highest density interval (HDI). Each of these methods has its own advantages and disadvantages. However, it is not clear which method is more suitable for constructing confidence intervals for hydrological variables. One criterion for investigating which method is more effective in constructing confidence intervals is to calculate the coverage probability of each of the constructed confidence intervals.

The coverage probability is defined as the percentage of times that the constructed intervals include the real value of the quantity of interest (here, quantile). This coverage probability should be consistent with the confidence level of the constructed interval, i.e., for a 95% confidence level, it is expected that 95% of the constructed intervals cover the real value of the quantity of interest [10]. Another criterion for evaluating the accuracy of the confidence intervals obtained by these methods can be the shorter length of the confidence intervals. The third criterion is the smaller number of confidence intervals constructed to achieve the desired coverage probability. That is, the method that achieves the desired coverage probability with the fewest number of constructed intervals performs better.

From previous research, [2] have evaluated the types of bootstrap confidence intervals and their accuracy. [6] examined the types of resampling methods and hypothesis testing using bootstrap. [12] and [14] proposed a method for calculating the shortest confidence interval with the highest density. [19] obtained the coverage probability and mean length of confidence intervals for various bootstrap methods for symmetric continuous distributions such as normal, uniform, Cauchy, and Laplace. [8] investigated the coverage probability for quantiles with different return periods and resampling methods, including PB, BC, BCA, and MSB, and six common distributions in hydrology. [3] discussed the estimation of confidence intervals using three bootstrap methods, including PB, BCA, and t-bootstrap, for the path coefficients in structural equation modeling and the coverage probability of the real population value. [18] also investigated the average length of the confidence intervals with four bootstrap methods PB, BCA, t-bootstrap and HDI for five meteorological stations in Gilan province of Iran; the HDI method was proposed for the first time in that article and we also use it here.

In this article, firstly, the best-fitted distribution from among 15 widely used distributions in hydrological sciences based on the past experiences of researchers (such as, [7–9, 16]), for 24-hour annual maximum precipitation data recorded during at least 45 years, between 1955 and 2018, obtained from the data used in [8], is selected with the BIC criterion. Then, the coverage probability and mean length of the confidence intervals obtained by HDI bootstrap method were calculated and compared with three other methods. To calculate the coverage probability of the confidence intervals based on quantiles, it is necessary to calculate quantiles of the primary data and quantiles of the simulated data. Since the distributions under consideration have multiple parameters, the parameters of the fitted distributions were first estimated for the original data using the maximum likelihood method. Then the parameters were substituted into the distributions to obtain quantiles. Subsequently, the confidence bounds were estimated by sorting quantiles.

In summary, the article aimed to evaluate the accuracy of HDI bootstrap method for constructing confidence intervals for hydrological variables, using the example of 24-hour annual maximum precipitation data. The results were compared with three other commonly used methods, and HDI bootstrap method was found to be an effective tool for constructing confidence intervals with high coverage probability (0.95) and short length. In previous studies (such as, [8, 13, 19]), the coverage probability was obtained by fixing the sample size and the number of bootstrap replications (e.g., 1000) and selecting the closest coverage probability to the nominal value of 0.95 as the desired outcome. However, in this article and [18], to avoid excessive bootstrap replications, we considered different bootstrap sample sizes and continued the bootstrap replications only until we achieved a coverage probability of 0.95.

The structure of the article is as follows: In section 2, Materials and Methods, information related to the real data of Mexico, bootstrap confidence interval methods, and the algorithm are explained. In the third section, findings of the study including data frequency analysis; and comparison of bootstrap methods are presented. Finally, in the fourth section, the result of this research is compared with the result of the climatological data of five stations in Gilan province of Iran and discussion and conclusions are made.

2 Materials and Methods

2.1 Climatological Data Information

The data used here consists of 24-hour annual maximum precipitation records (in millimeters) at five weather stations in Mexico during at least 45 years; between 1955 and 2018, obtained from the data used by [8]), which is provided in the Appendix. A summary of the data is presented in Table 1 , along with descriptive statistics for the five weather stations.

Table 1. Descriptive statistics for five weather stations in Mexico

station	mean	standard deviation	Skewness
14067	111.49	67.92	1.52
14011	124.64	62.02	0.91
14148	102.14	32.87	1.04
10024	47.09	23.18	1.48
12069	122.71	67.11	1.64

2.2 Bootstrap Confidence Interval Methods

Several methods have been introduced for constructing confidence intervals from bootstrap samples. Among them, we used the percentile bootstrap (PB), the bias-corrected and accelerated bootstrap (BCA), the t-bootstrap and the bootstrap highest density interval (HDI), these methods are discussed in [18].

2.3 Determining the Coverage Probability and the Length of the Confidence Interval

To determine the coverage probability and the length of the confidence intervals, we follow a similar process used in [11], [8] and [18]. We select the best-fitted distribution to the data among 15 commonly used distributions in hydrology based on the BIC criterion (table 2). The list of these distributions includes log-gamma (LOGGM), log-Pearson type 3 (LP3), Gumbel (GMB), generalized extreme values (GEV), log-logistic (LLOG), generalized logistic (GLOG), Weibull, log-normal (lnorm), gamma, normal, exponential, Frechet, generalized Pareto (genPareto), Pearson type 3, and three-parameter log-normal (lnorm3). To calculate the coverage probability of the confidence intervals based on the bootstrap quantiles, we need to calculate quantile of initial sample and quantile of simulated data. The corresponding quantiles to return periods of T="10, 100, 1000" are calculated. According to Table 3 in subsection 3.1, among 15 distributions, the Gumbel distribution often provided the best fit to the data, being selected as the best distribution in 3 out of 5 stations, while in one station, two-parameter log-normal distribution and another station two-parameter distribution of Frechet was selected as the best distribution. quantiles of these distributions are calculated using Equations (1) to (3):

$$\hat{x}_T = \hat{\alpha} - \hat{\beta}\ln\left[\ln\left(\frac{T-1}{T}\right)\right], \tag{1}$$

$$\hat{x}_T = \exp\left[\hat{\alpha} + \hat{\beta}\Phi^{-1}\left(\frac{T-1}{T}\right)\right], \tag{2}$$

$$\widehat{x}_T = \hat{\beta}\left(\ln\left(\frac{T}{T-1}\right)\right)^{-\frac{1}{\hat{\alpha}}}, \tag{3}$$

where β is the scale, α is the location parameter in the Gumbel distribution and also, α is the shape parameter in log-normal distribution. Since quantile of the above distributions includes multiple parameters, we first estimate the parameters of the fitted distributions to the original data using the maximum likelihood method; and then obtain quantiles based on them. Then, we estimate the confidence intervals by ordering quantiles. It should be noted that if we try different values n and B=1000 to see for what values of n and B, we can achieve a 95% coverage probability. We used the R software for simulating data and bootstrap repetitions. We used the gofstat function in the fitdist package for fitting distributions to the data and the boot.ci function in the boot package for the bootstrap methods BP, BCA, and t-bootstrap. We also used the hdi function in the HDInterval package for the HDI method ([17] and [15]).

3 Findings

Figure 1 shows the box plot of the annual maximum precipitation amount (in millimeters) at different stations during at least 45 years. As observed, the average annual maximum precipitation at 14011 station is higher than the others, and the average annual maximum precipitation at 10024 station is lower than the others. This figure is drawn with RStudio software.

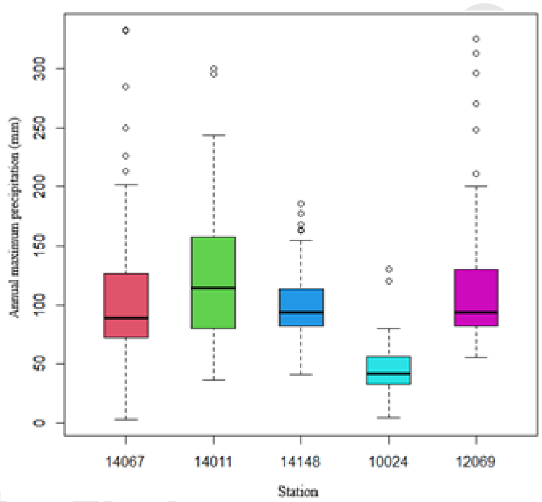

Figure 1. Annual maximum precipitation recorded at different stations

3.1 Frequency Analysis

The first step in frequency analysis is determining which distribution best fit annual maximum precipitation data. The results show that among the 15 selected two-parameter and three-parameter distributions, the Gumbel distribution often provided the best fit to the data, being selected as the best distribution in 3 out of 5 stations. The selection of the best distribution is based on the BIC criterion, which under equal goodness-of-fit conditions for two-parameter and three-parameter distributions for a given data series, selects the two-parameter distribution, i.e., the model with fewer parameters. Table 2 shows the BIC values for the fitted distributions at 14067 station, where the best two-parameter is Gumbel, that bolded in Table 2 . The best fitted distribution was selected for each station as the parent distribution for generating random samples. Table 3 shows the estimated parameters of the best fitted distribution.

Table 2. The BIC values for the fitted distributions at 14067 station

distribution	**Gumbel**	log-logistic	generalized extreme values	generalized logistic	log-gamma
BIC	**706.61**	708.24	709.36	708.41	746.03
distribution	log-Pearson type 3	Weibull	log-normal	gamma	normal
BIC	733.06	712.53	721.93	710.45	728.87
distribution	exponential	Frechet	generalized Pareto	Pearson type 3	three-parameter log-normal
BIC	735.54	771.92	714.56	712.65	710.58

Table 3. The estimated parameters of the best fitted distribution

station	the best distribution	α	β
14067	Gumbel	82.96	46.74
14011	log normal	4.70	0.49
14148	Gumbel	87.66	24.75
10024	Gumbel	37.08	17.89
12069	Frechet	89.19	2.93

3.2 Comparison of Bootstrap Methods

After selecting the best fitted distribution for the data, it was used to estimate the corresponding quantiles for different return periods T=10, 100 and 1000 years. Then, the average length of confidence intervals was constructed using four bootstrap methods with B repetitions and different sample sizes n. According to Table 4, the average length of these confidence intervals increases with increasing return periods. The widest confidence intervals belong to the t-bootstrap method, and the shortest ones belong to the HDI method. The widest average length of confidence intervals belongs to Frechet distribution, while the shortest average length of confidence intervals belongs to the Gumbel distribution. A combination of the best fitted distributions, return periods, sample size (n) from the parent distribution, and four bootstrap methods (with the number of repetitions in parentheses) for calculating the average length of confidence intervals in five stations is presented in Table 4. According to Table 4, the average length of confidence intervals using bootstrap methods shows that the HDI method has the shortest average length of confidence intervals among the four methods and fitted distributions in all stations. For example, Figure 2, which is drawn with Excel, shows the average length of confidence intervals using different bootstrap methods and different return periods at 14067 station for the Gumbel distribution. Similar figures can be plotted for other distributions and stations. In terms of the required number of repetitions (B) to achieve a nominal 95% confidence level, the HDI and t-bootstrap methods usually require fewer repetitions than other methods. For example, in Figure 3, which is drawn with RStudio software, the required number of repetitions to achieve a 95% confidence level for the Gumbel distribution with T=10 and n=80 using the BP method at 14067 station is shown.

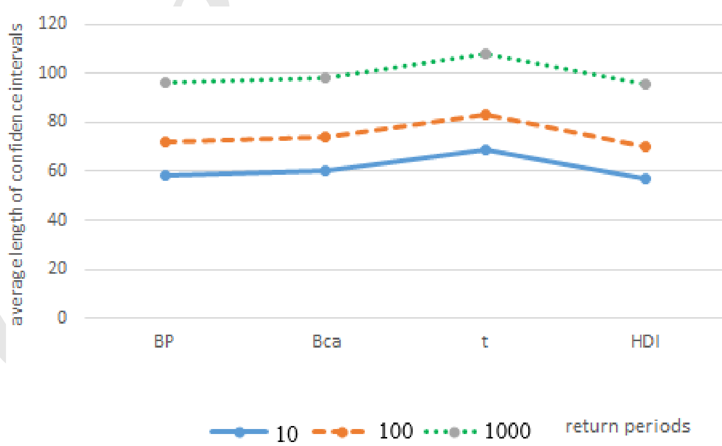

Figure 2. Average length of confidence intervals with different bootstrap methods and different return periods at 14067 station for the Gumbel distribution

4 Comparison with Five Stations in Gilan Province

The data used in the stations of Mexico are the 24-hour annual maximum precipitation records in five climatological stations during at least 45 years; between 1955 and 2018, and the data used in the stations of Gilan province are the 24-hour annual maximum precipitation records in five climatological stations during 20 years, from 1999 to 2020. According to the results obtained from the data analysis of five stations in Mexico and five stations in Gilan province, respectively, are shown in Table 4 and the table obtained in [18], it was observed that

Table 4. Average length (with repetition in parenthesis) of confidence intervals using bootstrap methods

Station	Fitted Distribution	T	n	BP	BCA	t	HDI
14067	Gumbel	10	80	58.66(20)	60.33(20)	69.03(20)	57.20(20)
		100	80	72.11(20)	73.92(20)	83.45(60)	70.36(20)
		1000	100	96.27(20)	98.50(20)	108.12(40)	95.36(20)
14011	log-normal	10	160	49.19(80)	51.26(60)	54.22(20)	48.05(100)
		100	100	196.95(40)	210.44(40)	178.52(20)	142.91(20)
		1000	80	334.62(160)	385.57(40)	424.49(60)	328.38(140)
14148	Gumbel	10	80	20.68(40)	20.90(40)	23.83(40)	20.48(20)
		100	220	24.61(20)	24.87(20)	25.78(20)	23.96(20)
		1000	200	39.37(20)	40.89(160)	48.03(40)	38.80(20)
10024	Gumbel	10	180	11.23(20)	11.39(20)	12.12(40)	11.00(20)
		100	180	19.96(20)	20.18(20)	22.36(20)	19.70(20)
		1000	180	28.87(20)	29.18(20)	32.61(20)	28.42(20)
12069	Frechet	10	120	57.03(20)	58.29(20)	62.72(20)	55.38(20)
		100	120	213.72(20)	222.90(20)	239.38(40)	206.23(20)
		1000	120	669.07(20)	702.61(20)	755.98(40)	637.83(20)

due to the large number of initial samples in Mexico, more samples are needed to reach a coverage probability of 0.95. In both countries, the best distributions fitted to the data among the 15 most used distributions in climatological data were two-parameter distributions; in Mexico, Gumbel, Frechet and Log Normal distributions were selected and in Gilan, Frechet, and gamma distributions were selected. In both countries, confidence interval length increases as the return period increases. Also, in both countries, the shortest confidence interval length among the four methods was related to the HDI method. In both countries, the widest confidence interval belonged to the t method, and the duration of the implementation of this method in R software was also longer than the other methods. In terms of the relationship between the used distribution and the length of the confidence interval, in both countries, the station that had the lowest maximum 24-hour precipitation between the stations also had the shortest length of the confidence interval; in Gilan province, Manjil station with gamma distribution and in Mexico, station of 10024 with Gumbel distribution had these conditions.

5 Conclusion

To analyze the frequency of maximum values of hydrological events, both two-parameter and three-parameter distributions should be considered, and a criterion for selecting the best fitted distribution to the data, such as BIC, should be used. Based on the fitted distributions for different stations data in Mexico and Gilan Province, two-parameter distributions always had a better fit for the examined data than three-parameter distributions; the best distribution for the data of Mexico is the Gumbel distribution, which was selected as the best distribution in 60% of the stations. Also, the best distribution for the data of Gilan province is the Frechet distribution, which was selected as

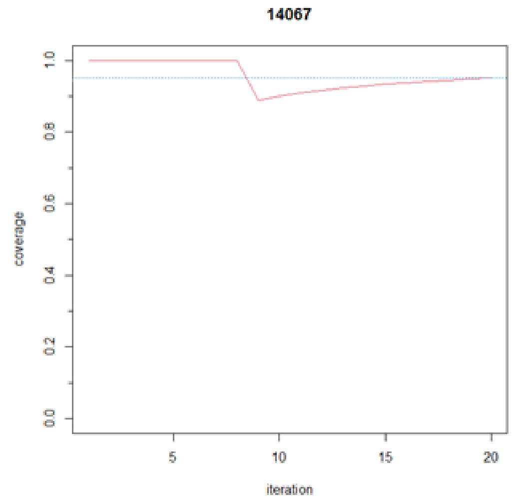

Figure 3. Number of repetition and coverage probability for the Gumbel distribution with $T = 10$ and $n = 80$ and BP method at 14067 station

the best distribution in 80% of the stations. It can be used for similar studies that deal with annual maximum precipitation values. Generally, the widest confidence intervals were obtained using BCA and t-bootstrap methods, while the shortest confidence intervals were obtained using HDI and BP methods. Finally, the HDI method had the shortest average confidence interval length among all methods. In terms of the distribution used and the length of the confidence interval, in Mexico, station 10024 with two-parameter Gumbel distribution and in Gilan, Manjil station with two-parameter gamma distribution has provided narrower confidence intervals that these stations had the lowest 24-hour annual maximum precipitation in both countries. To obtain the probability coverage of quantile exactly equal to the confidence level of the 95%, the number of different samples can be considered from the simulated parent distribution, which usually two-parameter distributions require a smaller sample size than three-parameter distributions. Also, the larger the number of initial samples, the more samples we needed to simulate the parent distribution and reach the 95% confidence level. In terms of the required number of repetitions (B) to achieve the nominal 95% confidence level, the t-bootstrap method usually requires fewer repetitions than other methods, although the implementation time of this method was longer in R software.

Authors' Contributions

The authors equally contributed to this work. All authors read and approved the final manuscript.

Data Availability

All data in the paper are available from the corresponding authors upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethical Considerations

The authors have diligently addressed ethical concerns, such as informed consent, plagiarism, data fabrication, misconduct, falsification, double publication, redundancy, submission, and other related matters.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or nonprofit sectors.

Acknowledgments

We thank our colleagues and agencies for their support.

References

[1] K.S. Cheng, J.L. Chiang, and C.W. Hsu, Simulation of probability distributions commonly used in hydrological frequency analysis, Hydrol. Process., 2, 51–60, (2007).

[2] A.C. Davison, and D.V. Hinkley, Bootstrap Methods and Their Application, 1st ed., Cambridge University Press, Cambridge, UK, (1997).

[3] T.J. DiCiccio, and B. Efron, Bootstrap confidence intervals, Stat. Sci., 11, 189–228, (1996).

[4] B. Efron, Bootstrap methods: Another look at the jackknife, Ann. Stat., (1979).

[5] B. Efron, Non-Parametric Standard Errors and Confidence Intervals, The Canadian Journal of Statistics, 9, 139–158, (1981).

[6] B. Efron, and R. Tibshirani, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Statistical Science, 1, 54–75, (1986).

[7] T. Fatih, and K. Ercan, Definition of the best probability distribution functions for annual minimum flows in the rivers of the Upper Euphrates River Basin, Turkey, IOP Conf. Series: Materials Science and Engineering, 737, 012166, (2020).

[8] R.S. Flowers-Cano, R. Ortiz-Gómez, J.E. León-Jiménez, R. López Rivera, and L.A. Perera Cruz, Comparison of Bootstrap Confidence Intervals Using Monte Carlo Simulations, Water, 10, 166, (2018).

[9] S. Grimaldi, S.C. Kao, A. Castellarin, S.M. Papalexiou, A. Viglione, F. Laio, H. Aksoy, and A. Gedikli, Statistical Hydrology, Peter Wilderer (ed.) Treatise on Water Science, 2, 479–517, Oxford: Academic Press, (2011).

[10] M.J. Hall, H.F.P. Van den Boogaard, R.C. Fernando, and A.E. Mynett, The construction of confidence intervals for frequency analysis using resampling techniques, Hydrol. Earth Syst. Sci, 8, 235–246, (2004).

[11] M. Hall, and M.S. Mayo, Bootstrap Confidence Intervals and Coverage Probabilities of Regression Parameter Estimates Using Trimmed Elemental Estimation, Journal of Modern Applied Statistical Methods, 7, 2, (2008).

[12] R.J. Hyndman, Computing and graphing highest density regions, American Statistician, 50, 120–126, (1996).

[13] K. Jung, J. Lee, V. Gupta, and G. Cho, Comparison of Bootstrap Confidence Interval Methods for GSCA Using a Monte Carlo Simulation, Front. Psychol, 10, 2215, (2019).

[14] J. Kruschke, Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan, ISBN: 978-0-12-405888-0, (2011).

[15] M. Meredith, and J. Kruschke, HDInterval R package version 0.2.2., https://github.com/mikemeredith/HDInterval.git, (2020).

[16] I. Ologhadien, Selection of Probabilistic Model of Extreme Floods in Benue River Basin, Nigeria, European Journal of Engineering and Technology Research, 6, 1, (2021).

[17] M.L. Rizzo, Statistical computing with R, Second edition, Boca Raton, CRC Press, Taylor & Francis Group, (2019).

[18] A. Shadrokh, M. Pejman, and A. Mohammadpour, Bootstrap Highest Density Confidence Interval with Application to Climatological Data, Journal of Andishe Amari, 28(1), 15–25, (2023).

[19] Y. Zhu, and J. Kolassa, Assessing and Comparing the Accuracy of Various Bootstrap Methods, Communications in Statistics - Simulation and Computation, (2017).