Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2016, 3(1):74-81

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Evaluation of Parameter Estimation Methods of Probability Distributions for Modelling of Surface Temperature

N Vivekanandan

Central Water and Power Research Station, Pune, Maharashtra, India

Abstract: Studies on extreme temperature are beneficial to human understanding of extreme events. Decision-makers and researchers in climatology will benefit from knowledge about the behaviour of extreme temperature, as appropriate policies and plans can be drawn to prepare the general public for changes due to extreme temperature. This can be achieved through Extreme Value of Analysis (EVA) by fitting probability distributions to the recorded surface temperature data. This paper illustrates the adoption of Extreme Value Type-1 (EV1), Extreme Value Type-2 (EV2), Log Normal (LN2) and Log Pearson Type-3 (LP3) distributions in EVA of surface temperature for Hissar. Method of Moments (MoM) and Maximum Likelihood Method (MLM) are used for determination of parameters of EV1, EV2, LN2 and LP3 distributions. In addition to MoM and MLM, order statistics approach is used for determination of parameters of EV1 and EV2 distributions. Goodness-of-Fit tests viz., Anderson-Darling and Kolmogorov-Smirnov and diagnostic test using D-index are applied for evaluation of parameter estimation methods of probability distributions adopted in EVA. The paper presents the LN2 (MLM) is better suited probability distribution for modelling the annual maximum temperature data recorded at Hissar whereas LP3 (MLM) for annual minimum temperature.

Keywords: Anderson-Darling, D-index, Kolmogorov-Smirnov, Log Normal, Log Pearson, Temperature

Introduction

Extreme weather and climate events severely influence ecosystems and human society. High temperatures are among the most frequently investigated extreme events; the domains in which they affect society include agriculture, water resources, energy demand and human mortality [1]. Many research activities focus on extreme climate phenomena both because of their current impacts and the threat of their possible increases in frequency, duration and severity in a climate perturbed by enhanced concentrations of greenhouse gases in the atmosphere. Impacts of climate change would result from changes in variability and extreme event occurrence rather than from an increase in mean temperature [2] and even relatively small changes in the means and variations of climate variables can induce considerable changes in the severity of extreme events [3-4]. Impacts of extreme events are more serious when extreme weather conditions prevail over extended periods.

Out of a number of probability distributions that are adopted in frequency analysis, Extreme Value Type-1 (EV1), Extreme Value Type-2 (EV2), 2-parameter Log Normal (LN2) and Log Pearson Type-3 (LP3) are generally used for EVA of surface temperature [5-6]. Based on the theoretical applicability, parameter estimation procedures viz., Method of Moments (MoM) and Maximum Likelihood Method (MLM) are generally used for determination of parameters [7]. In addition to MoM and MLM, Order Statistics Approach (OSA) is used for determination of parameters of EV1 and EV2 distributions. Atomic Energy Regulatory Board (AERB) [8] guidelines described that the OSA estimates are having less bias and minimum variance though number of methods are available for determination of parameters of probability distributions. AERB guidelines suggested the 1000-year return period Mean+SE (where Mean denotes the estimated temperature (X_T) and SE the Standard Error) and Mean-SE values will be generally considered for arriving at a design value of maximum and minimum temperature respectively.

In the recent past, number of studies has been carried out by researchers adopting probability distributions for EVA of surface temperature. Hughes et al. [9] carried out statistical analysis using Generalized Extreme Value (GEV) distribution and time-series models for minimum and maximum temperatures in the Antarctic Peninsula. Hasan et al. [10] applied GEV distribution for modelling the annual maximum temperature recorded at twenty-two meteorological stations in Malaysia. Vivekanandan [11] applied EV1 and EV2 distributions for modelling of extreme rainfall, temperature and wind speed for Kanyakumari region. He found that the EV1 (using OSA) is better suited probability distribution for modelling the series of annual extreme rainfall and annual maximum (or) minimum temperature whereas EV2 (using OSA) for modelling the series of annual hourly maximum wind speed for Kanyakumari. Generally, when different probability distributions are used for EVA, a common problem that arises is how to determine which model fits best for a given set of data. This can be evaluated by quantitative assessment using Goodness-of-Fit (GoF) and diagnostic tests. GoF tests viz., Anderson-Darling (A^2) and Kolmogorov-Smirnov (KS) are applied for checking the adequacy of fitting probability distributions to the surface temperature data. A diagnostic test (using D-index) is applied for the selection of most suitable probability distribution for estimation of extreme temperature. This paper details the procedures involved in assessing the suitable probability distribution for estimation of extreme temperature though GoF and diagnostic tests with illustrative example.

Methodology

The objective of the study is to assess the adequacy of Probability Density Function (PDF) for EVA of surface temperature. In this context, various steps followed for data processing, validation and analysis include: (i) prepare the Annual Maximum and Minimum Temperature (AMAXT and AMINT) data series from the hourly temperature data; (ii) select the PDFs for EVA (say, EV1, EV2, LN2 and LP3); (ii) select parameter estimation methods (say, MoM, MLM and OSA) wherever applicable; (iii) select quantitative GoF and diagnostic tests and (iv) conduct EVA and analyse the results obtained thereof. The PDF and quantile estimator (X_T) of the distributions are presented in Table 1.

Table 1: PDF and quantile estimator of probability distributions

Distribution	PDF	Quantile estimator
EV1	$f(X;\alpha,\beta)=\dfrac{e^{-(X-\alpha)/\beta}e^{-e^{-(X-\alpha)/\beta}}}{\beta}$, $\beta>0$	$X_T=\alpha+Y_T\beta$
EV2	$f(X;\beta,\gamma)=\dfrac{\gamma}{\beta}\left(\dfrac{\beta}{X}\right)^{\gamma+1}e^{-\left(\frac{X}{\beta}\right)^{-\gamma}}$, $\beta>0$	$X_T=\beta e^{(Y_T/\gamma)}$
LN2	$f(X;\alpha,\beta)=\dfrac{1}{\beta X\sqrt{2\pi}}\exp\left(-\dfrac{(\ln(X)-\alpha)^2}{2\beta^2}\right)$, $-\infty<X<\infty,\beta>0$	$X_T=e^{\alpha+K_P\beta}$
LP3	$f(X;\alpha,\beta,\gamma)=\dfrac{1}{\beta X\Gamma\gamma}\left(\dfrac{\ln(X)-\alpha}{\beta}\right)^{\gamma-1}e^{-\left(\frac{\ln(X)-\alpha}{\beta}\right)}$, $\beta,\gamma>0$	$X_T=Exp((\alpha+\beta\gamma)+K_P\beta\sqrt{\alpha})$

In Table 1, α, β and γ denotes the location, scale and shape parameters of the distributions respectively. For EV1 and EV2 distributions, the reduced variate (Y_T) for a given return period (T) is defined by $Y_T=-\ln(-\ln(1-(1/T)))$ while in the mathematical representation of LN2 and LP3, K_P denotes the frequency factor corresponding to the probability of exceedance. The Coefficient of Skewness (C_S) is C_S=0.0 for LN2 whereas C_S is based on the log transformed series of the recorded data for LP3 [12]. For the data series with AMINT, the value of Y_T will be read as $Y_T = -\ln(-\ln(1/T))$ and K_P the frequency factor corresponding to the probability of non-exceedance for LN2 and LP3 distributions.

Goodness-of-Fit Tests

Generally, A^2 test is applied for checking the adequacy of fitting of EV1 and EV2 distributions. The procedures involved in application of A^2 test for LN2 and LP3 are more complex though the utility of the test statistic is extended for checking the quantitative assessment. In view of the above, KS test is widely applied for the purpose of quantitative assessment. Theoretical descriptions of GoF tests are as follows:

A^2 test statistic is defined by:

$$A^2 = (-N) - (1/N)\sum_{i=1}^{N}\{(2i-1)\ln(Z_i) + (2N+1-2i)\ln(1-Z_i)\} \qquad \ldots (1)$$

Here, $Z_i = F(X_i)$ for i=1,2,3,…,N with $X_1 < X_2 < \ldots < X_N$, $F(X_i)$ is the Cumulative Distribution Function (CDF) of ith sample (X_i) and N is the sample size.

KS test statistic is defined by:

$$KS = \underset{i=1}{\overset{N}{Max}}\,(F_e(X_i) - F_D(X_i)) \qquad \ldots (2)$$

Here, $F_e(X_i)$ is the empirical CDF of X_i and $F_D(X_i)$ is the derived CDF of X_i by PDFs. In this paper, Weibull plotting position formula is used for computation of empirical CDF [13]. The theoretical values of A^2 and KS statistic for different sample size (N) at 5% significance level are available in the technical note on "Goodness-of-Fit Tests for Statistical Distributions" by Charles Annis [14].

Test criteria: If the computed value of GoF tests statistic given by the distribution is less than that of theoretical value at the desired significance level then the distribution is found to be suitable for EVA of surface temperature at that level.

Diagnostic Test

Sometimes the GoF test results would not offer a conclusive inference thus posing a problem for the user in selecting a suitable PDF for their application. In such cases, a diagnostic test in adoption to GoF is applied for making inference. The selection of most suitable probability distribution for estimation of extreme temperature is performed through D-index test (USWRC, 1981), which is defined as below:

$$\text{D-index} = (1/\overline{X})\sum_{i=1}^{6}|X_i - X_i^*| \qquad \ldots (3)$$

Here, $\overline{X}$ is the average value of the recorded data whereas X_i (i= 1 to 6) and X_i^* are the six highest recorded and corresponding estimated values by different PDFs. The distribution having the least D-index is considered as better suited distribution for estimation of temperature [15].

Application

EVA of surface temperature is carried out to estimate extreme (maximum or minimum) temperature for different return periods adopting four different PDFs viz., EV1, EV2, LN2 and LP3. MoM, MLM and OSA are used for determination of parameters of EV1 and EV2 distributions whereas MoM and MLM for LN2 and LP3 distributions. Hourly surface temperature data (with missing values) for the period 1970 to 2007 is used. The series of AMAXT and AMINT is extracted from the hourly temperature data and used for EVA. From the scrutiny of the hourly temperature data, it is observed that the data for the period of six years (1978, 1979, 1981, 1983, 1987 and 1989) are missing. So, for the series of AMAXT, the missing data for the years are imputed by the series maximum value i.e., 48.4 ^{o}C. Likewise, for the series of AMINT, the missing data for the years are imputed by the series minimum value i.e., 2.7 ^{o}C. After replacing the missing values as per AERB guidelines, the entire data series is used for EVA. Table 2 gives the descriptive statistics such as average, standard deviation, coefficient of skewness and coefficient of kurtosis of the data series of AMAXT and AMINT for Hissar.

Table 2: Descriptive statistics of AMAXT and AMINT for Hissar

Data series	Statistical parameters (SD: Standard Deviation)			
	Average (°C)	SD (°C)	Skewness	Kurtosis
AMAXT	45.7	1.2	0.708	-0.503
AMINT	5.3	1.1	-0.465	-0.511

Results and Discussions

Based on the parameter estimation procedures of EV1, EV2, LN2 and LP3 distributions, a computer code was developed in FORTRAN language and used for EVA. The estimated maximum temperature (X_T) with SE computed from four probability distributions is presented in Table 3 and 4. Similarly, the estimated minimum temperature (X_T) with SE for different return periods is presented in Table 5. Since there was no existence of OSA for LN2 and LP3 distributions, the EVA results of these two distributions are not presented in Tables 4 and 5. Similarly, from the EVA of AMINT data, it was observed that the EV2 distribution is not found to be feasible for fitting and therefore EVA results of EV2 distribution is not presented in Table 5.

Table 3: Estimated maximum temperature adopting EV1 and EV2 distributions

Return period (year)	Maximum temperature (°C)											
	EV1						EV2					
	MoM		MLM		OSA		MoM		MLM		OSA	
	X_T	SE	X_T	SE	X_T	SE	X_T	SE	X_T	SE	X_T	SE
2	46.0	0.2	46.0	0.2	46.0	0.3	46.5	0.2	45.9	0.3	46.0	0.3
5	47.3	0.5	47.4	0.4	47.5	0.5	47.7	0.5	47.3	0.4	47.5	0.5
10	48.2	0.6	48.3	0.6	48.5	0.6	48.5	0.7	48.2	0.6	48.5	0.7
20	49.1	0.7	49.2	0.7	49.4	0.8	49.4	0.8	49.1	0.8	49.5	0.8
50	50.2	0.9	50.3	1.0	50.7	0.9	50.5	1.0	50.3	1.0	50.9	1.0
100	51.0	1.1	51.2	1.1	51.6	1.1	51.3	1.2	51.2	1.2	51.9	1.2
200	51.9	1.2	52.1	1.2	52.5	1.2	52.1	1.4	52.1	1.4	53.0	1.4
500	53.0	1.4	53.2	1.5	53.7	1.4	53.3	1.6	53.4	1.6	54.4	1.7
1000	53.8	1.6	54.1	1.6	54.6	1.6	54.1	1.9	54.3	1.9	55.5	1.9

Table 4: Estimated maximum temperature adopting LN2 and LP3 distributions

Return period (year)	Maximum temperature (°C)							
	LN2				LP3			
	MoM		MLM		MoM		MLM	
	X_T	SE	X_T	SE	X_T	SE	X_T	SE
2	46.2	0.3	46.2	0.3	46.1	1.6	46.1	1.5
5	47.5	0.4	47.5	0.3	47.5	1.9	47.4	2.0
10	48.2	0.4	48.2	0.4	48.3	2.3	48.2	2.3
20	48.8	0.4	48.8	0.4	48.9	2.7	48.9	2.6
50	49.5	0.5	49.4	0.5	49.7	3.2	49.7	3.1
100	49.9	0.6	49.8	0.6	50.2	3.6	50.2	3.5
200	50.3	0.7	50.2	0.7	50.7	3.9	50.8	3.8
500	50.8	0.7	50.7	0.7	51.4	4.3	51.4	4.3
1000	51.2	0.7	51.1	0.7	51.8	4.7	51.9	4.7

Table 5: Estimated minimum temperature adopting EV1, LN2 and LP3 distributions

Return period (year)	Minimum temperature (°C)													
	EV1						LN2				LP3			
	MoM		MLM		OSA		MoM		MLM		MoM		MLM	
	X_T	SE	X_T	SE	X_T	SE	X_T	SE	X_T	SE	X_T	SE	X_T	SE
2	3.7	0.2	5.0	0.2	5.0	0.2	4.6	0.2	4.6	0.3	4.4	0.4	4.4	0.4
5	2.5	0.4	3.7	0.4	3.7	0.4	3.6	0.2	3.5	0.3	3.4	0.3	3.4	0.4
10	1.6	0.5	2.8	0.6	2.9	0.5	3.1	0.2	3.0	0.2	3.1	0.4	3.0	0.4
20	0.8	0.7	1.9	0.7	2.1	0.7	2.8	0.2	2.7	0.3	2.8	0.4	2.7	0.4
50	-0.2	0.9	0.8	0.9	1.1	0.9	2.5	0.2	2.3	0.2	2.6	0.4	2.5	0.5
100	-1.0	1.1	0.0	1.1	0.3	1.0	2.3	0.2	2.1	0.2	2.4	0.4	2.3	0.4
200	-1.8	1.2	-0.9	1.2	-0.5	1.1	2.2	0.3	2.0	0.3	2.3	0.4	2.2	0.4
500	-2.9	1.3	-2.0	1.4	-1.5	1.3	2.0	0.3	1.8	0.3	2.2	0.4	2.1	0.5
1000	-3.6	1.5	-2.8	1.6	-2.3	1.5	1.8	0.2	1.7	0.3	2.1	0.4	2.0	0.5

Analysis Based on GoF Tests

GoF tests viz., A^2 and KS were applied for quantitative assessment of fitting of EV1, EV2, LN2 and LP3 distributions by adopting different parameter estimation methods to the data series of surface temperature. The GoF tests results were computed from Eqs. (1) and (2), and given in Table 6.

Table 6: Computed and theoretical values of GoF tests

Distri-bution	A^2 (Theoretical value: 0.757)						KS (Theoretical value: 0.238)					
	AMAXT			AMINT			AMAXT			AMINT		
	MoM	MLM	OSA	MoM	MLM	OSA	MoM	MLM	OSA	MoM	MLM	OSA
EV1	1.213	1.064	0.921	2.74	2.698	3.365	0.123	0.118	0.104	0.188	0.518	0.482
EV2	3.105	1.181	0.894	NF	NF	NF	0.011	0.128	0.101	NF	NF	NF
LN2	1.154	1.372	-	1.352	1.453	-	0.165	0.182	-	0.162	0.171	-
LP3	1.622	1.199	-	1.415	1.573	-	0.162	0.174	-	0.122	0.136	-

From Table 6, it may be noted that the computed values of KS test results obtained from four probability distributions with different parameter estimation methods are not greater than the theoretical values at 5% level of significance, and at this level, all four distributions are found to be acceptable for modelling the series of AMAXT and AMINT recorded at Hissar whereas A^2 test results did not support the use of these distributions for modelling the data series of annual maximum and minimum temperature.

Analysis Based on Diagnostic Test

For the selection of most suitable probability distribution for estimation of extreme temperature, the diagnostic index, say D-index values of four probability distributions with different parameter estimation methods were computed from Eq. (3) and given in Table 7.

Table 7: D-index values of four probability distributions

Distribution	AMAXT			AMINT		
	MoM	MLM	OSA	MoM	MLM	OSA
EV1	0.099	0.103	0.110	1.715	0.962	0.893
EV2	0.119	0.104	0.117	NF	NF	NF
LN2	0.069	0.068	-	0.562	0.462	-
LP3	0.076	0.078	-	0.447	0.403	-

From Table 7, it may be noted that the D-index values computed from LN2 (MLM) for AMAXT and LP3 (MLM) for AMINT are found to be minimum when compared to the corresponding values of other probability distributions. The plots of estimated extreme (maximum or minimum) temperature for different return periods with recorded data for Hissar were presented in Figures 1 and 2.

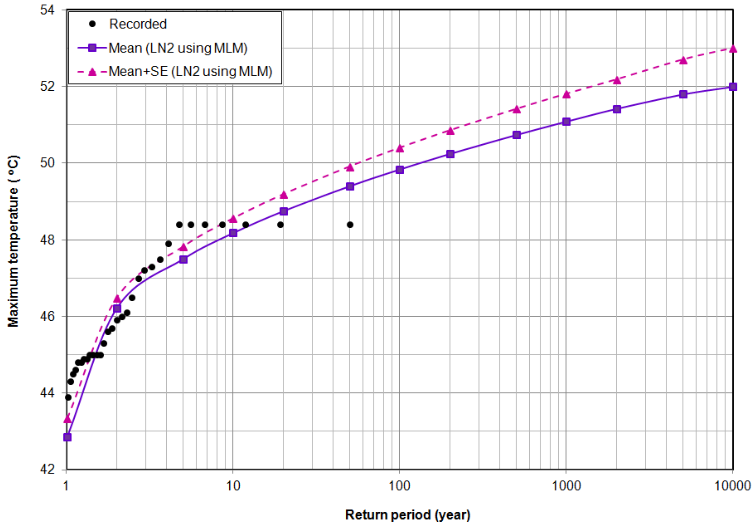

Figure 1: Plots of estimated Mean and Mean+SE values of annual maximum temperature using LN2 (MLM) distribution with recorded data for Hissar

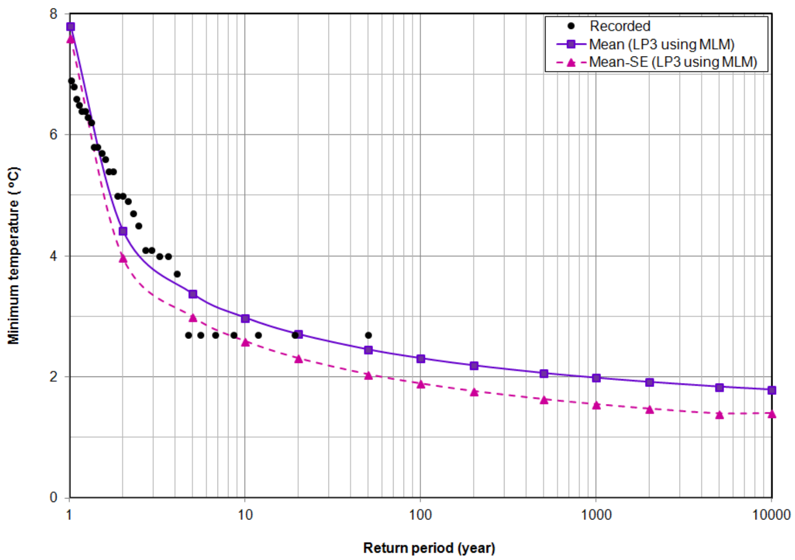

Figure 2: Plots of estimated Mean and Mean-SE values of annual minimum temperature using LP3 (MLM) distribution with recorded data for Hissar

Conclusions

The paper presented the study on EVA of surface temperature for Hissar adopting EV1, EV2, LN2 and LP3 distributions with applicable parameter estimation methods. From EVA results, the following conclusions were drawn from the study:

i) The estimated maximum temperature using EV1 and EV2 distributions were consistently higher than the corresponding values of LN2 and LP3 distributions in upper tail region when MoM and MLM is adopted for determination of parameters of distributions adopted in EVA of surface temperature.

ii) For LN2 and LP3 distributions, it was found that there is no appreciable difference between the estimated maximum temperature using MoM and MLM.
iii) Suitability of probability distribution was evaluated by GoF (using A^2 and KS) and diagnostic (using D-index) tests. The KS test results confirmed the applicability of EV1, EV2, LN2 and LP3 distributions for EVA of surface temperature though A^2 test rejects the use of these four distributions for EVA.
iv) For the series of annual maximum temperature, the D-index value of LN2 (MLM) distribution was found to be a minimum when compared to the corresponding values of EV1, EV2 and LP3 distributions.
v) The D-index value of LP3 (MLM) distribution was found to be a minimum when compared to the corresponding values of EV1, EV2 and LN2 distributions for the series of annual minimum temperature.
vi) On the basis of GoF and diagnostic test results, the study identified the LN2 (MLM) distribution is better suited amongst four distributions studied for estimation of maximum temperature at Hissar whereas LP3 (MLM) for estimation of minimum temperature.
vii) The study suggested that the 1000-year return period Mean+SE value of maximum temperature of 51.8^{o} C using LN2 (MLM) and Mean-SE value of minimum temperature of 1.5^{o} C using LP3 (MLM) could be used for design purposes while designing the hydraulic structures in Hissar.

Acknowledgments

The author is grateful to the Director, Central Water and Power Research Station, Pune, for providing the research facilities to carry out the study. The author is thankful to M/s Nuclear Power Corporation of India Limited, Mumbai for supply of surface temperature data of Hissar.

References

[1]. Watson, R.T., Zinyowera, M.C., Moss, R.H. & Dokken, D.J. (1996). Impacts, Adaptations, and Mitigation of Climate Change: Scientific-Technical Analyse. Published for Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 878 pp.
[2]. Parmesan, C., Root, T.L. & Willig, M.R. (2000). Impacts of extreme weather and climate on terrestrial biota. Bulletin of the American Meteorological Society, 81(3):443−450.
[3]. Colombo, A.F., Etkin, D. & Karney, B.W. (1999). Climate variability and the frequency of extreme temperature events for nine sites across Canada: Implications for power usage. Journal of Climate, 12(8):2490−2502.
[4]. Huth, R., Kyselý, J. & Pokorná, L. (2000). A GCM simulation of heat waves, dry spells, and their relationships to circulation. Climate Change, 46(1):29−60.
[5]. Cooley, D. S. (2005). Statistical analysis of extremes motivated by weather and climate studies: applied and theoretical advances, PhD thesis, University of Colorado.
[6]. Hasan, H. and Yeong, W.C. (2014). Extreme value modelling and prediction of extreme rainfall: A case study of Penang. *AIP Conference Proceedings.* V1309, pp. 372-393.
[7]. Bobee, B. and Askhar, F. (1991). The Gamma family and derived distributions applied in hydrology, Water Resources Publications.
[8]. AERB (2008). Extreme values of meteorological parameters, Atomic Energy Regulatory Board (AERB) Safety Guide No. AERB/ NF/ SG/ S-3.
[9]. Hughes, G.L., Rao, S.S. & Rao, T.S. (2007). Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula. Proceedings of the Royal Society, Series A, 463: 241–259
[10]. Hasan, H., Salam, N. and Adam, M.B. (2013). Modelling extreme temperature in Malaysia using generalized extreme value distribution. International Journal of Mathematical, Computational, Natural and Physical Engineering, 7(6):618-624.

[11]. Vivekanandan, N. (2015). Modelling of annual extreme rainfall, temperature and wind speed Using OSA of EV1 and EV2 distributions. International Journal of Innovative Research in Computer Science & Technology, 3(4):57-60.
[12]. Rao, A.R. and Hameed, K.H. (2000). Flood Frequency Analysis, CRC Publications, Washington, New York.
[13]. Cook, N. (2011). Comments on plotting positions in extreme value analysis. Journal of Applied Meteorology and Climatology, 50 (1):255-266.
[14]. Charles Annis, P.E. (2009). Goodness-of-Fit tests for statistical distributions, [http://www.statistical engineering. com/goodness.html]
[15]. USWRC (1981). Guidelines for determining flood flow frequency, United States Water Resources Council (USWRC) Bulletin No. 17B.