Int.J.Curr.Microbiol.App.Sci (2019) *8*(6): 1390-1396

International Journal of Current Microbiology and Applied Sciences
***ISSN: 2319-7706* Volume 8 Number 06 (2019)**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2019.806.168

Frequency Analysis of Rainfall Data Using Probability Distribution Models

H. Baghel*, H.K. Mittal, P.K. Singh, K.K. Yadav and S. Jain

Department of Soil and Water Engineering, CTAE, MPUAT, Udaipur 313001, India

**Corresponding author:*

Keywords

Rainfall, distribution models, frequency analysis

Article Info

Accepted:
12 May 2019
Available Online:
10 June 2019

A B S T R A C T

Rainfall is a prime input for various engineering design such as hydraulic structures, water conservation structures, bridges and culverts, canals, storm water sewer and road drainage system. The detailed statistical analysis of each region is essential to estimate the relevant input value for design and analysis of engineering structures and also for crop planning. The present study comprises statistical analysis i.e. frequency analysis of daily maximum rainfall data of Udaipur district. The daily rainfall data for a period of 56 years is collected to evaluate designed value of rainfall using probability distribution models. Around 07 different probability distributions (Gamble's extreme value type I, Logpearson type III, Lognormal, Normal, Exponential, Pearson type III and Gamma distribution) were used to evaluate maximum daily rainfall. Chi-squared tests were used for the goodness of fit of the probability distributions. Results showed that Lognormal distribution and Gumbel distribution found to be have least critical values in the tests hence consider as the best fit distribution for given sample population. Also maximum daily expected value of rainfall for various return periods were evaluated using all distribution model under consideration.

Introduction

Analysis of daily maximum rainfall of different return periods is a basic tool for safe and economical planning and design of small dams, bridges, culverts, irrigation and drainage work etc. Though the nature of rainfall is erratic and varies with time and space, yet it is possible to predict design rainfall fairly accurately for certain return periods using various probability distributions (Upadhaya and Singh, 1998). Probability analysis can be used for predicting the occurrence of future events of rainfall from the available data with the help of statistical methods (Kumar and Kumar, 1989). Anaya Kalita *et al.*, (2017) worked on frequency analysis of daily rainfall data of 24 years to determine the annual one day maximum rainfall and discharge of Ukiam (Brahmaputra River). Weibull's plotting position Gumbel, Log Pearson and Log normal probability distribution functions were fitted. For determination of goodness of fit chi square test was carried out. The results found showed that the Log Pearson and Log Normal were the best fit probability distribution. Esberto (2018) determined the best fit frequency distribution of rainfall patterns for event forecasting in order to address potential

disasters using 60 Probability Distribution Functions (PDF). Rainfall data were analyzed using Chi-Square and K-S goodness-of-fit tests. Amin *et al.*, (2016) analyzed to find the best-fit probability distribution of annual maximum rainfall based on a twenty-four-hour sample in the northern regions of Pakistan using four probability distributions: normal, log-normal, log-Pearson type-III and Gumbel max. Based on the scores of goodness of fit tests, the normal distribution was found to be the best-fit probability distribution at the Mardan rainfall gauging station. The log-Pearson type-III distribution was found to be the best-fit probability distribution at the rest of the rainfall gauging stations. This project is an effort to summaries the rainfall features for the Udaipur district. The total rainfall received in a given period at a location is highly variable from one year to another. The variability depends on the type of climate and the length of the considered period, the statistical inferences found in this study are important for designing optimum flood control facilities. Basically frequency analysis of rainfall is used for different purposes as mentioned below:

Probability of exceedance for design purposes

The selection of the probability of exceedance or return period for design purposes is related to the damage the excess or the shortage of rainfall may cause the risk one wants to accept and the life time of the project.

Probabilities of exceedance for management purposes

Information on the rainfall depth that can be expected in a specific period under various weather conditions is required for management and planning purposes. For rain-fed agriculture, rainfall is the single most important agro-meteorological variable influencing crop production.

Materials and Methods

Udaipur district is situated between $23^0$40’ and 25^0 30’ north latitude and 73^0 0’ and 74^0 35’ east longitude. It is located in the south eastern part of Rajasthan and lies in Aravali ranges. The district is having 1, 89,746 ha area surrounded by hills (Google map, cited on 25 May. 2019). 56 years of daily mean rainfall data from 12 raingauge stations of Udaipur district have collected from ‘Rainfall Profile of Udaipur’ Manual published by Indian Meteorological Department Jaipur (2014).

$\overline{X}$ is the arithmetic Mean, X_i is Variate, N is the total number of observations, S is Standard Deviation, C_v is the coefficient of Variation and C_s is the Coefficient of skewness.

Tests for goodness of fit (verification of sample population)

The goodness of fit of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. In stochastic hydrology there is a method whether or not a particular distribution adequately fits a set of observation-

Compare observed relative frequency with theoretical relative frequency using Chi-square test

Chi-square test

The chi-squared test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories.

$$\chi_c^2 = \sum_{i=1}^{N} (N_i - E_i)^2 / E_i$$

Where N is the total number of observations, N_i is the observed relative frequencies, and E_i is the theoretical or probable relative frequencies. If $\chi_c^2 = 0$, it indicates that observed and theoretical frequencies agree exactly while if $\chi_c^2 > 0$, they do not agree exactly. The hypothesis that the data follows a specific distribution is accepted if,

$$\chi^2_{data} < \chi^2_{\alpha-1,K-p-1}$$

Where α is the significance level and K-P-1 is the degree of freedom. Test is carried out at 10% significance level. Critical values of chi-square test for a particular degree of freedom and at particular significance level can be obtained from Chi-square distribution table.

Frequency distribution models

Gumbel's extreme value distribution model

Gumbel found that the probability of occurrence of an event, equal or larger than a value is given by the equation,

$$P(X > x_0) = 1 - e^{-e^{-y}}$$
$$y_t = -(\ln\ln\frac{T}{T-1})$$
$$X_T = \bar{X} + K\sigma_{n-1}$$

For N=56 the values for $\overline{y_n}$ and σ_n are 0.551 and 1.1696 respectively from standard tables (Ghanshyamdas, 2014).

Log-Pearson type lll distribution

$$z = \log x$$

For any recurrence interval T above equation can be expressed as

$$z_t = \log x_t$$

Applying general equation chow, z_T data series can be expressed a

$$z_T = \bar{z} + K_f \sigma_z$$

Where, K_f is the frequency factor, C_z is the coefficient of skewness, $\bar{z}$ is the mean of the representative variate sample z, σ_z is the standard deviation of the representative variate sample z. value of K_f can be determined by using the standard table for a specific value of C_z and recurrence interval T.

Log normal probability distribution method

The flood or rainfall of any return period which follows the log normal probability law is computed from:

$$Q_T = \bar{Q} + K\sigma_n$$

Where K is log normal frequency factor. A function of skewness coefficient, given by

$$C_s = 3C_v + C_v^3$$

Where C_v is a coefficient of variation and given by

$$C_v = \frac{\sigma}{Q}$$

The value of K can be determined from the normal probability table.

Normal distribution

It is also a most widely used method in extreme value distributions.

$$X_T = \bar{X} + K_T\sigma$$
$$K_T = Z = \frac{X_T - \bar{X}}{\sigma}$$

$$K_T = w - \frac{2.515517 + 0.802853w + 0.010328w^2}{1 + 1.432788W + 0.189269w^2 + 0.001308w^3}$$

Gamma distribution

Gamma distribution – a distribution of sum of b independent and identical exponentially distributed random variables.

$$f(x) = \frac{\lambda^{\beta}(x-\varepsilon)^{\xi-1}e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)}$$

Γ=Gamma function

$$\Gamma(\eta) = \int_0^{\alpha} t^{n-1}e^{-t}dt$$

Pearson type III

Named after the statistician Pearson, it is also called three-parameter gamma distribution. A lower bound is introduced through the third parameter (e).

$$f(x) = \frac{\lambda^{\beta}(x-\varepsilon)^{\xi-1}e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)}$$

Exponential distribution

In hydrology, the inter arrival time (time between stochastic hydrologic events) is described by exponential distribution.

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0, \lambda = \frac{1}{x}$$

Variance = $1/\lambda^2$

Results and Discussion

56 years of daily rainfall data is taken from the IMD manual published in 2014. For the series of daily rainfall data, annual maximum daily rainfall data is arranged. The seven probability distributions were subjected to test from goodness of fit tests (Chi-squared test). The purpose of the study was to find the best-fit probability distributions for district Udaipur (Table 1–3).

Table 1. Formula of Statistical Parameters

Sr.No.	Parameter name	Formula
1	Arithmetic mean	$\bar{X} = \frac{\sum_{N}^{i=1} X_i}{N}$
2	Standard deviation	$S = \sqrt{\frac{\sum_{i=1}^{N} X_i}{N-1}}$
3	Coefficient of variation	$C_v = \frac{S}{\bar{X}}$
4	Coefficient of skewness	$C_s = \frac{N\sum (X - \bar{X})^3}{(N-1)(N-2)S^3}$

Table 2. Goodness of fit result summary

Sr.no.	Distribution Model	Test Performed	Calculated values for χ_c^2 test	Degree of freedom	Critical values at 10 % significance level	Result
1	Gumbel's distribution	Chi-square Test	9.406	7	12.02	Accepted
2	Log-Pearson Type-III distribution	Chi-square Test	22.793	6	10.64	Rejected
3	Normal distribution	Chi-square Test	20.851	7	12.02	Rejected
4	Lognormal distribution	Chi-square Test	8.444	6	10.64	Accepted
5	Exponential distribution	Chi-square Test	48.331	8	13.362	Rejected
6	Pearson-III distribution	Chi-square Test	54.742	6	10.64	Rejected
7	Gamma distribution	Chi-square Test	10.163	7	12.02	Accepted

A. Magnitude of Daily Rainfall (mm) For Various Distribution Models

Table.3 Magnitude of designed value of daily rainfall for various distributions models and return

Table 4. Magnitude of designed value of daily rainfall for various distribution m**odels and return periods.**

Distribution model	Return period in years									
	5	**10**	**25**	**50**	**100**	**200**	**300**	**400**	**500**	**1000**
Gumbel distribution	73.85	89.77	109.88	124.80	139.61	154.37	162.99	169.10	173.84	188.5
Log-Pearson Type-III distribution	69.40	86.03	109.52	128.96	150.09	173.21	180.03	187.12	188.73	235.9
Normal distribution	74.60	85.52	97.17	104.69	111.45	117.64	121.04	123.37	125.14	130.4
Lognormal distribution	70.16	84.74	103.63	118.01	132.65	147.62	156.57	163.00	168.05	184.0
Exponential distribution	86.53	123.79	173.06	210.32	247.59	284.86	306.65	322.12	334.12	371.3
Pearson-III distribution	72.53	86.97	104.41	116.79	128.70	140.28	143.41	146.54	149.66	165.2
Gamma distribution	72.74	86.95	103.95	115.96	127.48	138.62	144.99	149.46	152.89	163.4

The maximum values of expected rainfall or rainfall estimates calculated using a probability distribution that does not provide the best-fit may yield values that are higher or lower than the actual values. These calculations may be used to influence decisions relating to local economics and hydrologic safety systems.

The tests were performed at 10% significance level. Out of 07 models 03 models have passed in the tests. The Log-normal distribution and Gumbel distribution provided the best-fit probability distribution with the least score for the test. The expected values of designed rainfall or rainfall estimates calculated using the best-fit probability distributions at the rainfall gauging stations might be used by design engineers to safely and feasibly design hydrologic projects.

References

Agarwal, M.C., Katiyar, V.S. and Babu, R. Probability analysis of annual maximum daily rainfall of U.P. Himalaya. *Indian journal of Soil Conservation*, 16(1): 35-43, 1995.

Al-suhili, R.H., and Khanbilvardi, R. Frequency Analysis of the Monthly Rainfall Data at Sulaimania Region, Iraq. *American Journal of Engineering Research (AJER)*, 03(05): 12-222, 2014.

Amin, T.A., Rizwan, M. and Alazba, A.A. A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan. Open Life Science, 11:432-440, 2016.

Arvind, G., Kumar, A.P., Girishkarthi, S. and Suribabu, C.R. Statistical Analysis of 30 Years Rainfall Data: A Case Study. IOP Conf. Series: Earth and Environmental Science, 80(2017):01-09, 2017.

Bhakar, S. R., Bansal, A.K., Chhajed N. and Purohit, R. C. Frequency Analysis of Consecutive Days Maximum Rainfall at Banswara, Rajasthan, India. ARPN Journal of Engineering and Applied Sciences, 31(3): 64-67, 2006.

Bhavyashree, S. and Bhattacharya, B. Fitting Probability Distributions for Rainfall Analysis of Karnataka, India. International Journal of Current Microbiology and Applied Sciences, 7(3): 1498-1506, 2018.

Dirk, R. Frequency Analysis of Rainfall Data. Note Submitted To Department of Earth And Environment Science, Katholieke Universiteit, Belgium, 2013.

Esberto, M.D.P. Probability Distribution Fitting of Rainfall Patterns in Philippine Regions for Effective Risk Management. Environment and Ecology Research, 6(3): 178-86, 2018.

Ghnshyamdas. Hydrology and Soil Conservation Engineering: Including Watershed Management, PHI Learning Private Limited.99-123p, 2014.

Ghosh, S., Roy, M. K., and Biswas, S. C. Determination of the best fit probability distribution for monthly rainfall data in bangladesh. American Journal of Mathematics and Statistics, 6(4):170–174, 2016.

Kalita, A., Bormudoi, A. and Saikia, M.D. Probability Distribution of Rainfall and Discharge of Kulsi River Basin. International Journal of Engineering and Advanced Technology (IJEAT), 6(4): 31-37, 2017.

Kumar, A. Probability Analysis for Prediction of Annual Maximum Daily Rainfall for Pantnagar. Indian Journal of Soil Conservation, 27(2):171-173, 1999.

Kumar, D. and Kumar, S. Rainfall Distribution Pattern Using Frequency Analysis. Journal of Agricultural Engineering, 26(1):33-38, 1999.

Kumar, R. and Bhardwaj, A. Probability analysis of return period of daily maximum rainfall in annual data set of Ludhiana, Punjab. Indian Journal of

Agricultural Research, 49(2):160-164, 2015.

Latitude and longitude of Udaipur. (https://www.google.co.in/maps/place/Udaipur,+Rajasthan/@24.5873424,73.6407606,23152m/data=!3m1!1e3!4m2!3m1!1s0x3967e56550a14411:0xdbd8c28455b868b0!6m1!1e1).

Nemichandrappa M., Balakrishnan and Senthilvel. Probability and Confidence Limit Analysis of Rainfall in Raichur Region. Karnataka.journal of Agriculture Science, 23 (5): (737-741), 2010.

Sharma, M.A. and Singh, J. Use of Probability Distribution in Rainfall Analysis. New York Science Journal 3(9): 40-49, 2010.

Singh, O.P., Singh, S.S. And Kumar, S. Rainfall Profile For Udaipur District, Meteorological Centre, Jaipur, IMD Department New Delhi, 2013.

Sreedhar, B.R. Fitting of Probability Distribution for Analyzing the Rainfall Data in the State of Andhra Pradesh, India. International Journal of Applied Engineering Research, 14(3): 835-839, 2018.

Upadhaya, A. and Singh, S.R. Estimation of consecutive day's maximum rainfall by various methods and their comparison. Indian Jornal of Soil Conservation, 26(3):193-200, 1998.

How to cite this article:

Baghel, H., H.K. Mittal, P.K. Singh, K.K. Yadav and Jain, S. 2019. Frequency Analysis of Rainfall Data Using Probability Distribution Models. *Int.J.Curr.Microbiol.App.Sci.* 8(06): 1390-1396. doi: https://doi.org/10.20546/ijcmas.2019.806.168