International Journal of Ecological Economics and Statistics;
Volume. 39, Issue No. 3; Year 2018; Int. J. Ecol. Econ. Stat.;
ISSN 0973-1385 (Print), ISSN 0973-7537 (Online)

Accurate Confidence Intervals for Weibull Percentiles Using Bootstrap Calibration: A Case Study of Weekly Rainfall in Sri Lanka

H. P. T. N. Silva[1] and **T. S. G. Peiris**[2]

[1]Department of Social Statistics, Faculty of Humanities and Social Science,
University of Sri Jayewardenepura, Sri Lanka
Email: thanuja@sjp.ac.lk

[2]Department of Mathematics, Faculty of Engineering,
University of Moratuwa, Sri Lanka
Email: sarathp@uom.lk

ABSTRACT

Modeling rainfall percentiles in the context of the confidence interval is an appropriate technique that can be employed to make inferences about the rainfall characteristic. The coverage probability of the confidence interval is one of the imperative factor that should be considered when making inferences. Accurate confidence bands enhance the degree of the awareness level of rainfall variability at high uncertainty. The main aim of this study is to find the accurate level of confidence intervals for weekly rainfall percentiles derived from Weibull distributions based on the real coverage probabilities which are formed using bootstrap calibration. Weekly rainfall data from 1970 to 2015 in the Colombo city were used for this analysis. A simulation was carried out based on the one weekly series (week 24; 11-17 June) using the bootstrapping approach. It was found that the data series pertaining to the week 24 is well fitted with the two parameter Weibull distribution. Furthermore, the result reveals that the real coverage probabilities of 95% confidence intervals of 50th, 60th, 70th, 80th and 90th weekly rainfall percentiles which were derived using maximum likelihood estimators of Weibull distribution can be attained on average at the levels 95.901%, 97.501%, 97.603%, 97.680% and 97.910% respectively.

Key Words: Percentiles, Coverage Probability, Bootstrapping, Confidence Intervals, Weekly Rainfall

Mathematics Subject Classification: 62F25

Journal of Economic Literature (JEL) Classification : C15

1. INTRODUCTION

An accurate analysis of the pattern of rainfall is essential to make effective decisions by utilizing water resources in many fields. Modelling rainfall percentile is one of the successful technique that can be applied to describe the rainfall characteristics in any region. Information on rainfall percentiles would enhance the level of awareness of rainfall behaviour, which can be used to reduce the difficulties that exist due to changes of atmospheric behaviour. The complexity of the temporal pattern of the rainfall is high in the short range scales of weekly and hourly than monthly, seasonal and annual. However, it is more important to have prior information on weekly rainfall in the urban areas which those were engaged with the large number of activities related in many fields such as industrial, constructions, health, rain water harvesting etc.. Moreover, the occurrences of extreme rainfall events in high population density areas have had a significant negative impact on the lives of the people and the infrastructure of the city. Most of the urban cities including the city of Colombo in Sri Lanka were vulnerable to many water related issues derived from the erratic rainfall events caused by changes in rainfall patterns, urbanization and installation of complex infrastructure by Lo and Koralegedera (2015).

The rainfall of the country is strongly governed by the seasonal varying monsoon system. According to Domroes (1974), a seasonal monsoon system of the country can be mainly divided into four major periods; First Inter Monsoon (FIM) from March to April, South West Monsoon (SWM) from May to September, Second Inter Monsoon (SIM) from October to November and North East Monsoon (NEM) from December to February.

Many researchers have made attempts to describe the temporal behaviour of the rainfall based on the point estimates for rainfall percentiles which derived fitted probability distributions (Sharda and Das, 2005, Sharma and Singh, 2010, Mishra et al., 2013), However, it is more imperative and practical to form a range for percentile rather than express it by a single value at the high uncertainty of climatic behaviour. Confidence intervals can be used to make inferences not only for the rainfall quantity, but also to have an idea about the its variability at the particular level of uncertainty. Some of the researchers employed confidence intervals to describe the characteristics of the rainfall quantile (Dunn, 2002, Park et al., 2001, Silva and Peiris, 2017).

Accurate estimates, either point or intervals is essential for better planning. The coverage probability of confidence intervals is one of the essential aspects to be considered to make more accurate interval estimates. In such a situation, the sample size is the other main factor which influences accurate inferences. In fact, it is more complicated to compose inferences and make decisions based on the small sample size. Most of the time, estimates derived from the fitted theoretical probability distributions becomes inaccurate due to the small sample size. To overcome this problem, the bootstrapping technique can be used. Past studies have shown three parameter and two parameter Weibull distributions were well fitted to the rainfall data, especially on the weekly scales (Sharda and Das, 2005; Silva and Peiris, 2017). In view of the above, the main objective of this study is to find the accurate confidence interval levels for weekly rainfall percentiles formed from two parameter Weibull distributions based on the real coverage probability derived from bootstrap calibration.

2. MATERIAL AND METHODS

2.1 Study area and data description

The city Colombo is situated with latitudes 6^0 93' N and Longitude 79^0 86' E in Sri Lanka and is selected as the study site . The city Colombo is the commercial capital of Sri Lanka. Daily rainfall data from 1970 to 2015 during the period of SWM in Colombo city were used for this analysis. These data were obtained from the Colombo meteorology station which is the only station with available meteorology data of Colombo from the Department of Meteorology in Sri Lanka. The daily rainfall (mm) data has been converted into weekly rainfall as Week 1 corresponding to 1-7 January, Week 2, Week 3 and so on corresponding to 8-14 January, 15-21 January and so on. Based on the above classification, the weeks 18-39 (30th of April to 30th of September) belongs to the SWM.

To work out the coverage probability, one weekly data series should be considered during the SWM as the main data series. In this study, the data which belongs to the week 24 (11-17 June) in SWM (The weekly data of 46 years for the time span from 1970 to 2015) was considered as the main data set. Those data were fitted to many theoretical probability distributions and two tests; Anderson Darling and Kolmogorov -Smirnov were used to test the goodness of fit of the parametric distributions.

2.2. Weibull distribution

The Weibull distribution is widely used for climatic data analysis due to its properties. The probability density function (Pdf) of a weibully variable, X with scale parameter 'α' and shape parameter 'β' is given by equations (1).

$$f(x) = \frac{\beta}{\alpha^{\beta}} x^{\beta-1} \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right) \qquad X \geq 0,\ \alpha > 0,\ \beta > 0 \qquad (1)$$

where, 'α' is the scale parameter and 'β' is the shape parameter of the distribution.

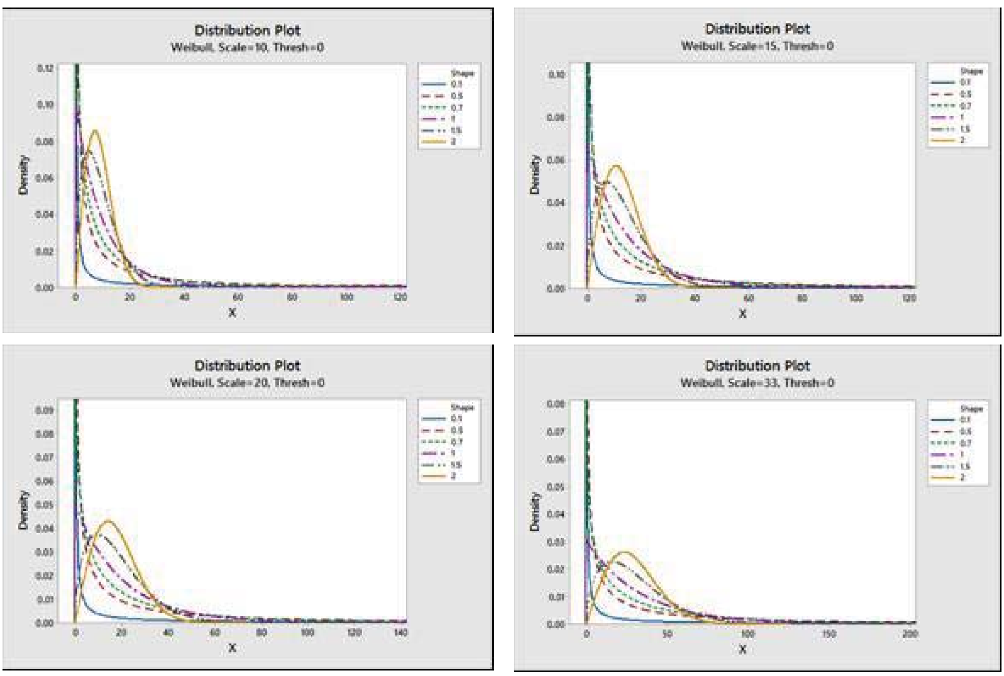

Figure 1: Probability density functions of Weibull distribution with different scale and shape parameters

The shape of the density function of the Weibull distribution changes drastically with the value of the shape parameter as shown in Figure 1.

2.3. (1-α) % Confidence intervals for the p^{th} percentile of Weibull distribution

The Weibull distribution can be approximated to the normal distribution when shape parameter is about 3.6 (Johnson and Kotz,1970). The p^{th} percentile of the Weibull distribution (Xp) and its variance are defined by equations (2) and (3) (Heo, et al., 2001).

$$\hat{X}_p = \hat{\alpha}\left[-\ln(1-p)\right]^{1/\hat{\beta}} \tag{2}$$

$$Var(\hat{X}_p) = \frac{\hat{X}_p^2}{\hat{\alpha}^2}Var(\hat{\alpha}) + \frac{\hat{X}_p^2}{\hat{\beta}^4}Z_p^2 Var(\hat{\beta}) - 2\frac{Z_p\hat{X}_p^2}{\hat{\alpha}\hat{\beta}^2}Cov(\hat{\alpha},\hat{\beta}) \tag{3}$$

where $Z_p = \ln[-\ln(1-p)]$ and $\hat{\alpha}$ *and* $\hat{\beta}$ are the maximum likelihood estimators for the α and β respectively. The equations that used to calculate the (1- α)% confidence intervals for the Weibull percentiles under the normal approximation is given by equation (4);

$$\left[\exp\left(\ln(\hat{X}_p) - Z_{\alpha/2}\frac{\sqrt{Var(\hat{X}_p)}}{\hat{X}_p}\right),\ \exp\left(\ln(\hat{X}_p) + Z_{\alpha/2}\frac{\sqrt{Var(\hat{X}_p)}}{\hat{X}_p}\right)\right] \tag{4}$$

2.4. Coverage probability

The coverage probability of a confidence interval can be briefly explained as the proportion of the time that interval contains the true value of interest. The coverage probability of a confidence interval can be calculated using simulation method. Firstly, number of samples of size n are simulated based on the main data series to compute the confidence intervals for interest parameter for each sample. After that, it should be computed for the proportion of samples for the known population parameters contained in the confidence intervals. That proportion is an estimate for the coverage probability for the confidence interval. However, a discrepancy can occur between the computed coverage probability and the nominal coverage probability. In this study, our interest parameter is percentile. The coverage probability will be calculated based on the confidence intervals of percentiles (P_{50}, P_{60}, P_{70}, P_{80} and P_{90}) .

2.5. Simulation

Assume that the major data series (The week 24) was well fitted with the two parameter Weibull distribution with scale parameter α and shape parameter β. Based on the size of the main data series (N=46), 2000 random samples (each sample size is also equal to 46) were generated using a bootstrapping approach called as Sample1, Sample2, Sample3, ... Sample2000 from the Weibull distribution (α, β). Furthermore, the scale and shape parameters of data sets pertaining to the Sample1 $(\hat{\alpha}_1, \hat{\beta}_1)$, Sample2$(\hat{\alpha}_2, \hat{\beta}_2)$, Sample3$(\hat{\alpha}_3, \hat{\beta}_3)$ and so on were estimated using maximum likelihood (MLE) method. Five percentiles (P_{50}, P_{60}, P_{70}, P_{80} and P_{90}) were calculated for each sample (Sample1 to Sample2000). It again generated 300 samples (Same sample size (n=46)) based on the generated Sample1, Sample2, Sample3 etc.. Those 300 samples which derived from the Sample1 could be indicated as $Sam1_1$, $Sam1_2$,...., $Sam1_{300}$. Here it describes only the coverage probability of randomly selected four samples (Sample68, Sample423, Sample802 and Sample1551). When considering the 300 samples generated based on the Sample1, firstly, the 50th percentiles and corresponding 95% confidence intervals were calculated of $Sam1_1$, $Sam1_2$, $Sam1_3$ and so on. The coverage probability was calculated based on the 300 confidence intervals (95%). The same procedure was carried out to calculate the coverage probability of confidence intervals at 95.2%, 95.4%, 95.6%, 95.8%, 96%, 96.2%, 96.4%, 96.6%, 96.8%, 97%, 97.2%, 97.4%, 97.6%, 97.8% and 98% confidence levels. Other samples which were generated from the Sample2,....., Sample2000 were applied the above procedure and the corresponding coverage probabilities for each confidence level listed above were calculated.

3. RESULTS

The summary statistic of the total rainfall during week 24 is presented in Table 1 along with the histogram (Figure 2).

Table 1: Descriptive Statistics of the weekly rainfall data (week 24)

Variable	No. of Data	Mean	StDev	Median	Mini	Max	Coefficient of Variance(%)	Skewn ess
Week 24	46	36.1	37.9	18.4	0.1	146.3	104.9	1.37

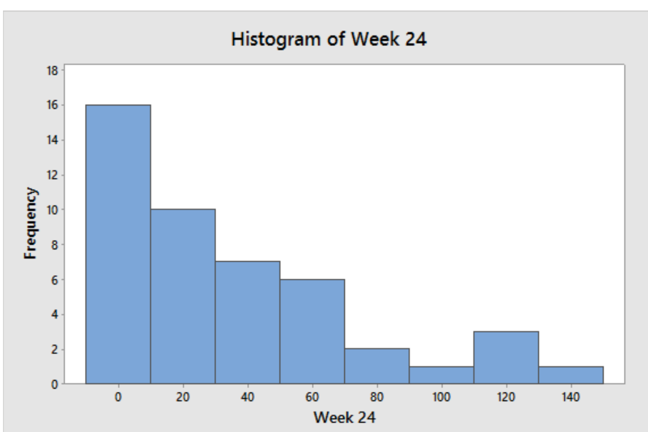

Figure 2: Histogram of weekly rainfall data (week 24)

From 1970 to 2015 total weekly rainfall in the 24th week varied from 0.1mm to 146.3mm with a mean 36.1mm. Figure 2 illustrates the weekly rainfall as are positively skewed with a longer tail to the right and the result was further confirmed as the coefficient of skewness is 1.37. The large coefficient of variance (104.9%) gives evidence to high fluctuations in weekly rainfall (week24).

The total rainfall during week 24 was fitted to different type of probability distributions such as Normal, Exponential, Gamma, Weibull and so on and those data were well fitted with the two parameter Weibull distribution. Corresponding Anderson Darling and Kolmogorov -Smirnov test statistics were 0.258(P-value =0.256) and 0.0673 (P-value = 0.941) respectively. The maximum likelihood estimates for the scale and shape parameters of the fitted Weibull distribution were 33.9286 and 0.8775 respectively. The coverage probabilities of confidence intervals of Sample68 at different uncertainty levels are presented in Table 2. The corresponding coverage probabilities of the confidence limits of Sample423, Sample802 and Sample1551 are presented in Table 3, Table 4 and Table 5 respectively.

Table 2 : Coverage Probabilities of five Percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample68

Confidence Level (%)	**Coverage Probability**				
	P_{50}	**P_{60}**	**P_{70}**	**P_{80}**	**P_{90}**
95.0	93.00	93.00	93.00	92.00	92.67
95.2	93.33	93.33	93.00	92.67	93.00
95.4	93.33	93.67	93.00	93.00	93.00
95.6	93.67	93.67	93.00	93.00	93.67
95.8	94.00	93.67	93.00	93.00	93.67
96.0	94.33	93.67	93.67	93.33	94.00
96.2	**95.33**	94.33	94.00	94.00	94.67
96.4	95.67	**95.00**	94.00	94.67	**95.00**
96.6	96.33	95.33	**94.67**	94.67	95.00
96.8	96.33	95.67	**95.00**	**94.67**	95.00
97.0	96.67	96.00	95.00	**95.33**	95.00
97.2	96.67	96.33	95.33	96.00	95.00
97.4	96.67	96.33	95.67	96.00	95.33
97.6	96.67	96.33	96.00	96.00	95.33
97.8	96.67	96.33	96.00	96.33	95.67
98.0	96.67	96.33	96.67	96.33	96.33

According to the above table, it is clear that 95% coverage probability of P_{50} can be attained at 96.2 % confidence level. The 95% coverage probability of P_{60}, P_{70}, P_{80} and P_{90} can be reached at the confidence levels 96.4%, 96.6%, 96.8% and 96.4% respectively.

Table 3: Coverage Probabilities of five Percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample 423

Confidence Level (%)	Coverage Probability				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
95.0	94.00	93.67	90.67	87.67	90.67
95.2	94.00	94.00	91.00	88.67	91.67
95.4	94.00	94.00	91.00	89.00	91.67
95.6	94.00	94.33	91.67	89.00	91.67
95.8	94.33	94.33	91.67	89.00	92.00
96.0	94.33	94.33	92.00	89.33	92.00
96.2	94.33	94.67	92.00	89.33	92.33
96.4	94.67	94.67	92.33	90.00	93.00
96.6	**95.00**	94.67	92.33	91.33	93.67
96.8	95.67	94.67	92.67	91.67	94.00
97.0	95.67	**95.00**	93.00	92.00	94.67
97.2	96.00	96.00	93.33	92.67	94.67
97.4	96.67	96.67	93.33	93.00	94.67
97.6	97.00	97.00	93.67	93.67	**95.00**
97.8	97.33	97.67	94.67	94.33	95.33
98.0	97.67	97.67	**95.00**	**95.00**	95.33

Based on the Table 3, the real 95% coverage probability of P_{50}, P_{60}, P_{70}, P_{80} and P_{90} can be obtained at the 96.6%, 97.0%, 98.0%, 98.0%, and 97.6% respectively.

Table 4: Coverage Probabilities of five Percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample 802

Confidence Level (%)	Coverage Probability				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
95.0	94.33	94.33	90.67	89.33	89.33
95.2	94.33	94.33	90.67	89.33	89.67
95.4	94.33	**94.67**	90.67	89.67	90.00
95.6	**95.33**	**95.33**	91.33	89.67	90.00
95.8	95.67	95.33	91.67	89.67	91.33
96.0	95.67	95.33	91.67	90.00	91.33
96.2	96.00	95.67	92.00	90.00	91.67
96.4	96.00	95.67	92.67	90.00	92.00
96.6	96.33	96.00	92.67	90.67	93.33
96.8	96.33	96.00	92.67	91.00	93.33
97.0	96.33	96.33	93.33	91.33	94.00
97.2	96.67	96.33	94.00	91.67	94.67
97.4	97.33	96.67	94.33	92.00	94.67
97.6	97.33	96.67	94.67	93.00	94.67
97.8	97.67	97.00	**95.00**	94.00	**95.00**
98.0	97.67	97.00	95.00	**95.00**	95.33

Table 4 illustrates the 95% coverage probability of P_{50}, P_{60}, P_{70}, P_{80} and P_{90} obtained at the 95.6%, 95.4%, 97.8%, 98.0%, and 97.8% confidence levels respectively.

Table 5: Coverage Probabilities of five Percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample 1551

Confidence Level (%)	Coverage Probability				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
95.0	92.67	92.33	92.00	92.00	89.67
95.2	93.00	92.33	92.00	92.33	89.67
95.4	93.33	92.67	92.00	92.67	89.67
95.6	93.33	92.67	93.33	92.67	92.33
95.8	**94.67**	93.00	93.33	93.00	92.33
96.0	95.33	93.00	93.33	93.00	92.33
96.2	96.00	93.00	93.67	93.67	93.33
96.4	96.00	93.33	93.67	93.67	93.33
96.6	96.00	93.33	94.00	94.00	93.67
96.8	96.00	93.67	94.00	94.33	93.67
97.0	96.33	93.67	94.67	94.67	94.67
97.2	96.33	93.67	94.67	94.67	94.67
97.4	96.67	94.67	**95.00**	94.67	94.67
97.6	96.67	94.67	95.00	**95.00**	**95.00**
97.8	96.67	**95.00**	95.33	95.00	95.33
98.0	97.67	95.00	95.33	95.33	95.33

Table 5 shows that the 95% coverage probability of P_{50}, P_{60}, P_{70}, P_{80} and P_{90} obtained at the 95.8%, 97.8%, 97.4%, 97.6%, and 97.6% confidence levels respectively.

The same procedure was applied for the remaining samples and the average accurate coverage probability were calculated of 300 samples derived from the each 2000 sample and result is presented in Table 6.

Table 6 : Average accurate confidence level based on the 95% confidence level for Weibull percentiles

Percentiles	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
Coverage probability	95.901	97.501	97.603	97.680	97.910

Based on the above real confidence levels, the calculated confidence bands of percentiles of week 24 are as follows.

Table 7: Confidence bands of percentiles of week 24 (nominal and actual)

Percentile	Value	Confidence limits at level of 95% (nominal)		Accurate confidence levels	Confidence limits (Actual)	
P_{50}	22.4	15.2	32.9	95.901	14.9	33.5
P_{60}	30.7	21.5	43.8	97.501	20.5	46.1
P_{70}	41.9	30.0	58.6	97.603	28.5	61.6
P_{80}	58.4	42.0	81.2	97.680	39.8	85.5
P_{90}	87.8	62.0	124.3	97.910	58.3	132.2

Based on the result formed from the simulation, there is a considerable difference between nominal and calculated coverage probabilities. Weibull distribution, drastically tends to be skewed to the right when the shape parameter less than one. Thus, the distribution of weekly rainfall deviates from the normal distribution with respect to the lower (less than one) value of shape parameter of the distribution. The deviation of the normality of the fitted distribution with the small size of sample is might be one of the reason for the discrepancy of the nominal and calculated coverage probabilities.

4. DISCUSSION AND CONCLUSION

Accurate estimates are important to draw reliable decisions which would be helped to minimize the issues that are caused due to heavy rainfall events and to utilize the water resources of the country. Weekly rainfall percentiles with 95% confidence bands are utilized in this study to compute the real coverage probability of the 95% confidence intervals. Rainfall total during the 24^{th} week (11-17 June) in SWM was considered as the main data series with size as 46. The data set were well fitted with the two parameter Weibull distribution and the maximum likelihood estimates for the scale and shape parameters of the fitted Weibull distribution were 33.9286 and 0.8775 respectively.

Based on the simulation carried out by bootstrapping approach, it is found that the most of the real coverage probabilities of 95% confidence intervals of percentiles (50^{th}, 60^{th}, 70^{th}, 80^{th} and 90^{th}) is less than 0.95. The accurate coverage probability of 95% confidence interval for the 50^{th} percentile is attained at the average level of 95.901%. The corresponding accurate coverage probabilities of 95% confidence intervals of 60^{th},70^{th}, 80^{th} and 90^{th} percentiles are given at the average levels of 97.501%, 97.603%, 97.680% and 97.910% respectively. Based on the above result, it can be concluded that the most of the coverage probabilities of the 95% confidence intervals for the rainfall percentiles get less value than the 0.95. This implies that the confidence interval of the percentile which derived from the skewed distribution as two parameter Weibull distribution at small sample size is not always give actual coverage probability. As a result of this, inferences make based on this facts do not provide much accurate estimates which require to get decisions at the high uncertainty. Thus, it is more suitable to consider much greater value for confidence level to get 0.95 coverage probability for percentiles of skewed distribution as two parameter Weibull distribution along with the small sample size.

ACKNOWLEDGEMENT

This study was partially funded by the University Research Grant, University of Sri Jayewardenepura, Sri Lanka under Grant (ASP/01/RE/HSS/2016/75).

REFERENCES

Dunn, P. K., 2001, Bootstrap confidence intervals for predicted rainfall quantiles. *Int.J. Climatol*, **21(1)**, 89-94.

Domroes, M., 1974, The Agroclimatic of Sri Lanka. Wiesbaden.

Heo, J.H., Salas, J.D. , Kim, K.D., 2001, Estimation of Confidence Intervals of Quantiles for the Weibull Distribution. *Stochastic Envirnmental Reserach and Risk Assesment*, **15**, 284-309.

Johnson, N.L., Kotz,S.,1970, Continous Univariate Distribution-1. houghton Mifflin Company, Boston.

Lo, K. F. A., Koralegedara, S. B., 2015, Effects of climate change on urban rainwater harvesting in Colombo city. *Sri Lanka. Journal of Environments*, **2(1),** 105-124.

Mayooran, T., Laheetharan, A., 2014, The Statistical Distribution of Annual Maximum Rainfall in Colombo District. *Sri Lankan Journal of Applied Statistics,* **15(2),** 107-130.

Mishra, P., Khare, D., Mondal, A., Kundu, S., Shukla, R., 2013, Satistical and probability analysis of rainfall for crop planning in a canal command. *Agriculture and Sustainable Development*, **1(1),** 95-102.

Sharda, V., Das, P., 2005, Modelling weekly rainfall data for crop planning in a sub-humid climate of India. *Agricultural water management*, **76 (2)**, 120-138.

Sharma, M. A., Singh, J. B., 2010, Use of probability distribution in rainfall analysis. *New York Science Journal,* **3 (9)**, 40-49.

Silva, H.P.T.N, Peiris, T.S.G., 2017, Analysis of Weekly Rainfall using Percentile Bootstap Approach. *International Journal of Ecology & Development*, **32(3)**, 97-100.

Silva, H.P.T.N, Peiris, T.S.G., 2017, Statistical Modeling of Weekly Rainfall: A case study in Colombo city in Sri Lanka, Proceedings of the 3rd International Multidisciplinary Engineering Research Conference; 2017 May 29-30; Moratuwa, Sri Lanka: IEEE.

Park, J. S., Jung, H. S., Kim, R. S., O.H, J. H., 2001, Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution. *International Journal of Climatology*, **21(11)**, 1371-1384.