IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 6 Ver. V (Nov. - Dec.2016), PP 08-12
www.iosrjournals.org

Modeling with Probability Distribution of Extreme Water Levels of the Brahmaputra in Dibrugarh, Assam

Pallab Changkakoti[1], Jiten Hazarika[2]
[1]*Department of Statistics, Kohima Science College, Jotsoma, Nagaland, India*
[2] *Department of Statistics, Dibrugarh University, Dibrugarh, Assam, India*

Abstract : *Daily water level data of the Brahmaputra of Dibrugarh gauge station in Assam has been obtained from Central Water Commission (CWC), Govt. of India. This study is aimed to explore the statistical probability distribution of the extreme water levels of the Brahmaputra in Dibrugarh, Assam. Descriptive statistics and six probability distributions, viz., Generalized Extreme Value, Generalized Logistic, Gumbel Max, Logistic, Log Logistic and Lognormal have been proposed and ranked using Akaike Information Criterion (AIC). Parameters of the distribution are estimated using the Maximum Likelihood Estimation (MLE) method.*
Keywords: *Extreme water level, probability distribution, AIC*

I. Introduction

Floods are one of the most damaging and dangerous natural hazards. There are a lot of different impacts that flood can bring: (1) Landslides or instability of earth - this is commonly caused when there is an oversaturation of liquid in the soil. This can be very dangerous, and almost always happen without warning. (2) Destruction of Crops - when they are flooded out or pushed out of the land from the roots being oversaturated with water. (3) chemicals and other industrial wastes being put into waterways, and into irrigation fields used for plantation, cultivation etc. In India 60% of the total damage due to floods is in the states of Assam, Bihar, Uttar Pradesh and West Bengal (*Central Statistical Organization)*[1]. With about 40 million hectares flood-prone areas coupled with the average annual damage to the extent of Rs 500 crores and affecting 10 per cent of total population, floods have been the most severe natural disaster in India. The flood-affected areas have increased 3 times since 1960. The average total damage incurred presently is about six times that incurred in 1950s. Ever-increasing catastrophic flood profiles in the country call for utmost attention at all levels.

Brahmaputra, the biggest river in the Indian subcontinent, is uniquely characterized by its narrow sized valley coupled with the steep slopes and transverse gradient along with catchments experiencing the highest rainfall and loading heaviest sediments, which are driven mainly due to deforestation and unsustainable agricultural practices. The recurring floods leave a trail of deaths, destruction and damages to existing infrastructure, including roads, bridges, embankments, irrigation canals, buildings and crops. Due to information gap during critical times and absence of precise damage assessment, the State and Central mechanisms are unable to address the flood management-related issues more judiciously and efficiently.

Extensive studies have been made in the fitting and modelling of hydrological data not only in the western countries but in India also. Normal, Lognormal, Gumbel and Pearson Type III distributions were fitted to flood flow at Garudeshwar station in Narmada river, Gujarat, India by Khan [2]. Izinyon *et. al.* [3] used Log Pearson Type III distribution for flood frequency analysis of Ipkoba river of Nigeria. Nooijen *et. al.*[4] compared fitting of several distributions for hydrological data. Abida *et. al.*[5] tested five most frequently used distributions in the analysis of hydrologic extreme variables in Tunisia using L-Moment method. Viz.,(i) Generalized Extreme Value (GEV), (ii) Pearson Type III (P3), (iii) Generalized Logistic (GLO), (iv) Generalized Normal (GN), and (v) Generalized Pareto (GPA) distributions. Kashani *et. al.*[6] estimated flood for Urmia lake basin in Iran using a regression model, an artificial neural network (ANN) model and a hybrid model. Abida *et. al.*[7] found that Generalised Normal and Generalised Extreme Value distributions are the best fitted distributions for hydrologic extreme variables using L-Moment method. Kale *et. al.*[8] evaluated the temporal variations in specific stream power and the total energy available for geomorphic work during the monsoon season for the Tapi river in central India. Shabri *et. al.* [9] recommended Generalised Logistic distribution as a standard for flood frequency analysis and tested for annual maximum flood series for the River Kelvin and River Spey. Kale [10] observed that the annual peak-discharge data indicate significant decreasing trends for the upper Ganga, Krishna and Brahmaputra rivers, and significant increasing trends for Narmada, Godavari and Satluj rivers using time-series analyses. Kroll *et. al.* [11] used the method of L-Moment and recommended Pearson Type III and Lognormal (3P) for describing low stremflow statistics in United States. Dhar *et. al.*[12] analysed flood data of the main Brahmaputra River and its 12 major tributaries to discover which sites and tributaries experience frequent floods in different monsoon months and how high flood magnitudes are.

In this study an attempt has been made to find an appropriate probability distribution to fit maximum water level data of gauge station of the Brahmaputra river in Assam, India and the findings of the same along with the methodology adopted are presented in this paper.

II. Materials and Method

2.1 Study Area:

The Brahmaputra is one of the largest river in the world. From its origin in a glacier east of Manasarovar in south western Tibet as the Yarlung Zangbo river, it flows across southern Tibet to break through the Himalayas in great gorges and into Arunachal Pradesh where it is known as Dihang. It flows southwest through the Assam Valley as Brahmaputra and south through Bangladesh as the Jamuna. In Bangladesh the river merges with the Ganges and splits into two: the Padma and Meghna River. When it merges with the Ganges it forms the world's largest delta, the Sunderbans and fall at Bay of Bengal. The total length of the river is 2880 km, out of which 918 km is in India. In Assam, the length is 640 km and its average width is just 8 km. Although two-third of its length and more than half of the catchments are in Tibet, about two-third contribution of its flow is from the drainage area in India (Brahmaputra Board) [13]. Brahmaputra river valley in Assam from Kobo to Dhubri is joined by about 20 important tributaries on its north bank and about 13 on its south bank. The water level gauge station at Dibrugarh of the Brahmaputra river has been selected for this study.

2.2 Data

Daily maximum and minimum water level data from 15th May to 15th October for ten years i.e. from 2001 to 2010 in meter (m) have been obtained from Central Water Commission (CWC), Govt. of India for Dibrugarh water level gauge station of the Brahmaputra in Assam.

2.3 Methodology:

Descriptive statistics such as maximum and minimum, mean, standard deviation, and coefficient of variation (COV) of water level data, and six continuous probability distribution viz. Generalized Extreme Value, Generalized Logistic, Gumbel Max, Logistic, Log Logistic, and Lognormal distributions are proposed for modelling the water level data of the Brahmaputra at Dibrugarh. Parameters of the distributions are estimated using Maximum Likelihood Estimation (MLE). If the continuous random variable X represents the water level then the probability density functions of different distributions are given as follows:

2.3.1 Generalized Extreme Value distribution:

$$f(x)=\begin{cases}\dfrac{1}{\sigma}\exp\left\{-(1+kz)^{-1/k}\right\}\left((1+kz)^{-1-1/k}\right); & 1+k\dfrac{(x-\mu)}{\sigma}>0 \quad for\ k\neq 0\\ \dfrac{1}{\sigma}\exp\left\{-z-\exp(-z)\right\} & ;\quad -\infty<x<\infty \qquad for\ k=0\end{cases}$$

$$Where\ \ z=\frac{x-\mu}{\sigma}\ ;\quad k: shape\ parameter,\ \sigma: scale\ parameter,\ \ \mu: location\ parameter$$

2.3.2 Generalized Logistic distribution:

$$f(x)=\frac{(1+kz)^{1-1/k}}{\sigma\left(1+(1+kz)^{-1/k}\right)^2}\qquad 1+k\frac{(x-\mu)}{\sigma}>0\qquad for\quad k\neq 0;$$

$$f(x)=\frac{\exp(-z)}{\sigma\left(1+\exp(-z)\right)^2}\qquad -\infty<x<\infty\qquad for\quad k=0$$

$$Where\ z=\frac{x-\mu}{\sigma};\qquad k=shape\ parameter\ \mu: location\ parameter, \sigma: scale\ parameter$$

2.3.3 Gumbel Max (Maximum Extreme Value Type-I) distribution:

$$f(x)=\frac{1}{\sigma}\exp\left\{-z-\exp(-z)\right\}\ ;\quad -\infty\leq x<+\infty\ ;\quad Where\ \ z=\frac{x-\mu}{\sigma}$$

2.3.4 Logistic distribution:

$$f(x) = \frac{exp(-z)}{\sigma\left(1+exp(-z)\right)^2} \qquad -\infty < x < \infty$$

$$Where\ z = \frac{x-\mu}{\sigma}; \qquad \mu : location\ parameter, \sigma : scale\ parameter$$

2.3.5 Log Logistic distribution:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{-2} \qquad 0 \le x < \infty$$

$$Where\ \ \alpha : shape\ parameter, \beta : scale\ parameter$$

2.3.6 Lognormal distribution:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right\}; \quad -\infty < x < +\infty$$

$$Where\ \ \mu : continuous\ location\ parameter; \quad \sigma : continuous\ scale\ parameter$$

These distributions are selected for the analysis due to the fact that they may be fitted to water level data as mentioned above.

2.3.1 Comparing Fitted Models: Akaike Information Criterion (AIC):

The Akaike information criterion (AIC) (Akaike, [14]) is a popular method for comparing the adequacy of multiple, possibly non nested models.

AIC= - 2Loglikelihood+2*k*, Where *k* is the number of parameters.

When *k* is large compared to *n* (sample size) second order bias corrected version of AIC called AIC_C is used.

$$AIC_C = -2Loglikelihood + 2k + \frac{2k(k+1)}{n-k-1} \qquad \text{(Sugiura et. al. [15])}$$

AIC_C should be used unless n/k > about 40. AIC_C converges to AIC as n gets large. In practice AIC_C should be used (Burnham et. al., [16]). Even in moderate sample sizes , AIC_C provides substantially better model selections than AIC (Hurvich et. al., [17]).

The above statistics have been used in our present study to compare adequacy of the proposed models.

2.3.2 Quantile-Quantile (Q-Q) Plot:

The basic idea of the Q-Q plot is to compute the theoretically expected value for each data point based on the distribution in question. If the data indeed follow the assumed distribution, then the points on the Q-Q plot will fall approximately on a straight line. A Q-Q plot is a plot of the quantiles of the first data set against the quantiles of the second data set.

III. Results And Discussion

A summary of daily gauge reading for the Dibrugarh gauge station of the Brahmaputra is provided in Table – 3.1. Very small standard deviation and coefficient of variation indicate that the variation in the water level is very less for the study period.

Table – 3.1 Descriptive Statistics Values

Minimum	Maximum	Mean	Standard Deviation	COV (100%)
102.61	106.26	104.65	0.5837	0.558

Table – 3.2 shows the estimated parameters, negative of the Loglikelihood functions, Akaike Information Criteria (AIC), second order bias corrected version of AIC viz. AIC_C, Δ AIC_C and Akaike Weights. According to the values of the AIC_C's, the six fitted distributions are ranked for Dibrugarh gauge station of the Brahmaputra.

Table – 3.2 Fitting of Probability Distributions

Rank	Distribution	Parameters	Negative Log Likelihood	AIC	AIC_C	$\Delta\ AIC_C = AIC_{Ci}\text{-Min}(AIC_C)$
1	Generalized Logistic	k=-0.10318 σ=0.31943 μ=104.7	1322.86	2651.72	2651.74	0.00
2	Generalized Extreme Value	k=-0.47848 σ=0.62218 μ=104.5	1349.46	2704.92	2704.94	53.20
3	Logistic	σ=0.32181 μ=104.65	1356.09	2716.18	2716.19	64.45
4	Log-Logistic	α=318.84 β=104.65	1357.57	2719.14	2719.15	67.41
5	Lognormal	σ=0.00558 μ=4.6506	1358.75	2721.50	2721.51	69.77
6	Gumbel	σ=0.45511 μ=104.39	1590.55	3185.10	3185.11	533.37

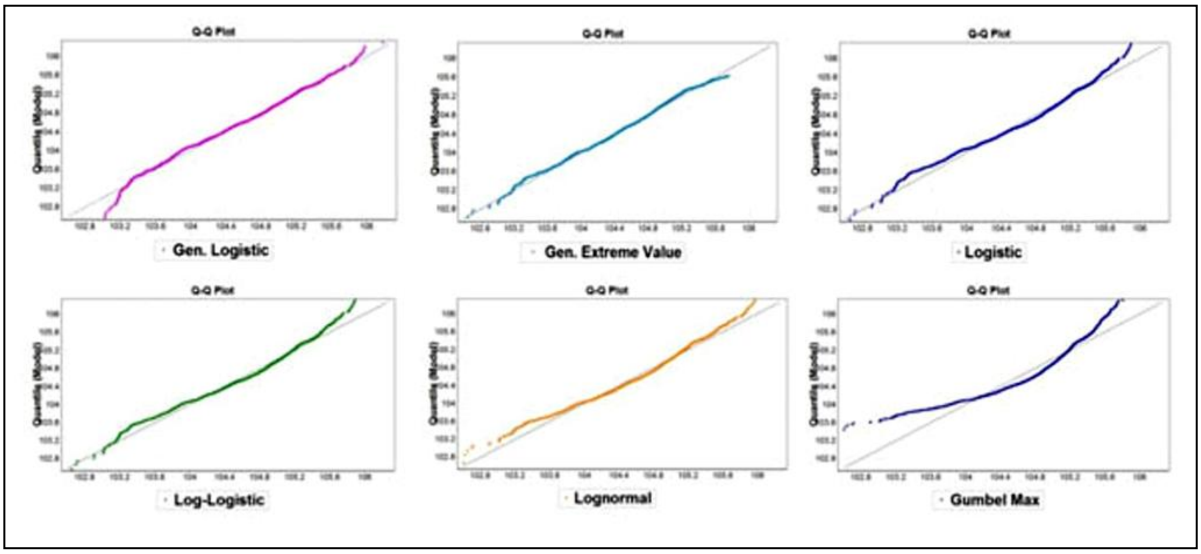

Fig. – 2.3.2: Q-Q Plot for Various Distributions

From the Q-Q Plot of the fitted distributions it is seen that all of the plots are arced or "S" shaped, indicating that either the observed or the fitted distribution is more skewed than the other.

IV. Conclusion

In this paper an attempt has been made to analyse daily maximum water level data of Dibrugarh gauge stations of the Brahmaputra situated in Assam in India, using probability models. In order to meet the stated objectives, based on empirical works of such types of study at national and international level, six probability models have been proposed. The study signifies the supremacy of Generalized Logistic distribution among the six probability distributions tested on the water level data of the bBrahmaputra at Dibrugarh gauge station in Assam, India.

References

[1] *Compendium of Environment Statistics, Central Statistical Organization, Department of Statistics, Ministry of Planning and Programme Implementation, Govt. of India Publication, New Delhi, 1997, pp. 115–120*

[2] Khan, M., (2013): Frequency Analysis of Flood Flow at Garudeshwar Station in Narmada River, Gujarat, India. *Universal Journal of Environmental Research and Technology, eISSN : 2249 0256 , Volume 3, Issue 6, 2013, pp: 677-684*

[3] Izinyon, O.C., Ihimekpen, N. and Igbinoba, G.F. (2011): Flood Frequency Analysis of Ikpoba River Catchment at Benin City Using Log Pearson Type III Distribution. *Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS) 2(1). Scholarlink Research Institute Journals, 2011 (ISSN: 2141-7016) pp :50-55*

[4] Nooijen, R.R.P.V., Gubareva, T. and Kolechkina, A.G. (2010): A Comparison of Fitting Methods and Tests for Several Distributions on Hydrological Data. *International Workshop: ADVANCES IN STATISTICAL HYDROLOGY. May 23-25, 2010 Taormina, Italy*.

[5] Abida, H. and Ellouze, M. (2008): Probability Distribution of Flood Flow in Tunisia.*Hydrology and Earth System Sciences, 12,2008. pp: 703-71*.

[6] Kashani, M.H., Montaseri, M. and Yaghin, M.A.L. (2008): Flood Estimation at Unguaged Sites Using a New Hybrid Model. *Journal of Applied Sciences 8(9) ISSN-1812-5654. pp: 1744-1749*

[7] Abida, H. and Ellouze, M. (2007). Probability distribution of flood flows in Tunisia. Hydrology and Earth System Sciences Discussion, 4, 957–981, 2007

[8] Kale, V.S. and Hire, P.S. (2007): Temporal Variations in the Specific Stream Power and Total Energy Expenditure of a Mosoonal River. *The Tapi River, India: Geomorphology, Vol. 92. pp:134-146.*

[9] Shabri, A. and Jemain, A.A. (2011): Fitting the Generalized Logistic Distribution by LQ-Moments. *Applied Mathematical Sciences, Vol. 5, 2011, No. 54, pp: 2663 - 2676*

[10] Kale, V.S. (2004): Research on Floods in India.

[11] Kroll, C.N., Vogel, R.M. (2002): Probability Distribution of Low Streamflow Series in the United States. *Journal of Hydrologic Engineering*. *March/April 2002. pp: 137-146*.

[12] DHAR, O.N. and NANDARGI, S. (2000): A Study of Floods in the Brahmaputra Basin in India. *International Journal of Climatology, 20, 2000. pp: 771-781*

[13] Master Plan for Brahmaputra Basin, Summary, Recommendation,Programme and Outlays, Brahmaputra Board, Ministry of Water Govt of India, 1986.

[14] Akaike, H., Petrov, B.N. and Csaki, F., (1973) : Information theory and an extension of the maximum likelihood principle". *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, Budapest: Akadémiai Kiadó, pp. 267–281*.

[15] Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. Comm. Statist. A 7, 13-26.

[16] Burnham, K. P., & Anderson, D. R. (2002) : Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. *Springer-Verlag : New York*.

[17] Hurvich, C. and Tsai, C. (1991) : Bias of the Corrected AIC Criterion for Under Fitted Regression and Time Series Models. *Biometrika ,78, 3, pp : 499-509*