International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934

Fitting hourly rainfall using various distribution and evaluation of best fit for mumbai region

Khan Mujiburrehman

Environmental Engineer, Dar al Handasah, Makkah, Saudi Arabia

Abstract

Mumbai city having an area of 437 sq km with a population of 12 million came to a complete halt due to the unprecedented rainfall of 944 mm during the 24 hours starting on 26th July 2005; with 380 mm falling in just 3 hours between 14:30 to 17:30 and hourly rainfall exceeding 190 mm/hr. The severity of flood was such that it took 447 lives in Mumbai alone. The financial cost of flood was unprecedented and affected the entire commercial, trading, and industrial activity for 3 days.

This paper presents a statistical study on fitting distribution of maximum hourly rainfall amount in Mumbai using several types of mathematical distributions for periods from 1969 to 2008. The Normal, Lognormal, Gamma, General maximum value, Weibull, Log pearson type III and Gumbell extreme value type I are proposed and tested together with their single distributions to identify the optimal model for maximum hourly rainfall. The selected model will be determined based on the minimum error produced by some criteria of goodness of-fit (GOF) tests. The results indicated General extreme value distribution is better than the other distributions in modeling hourly rainfall amount in Mumbai Region. These results however can vary between the rain gauge stations which are strongly influenced by their geographical, topographical and climatic changes. The following study shows that General Extreme value distribution is most suitable relatively for fitting of maximum hourly rainfall data based on GOF test.

The following study can be used in evaluating the more accurate Intensity Duration Frequency Curve for Mumbai region using the best fit. Subsequently the flood can be managed by proper design of storm water management system.

Keywords: Mumbai, Flood, Frequency distribution, Goodness of Fit test, General Extreme value.

1. Introduction

Many of the highly populated cities in the developing world that are located on the coast, for example, Mumbai, Bangkok, Dhaka and Jakarta, are highly susceptible to urban flooding. One of the main reasons cited for this is rapid urbanization, which causes changes in landscape owing to construction of urban infrastructures and changes in urban conveyances networks [1]. Mumbai city having an area of 437 sq km with a population of 12 million came to a complete halt due to the unprecedented rainfall of 944 mm during the 24 hours starting on 26th July 2005; with 380 mm falling in just 3 hours between 14:30 to 17:30 and hourly rainfall exceeding 190 mm/hr [2]. Determining the extent of flooding is an important role of the water resources community and provides a vital service to planners and engineers [3]. Engineers and Consultant faces the problem about the design rainfall value and so many times they assume the value on the basis of their experience. Hence it becomes very important to analyses the rainfall data and produce the curve and graph which can be used by engineers for their design purpose.

This paper tries to arrive on the best fit so that it can be used for further studies of rainfall data. Modeling daily rainfall data using various mathematical models has been an important research in hydrology for the last 30 years. The use of mathematical models of rainfall has been applied worldwide in order to give a better understanding about the rainfall pattern and its characteristics. This process involves the study on the sequence of dry and wet days as well as the study of rainfall amount on the wet days. Markov chain models are normally used to model the sequence of dry and wet days [4, 5, 6, 7] while rainfall

amount is often modeled using a two-parameter gamma distribution [8, 9, 10, 11]. Other theoretical distributions that have also been employed in the analysis of rainfall are exponential distribution [12], kappa B distribution [13], S distribution [12], mixed exponential distribution [14, 15, 16, and 17], weibull distribution [1] and skew normal distribution [16, 17]. The selection of the best fitting distribution has always been a key interest in the study of rainfall amount.
In this paper, we will focus on basic two and three parameters distributions in order to find the best model in fitting rainfall data. In order to verify the suitable distribution that best describes the rainfall amount, the new method of goodness-of-fit tests (GOF) based on the likelihood ratio statistics which has been developed by Zhang[18]. An additional criterion that is included in the analysis is the median of absolute difference between the hypothesized and empirical distribution function. The final result on the best fitting distribution will be chosen based on the minimum error specified by these GOF criteria.

2. Materials and methods

2.1 Study area description

Greater Mumbai comprises Mumbai, South Salsette and Trombay Islands, bounded by 18° 53' and 19° 20' north latitude and 72°45' and 73°00' east longitude. Mumbai has ridges along its western and eastern sides. Mumbai city receives seasonal rainfall for four months, from June to September. Average rainfall is 2500 mm, out of which 70 per cent is during July and August only. Mumbai is lined on the west by Arabian sea and is intercepted by number of creeks (Mahim, Mahul and Thane creeks), rivers (Mithi, Dahisar, Poisar and Oshiwara rivers, and their tributaries) and a complex nallah (drain) system [9]. The catchment considered in this study is Mithi River catchment which is located between north latitudes of 19°1'36"and 19°10'9" and east longitudes of 72°49'59" and 72°56'33". The location of the Mithi River has been shown in Figure 1.

2.2 Climate of study area

The Climate of Mumbai is a tropical wet and dry climate. Mumbai's climate can be best described as moderately hot with high level of humidity. Its coastal nature and tropical location ensures temperatures won't fluctuate much throughout the year, the mean average of 27.2 °C and average precipitation of 242.2 cm (95.35 inches).The mean maximum average temperatures in about 32 °C (90 °F) in summer and 30 °C (86 °F) in winter, while the average minimums are 25 °C (77 °F) in summer and 20.5 °C (68.9 °F) in winter. Mumbai experiences four distinct seasons: Winter (December–Feb); Summer (March–May); Monsoon (June–Sep); and Post-Monsoon (Oct–Dec) [19].

Figure 1: Location of Study area

2.3 Data collection

Mumbai has two weather monitoring stations, one at Santa Cruz Airport and the other at Navy Nagar near Colaba towards the southern tip of Mumbai. New monitoring equipment's are being installed at all the firefighting stations Mumbai. The data is shown on real-time on the new BMC. Daily rainfall series data for this study have been obtained from the Mumbai Meteorological Department for the periods ranging from 1969 to 2008 years. For this study, Santa Cruz rain gauge station was used based on the completeness of the data. The details have been shown in Figure 2.The stations are selected to represent rainfall pattern of rainfall for the whole Mumbai Region.

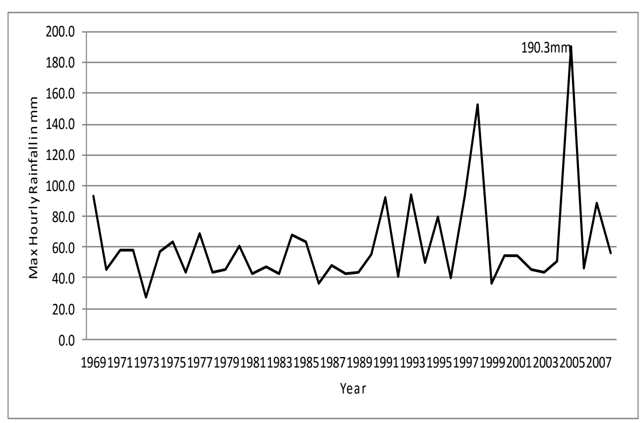

Figure 2: Maximum annual hourly rainfall for Mumbai Region

3. Modeling daily rainfall amount

Models of rainfall amount are described as follows with their probability density functions and cumulative distribution functions. Note that X is the random variable representing the hourly maximum rainfall amount. The model distribution was chosen on the basis of their relative advantage for analysis of rainfall data. The chosen distribution has been described in the Table. The description of model can be found in any basis statistics books. The Table 1 shows the properties of selected distributions.

3.1 Estimation of parameters

Many methods are available for parameter estimations, which include the method of moments (MM), maximum likelihood estimation (MLE), the least squares method (LS), L-moments and generalized probability weighted moments (GPWM).The MLE method is considered in this study because it provides the smallest variance as compared to other methods. The idea of this method is to find a set of parameters that will maximize the likelihood function. The parameters are obtained by differentiating the log likelihood function with respect to the parameters of the distribution. The all parameters was estimated by creating formulas in Microsoft excel 2010 and have been shown in Table 2.

4. Goodness-of-fit tests (GOF)

Three different mostly used GOF tests have been used in this study to identify the best fit models. The chosen distribution that best fits the daily rainfall amount is based on the minimum error indicate by all these seven tests. The description of all tests can be found in any basic statistics books. The results have been shown in Table 3.

5. Results and discussions

The excel sheet was developed for calculation of all statistics and result were prepared. The results and calculations were verified by using Easyfit software. The results have been summarized in Table. The Goodness of fit test was done for all distribution using three methods. The rank has been given on the basis of minimum value of error given by GOF test. First, we will proceed to give comments on the results of fitting distributions that are based on GOF criteria. Finally the remarks on the estimated parameters for the best model will be made. Mumbai received the highest maximum hourly rainfall amount during the year 2005 which was extreme flood event in the history of Mumbai. The average mean maximum hourly rainfall was found as 40 mm for the given data.

5.1 Fitting Distribution Based on GOF Criteria

The values of three goodness-of-fit criteria have been calculated and the best distribution was chosen based on the minimum error of GOF tests. The distributions were then ranked in ascending order based on those values. Unfortunately, when many criteria are used to identify the best distribution, it is more difficult to for the same data may be different for different analysis. In this study, we chose the best fitting distribution based on the majority of the tests, since we did not investigate which is the most effective test. Based on the results, General Extreme value distribution was found to be best fitting curve for maximum hourly rainfall data of Mumbai. Based on the result Gamma and Lognormal distribution was found to be suitable for fitting of data while Log Pearson type III fails in Chi Square test.

The Figure 3 and 4 shows the Probability distribution function f(x) and Cumulative distribution function for all selected probability distribution functions. The General extreme value catches the peak of the rainfall data properly and at the same time try to cover the all rainfall values under it. Hence it is found to be most suitable for fitting rainfall data.

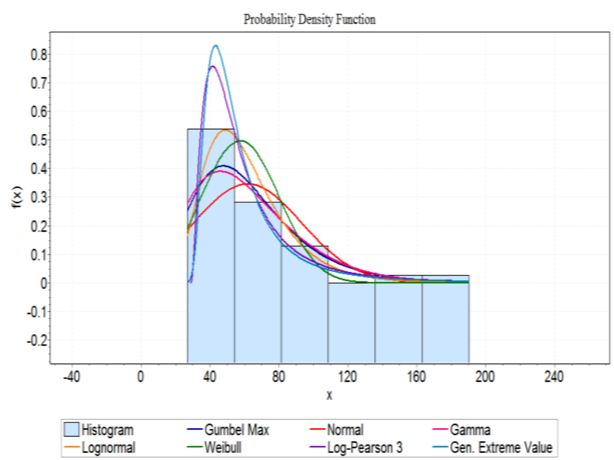

Figure 3: Probability density functions for selected distribution

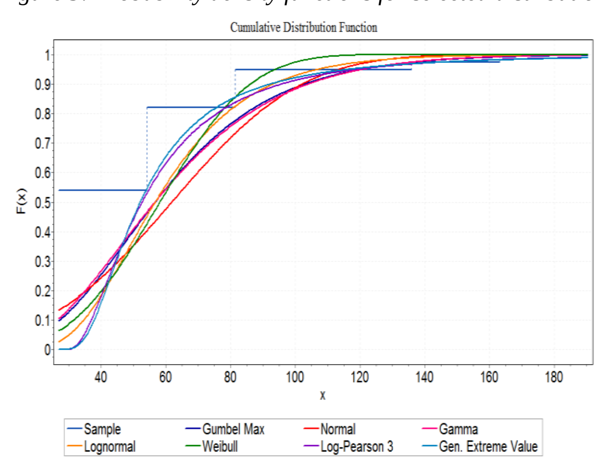

Figure 4: Cumulative distribution functions for selected distribution

Table 1: Density functions of selected distributions

S.No	Distributions	f(X)	F(X)
1	Normal	$\frac{\exp -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}{\sigma\sqrt{2\pi}}$	$\varphi\left(\frac{x-\mu}{\sigma}\right)$
2	Lognormal	$\frac{\exp -\frac{1}{2}\left(\frac{ln\,(x)-\mu}{\sigma}\right)^2}{x\sigma\sqrt{2\pi}}$	$\varphi\left(\frac{\ln(x)-\mu}{\sigma}\right)$
3	Gamma	$\frac{x^{\alpha-1}}{\beta^{\alpha}\,\Gamma(\alpha)}\exp\,(-x/\beta)$	$\frac{\Gamma_{x/\beta}(\alpha)}{\Gamma\,(\alpha)}$
4	Gumbell	$\frac{1}{\sigma}\exp(-z-\exp(-z))$	$\exp(-\exp(-z))$
5	Weibull	$\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$	1- $\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$
6	Logpearson Type III	$\frac{1}{xI\beta I\tau\,(\alpha)}\left(\frac{\ln(x)-\gamma}{\beta}\right)^{\alpha-1} exp\left(-\frac{\ln(x)-\gamma}{\beta}\right)$	$\frac{\tau\,(\ln(x)-\gamma)/(\beta)^{\alpha}}{\tau\,(\alpha)}$
7	General Extreme value	$\begin{cases}\frac{1}{\sigma}\exp\,\left(-\left(1+kz\right)^{-\frac{1}{k}}\right)\,(1+kz)^{-1-\frac{1}{k}} & k\neq 0\\ \frac{1}{\sigma}\exp(-z-\exp(-z)) & k=0\end{cases}$	$\begin{cases}\exp\,\left(-\left(1+kz\right)^{-\frac{1}{k}}\right) & k\neq 0\\ \exp(-z-\exp(-z)) & k=0\end{cases}$

Table 2: Parameters of selected distributions

S. No	Distributions	Parameters					
		μ	σ	α	β	ϒ	k
1	Normal	61.862	31.454	--	--	--	--
2	Lognormal	4.0383	0.38644	--	--	--	--
3	Gamma	--	--	3.8679	15.993	--	--
4	Gumbell	24.525	47.705	--	--	--	--
5	Weibull	--	--	3.0619	65.598	--	--
6	Log Pearson Type III	--	--	2.9354	0.2285	3.3675	--
7	General Extreme value	47.068	12.842	--	--	--	0.37226

Table 3: GOF value for selected probability distributions

S.No	Distributions	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Normal	0.22029	6	3.3076	6	13.109	5
2	Lognormal	0.14199	4	1.2929	3	2.4429	3
3	Gamma	0.12927	3	1.1503	2	1.4223	2
4	Gumbell maximum value	0.18167	7	2.004	4	2.5473	4
5	Weibull	0.16844	5	3.0654	5	13.547	6
6	Log Pearson Type III	0.12097	2	4.3788	7	Fails	
7	General Extreme value	0.10296	1	0.53227	1	1.118	1

6. Conclusions

The search for the best fitting distribution for maximum hourly rainfall data amount has been the main interest in several studies. Various forms of distributions have been tested in order to find the best fitting distribution. Different tests of goodness-of-fit have been attempted in the studies. In this study, the General maximum value has been identified as the best fitting distribution for rainfall data in Mumbai. However, the gamma and lognormal distribution were also considered as the best based on relative higher rank. Based on these findings, we can conclude that the pattern of rainfall distribution Mumbai could be fit and analyzed for various statistics with General extreme value.

References

[1] J. Marsalek, B. E. Jimenez-Cisneros, P. A. Malmquist, M. Karamouz, J. Goldenfum, and B. Chocat, "Urban Water Cycle Processes and Interactions, IHP-VI Technical Document in Hydrology No 78, 2006, (UNESCO, Paris).

[2] K. Gupta, " Urban flood resilience planning and management and lessons for the future: a case study of Mumbai, India, Urban Water Journal, Vol. 4, No. 3, September 2007, 183-194.

[3] P. F. Hudson, and R.R Colditz, "Flood delineation in a large and complex alluvial valley, lower Panuco basin, Mexico", Journal of Hydrology, 280: 229–245, 2003.

[4] Gabriel, K.R. and J. Neumann, 1962. A Markovchain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of Royal Meteorology Society, 88: 90-95.

[5] 2. Roldan, J. and D.A. Woolhiser, 1982. Stochastic daily precipitation models 1. A comparison of occurence process. Water Resources Research, 18(5): 1451-1459.

[6] Stern, R.D. and R. Coe, 1984. A model fitting analysis of daily rainfall data. Journal of Royal Statistical Society Series A, 147: 1-34.

[7] Jimoh, O.D. and P. Webster, 1996. Optimum order of Markov chain for daily rainfall in Nigeria. Journal of Hydrology, 185: 45-69.

[8] Ison, N.T., A.M. Feyerherm and L.D. Bark, 1971.Wet period precipitation and the gamma distribution. Journal of Applied Meteorology, 10: 658-665.

[9] Katz, R.W., 1977. Precipitation as chaindependent process. Journal of Applied Meteorology, 16: 671-676.

[10] Buishand, T.A., 1978. Some remarks on the use of daily rainfall models. Journal of Hydrology, 36: 295-308.

[11] Aksoy, H., 2000. Use of gamma distribution in hydrological analysis. Turkey Journal of Engineering Environmental Sciences, 24: 419-428.

[12] Mielke, P.W., 1973. Another family of distributions for describing and analyzing precipitation data. Journal of Applied Meteorology,12: 275-280.

[13] Swift Llyold W, Jr. and H.T. Schreuder, 1981. B Fitting daily precipitation amounts using the S distribution. Monthly Weather Review, 109(12): 2535-2540.

[14] Woolhiser, D.A. and J. Roldan, 1982. Stochastic daily precipitation models 2. A comparison of distribution of amounts. Water Resources Research, 18(5): 1461-1468.

[15] Chapman, T.G., 1997. Stochastic models for daily rainfall in the Western Pacific. Mathematics andComputers in Simulation, 43: 351-358.

[16] 15. Chapman, T.G., 1998. Stochastic modelling of daily rainfall: the impact of adjoining wet days on the distribution of rainfall amounts. Environmental Modelling and Software, 13: 317-324.

[17] Wilks, D.S., 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agricultural and Forest Meteorology, 93: 153-169.

[18] Zhang, J., 2002. Powerful goodness-of-fit tests based on the likelihood ratio. Journal of Royal Statistical Society Series B, 64(2): 281-294.

[19] Vega, Anthony J (2007). Climatology. Jones and Bartlett Publishers. p. 267.