

Original Research Article

<https://doi.org/10.20546/ijcmas.2018.703.178>

## Fitting Probability Distributions for Rainfall Analysis of Karnataka, India

S. Bhavyashree\* and Banjul Bhattacharyya

*Department of Agricultural Statistics, BCKV, West Bengal-741252, India*

*\*Corresponding author*

### ABSTRACT

#### Keywords

Rainfall, Probability distributions and Goodness-of-fit test

#### Article Info

##### Accepted:

12 February 2018

##### Available Online:

10 March 2018

Distribution of Rainfall plays important role in agriculture and efficient utilization of the water resources. The present study was conducted to know the distribution of rainfall during Monsoon period of different districts of Karnataka. The secondary data of Rainfall over a period of 34 years (1980-2013) was collected from AICRP on Agro Meteorology, UAS Bangalore. Around 20 different probability distributions were used to evaluate the best fit for maximum daily rainfall (mm). Kolmogorov-Smirnov, Anderson Darling and Chi-squared tests were used for the goodness of fit of the probability distributions showed that, for Log logistic (3P), Gen gamma (4P), Dagum (3P,4P), Gamma (2P), Pearson 5(3P), Weibull (3P), Johnson SB (4P) were found to be the best fit for different districts of the state.

### Introduction

Indian economy is mainly based on Agriculture and allied, Industry and Service Sectors. Total production of agriculture sector is \$366.92 billion. India is 2nd larger producer of agriculture product. India accounts for 7.68 percent of total global agricultural output.

Contribution of Agriculture sector in Indian economy is much higher than world's average (6.1%). Contribution of Industry and Services sector is lower than world's average 30.5% for Industry sector and 63.5% for Services sector. Karnataka is one of the fastest growing states in India which contributes 13% GDP for

agriculture. Agriculture in Karnataka is heavily dependent on the southwest monsoon because of the extent of arid land in the state. The average annual rainfall in Karnataka is 1248 mm. The analysis of rainfall data mainly based upon its distribution type. Therefore, the study of the rainfall distribution is very important for the state which adds up to Indian economy.

In our life, the probability distributions are used in different fields of science such as engineering, medicine, economic and agricultural science. Probability distributions of rainfall have been studied by many researchers

Tariq Mahgoub Mohamed and Abbas Abd Allah Ibrahim (2016) analyzed Annual rainfall data for fourteen rainfall stations in Sudan during the period 1971 to 2010 to select the best probability distribution for every station. They tested five distributions, namely Normal, Log normal, Gamma, Weibull and exponential distribution. The normal and gamma distribution were selected as the best fit probability distribution for the annual rainfall in Sudan during the period of the study. Sanjib Ghosh et.al., (2016) made an attempt to determine the best fitted distribution to describe the monthly rainfall data for the period of 1979 to 2013 of distantly located stations in Bangladesh such as Chittagong, Dhaka, Rajshahi and Sylhet. The Normal, Lognormal, Gamma, Weibull, Inverse Gaussain and Generalized Extreme Value distributions were fitted for these purposes using the method of L-moment. Mandal *et al.*, (2014) analyzed 16 years of rainfall (1995-2010) data at Daspalla region in Odisha of eastern India for prediction using six probability distributions, forecasting the probable date of onset and withdrawal of monsoon, occurrence of dry spells by using Markov chain model and crop planning for the region. It was found that for prediction of monsoon and post-monsoon rainfall, log Pearson type III and Gumbel were the best-fit probability distribution. Lala I.P. Ray *et al.*, (2013) analyzed daily rainfall data for 28 years (1983-2010) of Central Meghalaya, Nongstoin station for estimating maximum daily rainfall. The annual maximum daily rainfall data was fitted to five different probability distribution functions i.e. Normal, Log-normal, Pearson Type-III, Log Pearson Type-III and Gumbel Type-I extreme. The probable rainfall value for different return periods was estimated. They found that, Gumbel distribution may be used to predict maximum rainfall, which will be a great importance for economic planning and design of small and medium hydraulic structures.

The main objective for this study is the analysis of the probability distribution of the monthly rainfall in 10 different districts of Karnataka. Different probability distributions are compared in this paper to best fit for different districts of Karnataka.

## **Materials and Methods**

### **Data**

Rainfall of Ten districts of Karnataka was selected with monthly rainfall series during the period 1980 to 2013. These 10 districts rainfall data were selected based sampling done using stratified random sampling among 16 reasonably long records for the daily and monthly rainfall data in locations which represent different climatic zones in Karnataka.

The districts were classified based on the amount of rainfall as <1000mm, 1000-3000 mm and >3000mm. Then the stratified random sampling is applied and allocation as been made using proportional allocation and the selected districted were, Chitradurga, Davanagere, Tumkur, kolar, mandya. Bangalore rural for <1000mm rainfall, Hassan and chikkamagaluru under 1000-2000mm rainfall and Dakshina kannada and Udapi districts for >3000mm rainfall.

### **Methodology**

The procedures used for this work can be summarized in the following steps:

### **Statistics of annual rainfall**

Using the sample data ( $i=1,2,\dots,n$ ) the basic statistical descriptors of the annual rainfall series, the mean, standard deviation, coefficient of variation, skewness, kurtosis, minimum and maximum values, have been estimated for each district.

Twenty probability distributions were used to select the best fit probability distribution for monthly rainfall of Karnataka. The description of the probability distribution functions are presented in Table 1.

**Testing the goodness of fit**

The goodness-of-fit tests namely, Kolmogorov Smirnov test, Anderson-Darling test and Chi-Squared were used at  $\alpha$  (0.05) level of significance for the selection of the best fit distribution.

The hypothesis under the GOF test is:

$H_0$ : MMR (Monthly monsoon rainfall) data follow the specified distribution,

$H_1$ : MMR data does not follow the specified distribution.

The best fitted distribution is selected based on the minimum error produced, which is evaluated by the following techniques:

**Kolmogorov-Smirnov test**

Is used to decide if a sample  $\{x_1, x_2, \dots, x_n\}$  with CDF  $F(x)$  comes from a hypothesized continuous distribution. The Kolmogorov-Smirnov statistic ( $D$ ) is based on the largest vertical difference between the theoretical CDF and the empirical (observed) CDF and is given by

$$D = \max_{1 \leq i \leq n} \left[ F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right]$$

A large difference indicates an inconsistency between the observed data and the statistical model.

The Anderson-Darling test it was introduced by Anderson and Darling (1952) to place more weight or discriminating power at the tails of

the distribution. This can be important when the tails of the selected theoretical distribution are of practical significance. It is used to compare the fit of an observed CDF to an expected CDF. This test gives more weight to the tails than the Kolmogorov -Smirnov test. The test statistic  $A^2$ , is defined as

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln F(x_i) + \ln(1-F(X_{n-i+1}))]$$

The Chi-Squared test is used to determine if a sample comes from a population with a specific distribution. The Chi-Squared statistic is defined as

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Where  $O_i$  is the observed frequency for bin  $i$ , and  $E_i$  is the expected (theoretical)  $x_2$  frequency for bin  $i$  calculated by  $E_i = F(x_2) - F(x_1)$ ,  $F$  is the CDF of the are the probability distribution being tested,  $x_1$  and  $x_2$  limits for bin  $i$ .

**Results and Discussion**

**Descriptive statistics**

Descriptive Statistics viz. minimum, maximum, range, mean, standard deviation (SD), coefficient of variation (CV), skewness and kurtosis of AMRAM for the 10 districts are summarized in Table 2.

Over the period of 34 years among the 10 districts of Karnataka, it has been observed that the maximum AMRAM was highest (2154.90mm) and lowest (300.80 mm) in Udapi and Tumkur districts respectively. Within the same period, the minimum MMR was as low as 2.50 mm in Chitradurga and highest (92.95mm) in Dakshina kannada district.

**Table.1** Description of continuous probability distributions

Sl. No.	Distributions	Probability density Function $f(x)$	Ranges/ Values	Parameters
1.	Burr (3P)	$f(x) = \frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha-1}}{\beta \left(1 + \left(\frac{x}{\beta}\right)^\alpha\right)^{k+1}}$	$k, \alpha, \beta > 0,$ $0 \leq x < +\infty$	$k, \alpha = shape$ $\beta = scale$
2.	Dagum (3P)	$\frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha k-1}}{\beta \left(1 + \left(\frac{x}{\beta}\right)^\alpha\right)^{k+1}}$	$k, \alpha, \beta > 0,$ $0 \leq x < +\infty$	$k, \alpha = shape$ $\beta = scale$
3.	Dagum (4P)	$\frac{\alpha k \left(\frac{x-\gamma}{\beta}\right)^{\alpha k-1}}{\beta \left(1 + \left(\frac{x-\gamma}{\beta}\right)^\alpha\right)^{k+1}}$	$k, \alpha, \beta > 0,$ $\gamma \leq x < +\infty$	$k, \alpha = shape$ $\beta = scale$ $\gamma = location$
4.	Fatigue Life (3P)	$\frac{\sqrt{(x-\gamma)/\beta} + \sqrt{\beta/(x-\gamma)}}{2\alpha(x-\gamma)} \cdot \phi\left(\frac{1}{\alpha} \left(\sqrt{\frac{x-\gamma}{\beta}} - \sqrt{\frac{\beta}{x-\gamma}}\right)\right)$	$\alpha, \beta > 0,$ $\gamma < x < +\infty$	$\alpha = Shape$ $\beta = Scale$ $\gamma = Location$
5.	Gamma (2P)	$f(x) = \frac{(x)^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp(-x/\beta)$	$\alpha, \beta > 0,$ $0 \leq x < +\infty$	$\alpha = Shape$ $\beta = Scale$
6.	Gamma (3P)	$f(x) = \frac{(x-\gamma)^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp(-(x-\gamma)/\beta)$	$\alpha, \beta > 0,$ $\gamma \leq x < +\infty$	$\alpha = Shape,$ $\beta = Scale$ $\gamma = Location$
7.	Gen. Extreme Value (3P)	$f(x) = \begin{cases} \frac{1}{\sigma} \exp(-(1+kz)^{-1/k})(1+kz)^{-1-1/k} & k \neq 0 \\ \frac{1}{\sigma} \exp(-z - \exp(-z)) & k = 0 \end{cases}$	$\alpha, \beta > 0,$ $\gamma \leq x < +\infty$	$\alpha = Shape$ $\beta = Scale$ $\gamma = Location$

8.	Gen. Gamma (3P)	$\frac{kx^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)} \exp(-(\frac{x}{\beta})^k)$	$k, \alpha, \beta > 0,$ $0 \leq x < +\infty$	$k, \alpha = shape$ $\beta = scale$
9.	Gen. Gamma (4P)	$f(x) = \frac{k(x-\gamma)^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)} \exp(-((x-\gamma)/\beta)^k)$	$k, \alpha, \beta > 0,$ $\gamma \leq x < +\infty$	$k, \alpha = shape$ $\beta = scale$ $\gamma = location$
10.	Gumbel Max (2P)	$f(x) = \frac{1}{\sigma} \exp(-z - \exp(-z))$  Where, $f(x) = \frac{x - \mu}{\sigma}$	$-\infty < x < +\infty$	$\sigma = Scale$ $\mu = location$
11.	Inv. Gaussian (3P)	$\sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right]$	$\lambda, \mu > 0,$ $0 < x < +\infty$	$\lambda = shape$ $\mu = location$
12.	Johnson SB (4P)	$f(x) = \frac{\delta}{\lambda\sqrt{2\pi z(1-z)}} \exp\left(-\frac{1}{2}\left(\gamma + \delta \ln\left(\frac{z}{1-z}\right)\right)^2\right)$	$\xi \leq x \leq \xi + \lambda$	$\delta, \gamma = Shape$ $\lambda = scale$ $?? = location$
13.	Log-Logistic (3P)	$\frac{\alpha}{\beta} \left[\frac{x-\gamma}{\beta}\right]^{\alpha-1} \left[1 + \left[\frac{x-\gamma}{\beta}\right]^\alpha\right]^{-2}$	$\alpha, \beta > 0,$ $\gamma \leq x < +\infty$	$\alpha = shape$ $\beta = scale$ $\gamma = location$
14.	Lognormal (3P)	$\frac{\exp\left[-\frac{1}{2}\left[\frac{\ln(x-\gamma)-\mu}{\sigma}\right]^2\right]}{(x-\gamma)\sigma\sqrt{2\pi}}$	$\sigma > 0,$ $-\infty < \mu < \infty,$ $\gamma < x < +\infty$	$\sigma = shape$ $\mu = scale$ $\gamma = location$
15.	Log-Pearson 3(3P)	$f(x) = \frac{1}{X \beta \Gamma(\alpha)} \left(\frac{\ln(X)-\gamma}{\beta}\right)^{\alpha-1} \exp\left(-\frac{\ln(X)-\gamma}{\beta}\right)$	$\alpha > 0,$ $\beta \neq 0,$ $\begin{cases} 0 < x \leq e^\gamma & \beta < 0 \\ e^\gamma \leq x < +\infty & \beta > 0 \end{cases}$	$\alpha = shape$ $\beta = scale$ $\gamma = location$

16.	Pearson 5 (3P)	$\frac{\exp(-\beta/(x-\gamma))}{\beta\Gamma(\alpha)((x-\gamma)/\beta)^{\alpha+1}}$	$\alpha, \beta > 0$ $\gamma < x < +\infty$	$\alpha = \text{shape}$ $\beta = \text{scale}$ $\gamma = \text{location}$
17.	Pearson 6 (3P)	$f(x) = \frac{(x/\beta)^{\alpha_1-1}}{\beta B(\alpha_1, \alpha_2)(1+x/\beta)^{\alpha_1+\alpha_2}}$	$\alpha_1, \alpha_2, \beta > 0$ $0 \leq x < +\infty$	$\alpha_1, \alpha_2 = \text{shape}$ $\beta = \text{scale}$
18.	Pearson 6 (4P)	$f(x) = \frac{((x-\gamma)/\beta)^{\alpha_1-1}}{\beta B(\alpha_1, \alpha_2)(1+(x-\gamma)/\beta)^{\alpha_1+\alpha_2}}$	$\alpha_1, \alpha_2, \beta > 0$ $\gamma \leq x < +\infty$	$\alpha_1, \alpha_2 = \text{shape}$ $\beta = \text{scale}$ $\gamma = \text{location}$
19.	Weibull (2P)	$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right)$	$\alpha, \beta > 0,$ $0 \leq x < +\infty$	$\alpha = \text{Shape}$ $\beta = \text{Scale}$
20.	Weibull (3P)	$f(x) = \frac{\alpha}{\beta} \left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^\alpha\right)$	$\alpha, \beta > 0,$ $\gamma \leq x < +\infty$	$\alpha = \text{Shape},$ $\beta = \text{Scale}$ $\gamma = \text{Location}$

Table.3 Score wise best fitted probability distribution with parameter estimates

Sl. No.	District	Distribution		
		Name of Distribution	Total Score	Parameter estimates
1.	Bangalore®	Weibull (3P)	56	$\alpha = 1.4669, \beta = 121.5, \gamma = 5.6734$
2.	Chikmagalur	Johnson SB (4P)	55	$\gamma = 1.8322, \delta = 1.4238, \lambda = 1741.8, \theta = -58.152$
3.	Chitradurga	Log-Logistic (3P)	53	$\alpha = 2.909, \beta = 81.133, \gamma = -8.4964$
4.	Davangere	Log-Logistic (3P)	57	$\alpha = 4.8726, \beta = 122.3, \gamma = -35.569$
5.	DNK	Dagum (4P)	58	$k = 0.1018, \alpha = 9.7576, \beta = 1180.3, \gamma = 92.636$
6.	Hassan	Pearson 5 (3P)	48	$\alpha = 24.732, \beta = 9246.7, \gamma = -213.14$
7.	Kolar	Gen. Gamma (4P)	60	$k = 2.3588, \alpha = 0.415, \beta = 180.2, \gamma = 5.2117$
8.	Mandya	Gen. Gamma (4P)	56	$k = 1.2086, \alpha = 1.1747, \beta = 86.015, \gamma = 5.0922$
9.	Tumkur	Gamma (2P)	47	$\alpha = 2.3215, \beta = 42.916$
10.	Udapi	Dagum (3P)	60	$k = 0.12855, \alpha = 8.1895, \beta = 1510.7$

**Table.2** Descriptive Statistics of AMRAM for the Selected 10 districts of Karnataka

District	Min	Max	Range	Mean	SD	CV	Skewness	Kurtosis
Bangalore Rural	6.85	459.30	452.45	115.71	76.56	0.6616	1.2086	2.0812
Chikmagalur	48.20	1142.50	1094.30	355.34	209.06	0.5883	0.8689	0.6409
Chitradurga	2.50	297.50	295.00	86.85	55.63	0.6406	1.0504	0.6763
Davanagere	3.02	263.40	260.38	94.64	47.59	0.5029	0.9110	0.9248
Dakshina Kannada	92.95	1776.30	1683.30	701.23	389.20	0.5550	0.2903	-0.8236
Hassan	15.53	478.45	462.92	176.47	81.62	0.4625	0.7966	0.8971
Kolar	5.21	431.78	426.57	102.21	70.81	0.6928	1.0785	2.0022
Mandya	5.73	376.59	370.87	98.12	70.95	0.7231	1.1524	1.3027
Tumkur	2.90	300.80	297.90	99.63	65.39	0.6563	1.0010	0.5746
Udupi	79.30	2154.90	2075.60	821.49	492.59	0.5996	0.1999	-0.8767

The highest and lowest ranges of AMRAM are observed in Udupi (2075.60mm) and Chitradurga (295.00 mm) district respectively. Udupi district has received the highest mean AMRAM of 821.4 mm over the time span of 34 years while in Chitradurga district, the lowest mean AMRAM of 86.85 mm has been observed during the same period.

The SD of AMRAM in 10 different districts of Karnataka varies from 47.59 mm to 492.59 mm. Udupi district obtains the highest SD implying the large variation in AMRAM while Davanagere district has the lowest SD which indicates comparatively smaller variation in AMRAM.

The CV values indicate the irregularities in distribution of AMRAM in the region. Mandya and Hassan district possess the highest (0.72) and lowest (0.46) values of CV respectively. Therefore, it can be concluded that AMRAM in Hassan is more consistent than that of any other district whereas in Mandya district, AMRAM is most irregular. Skewness measures the asymmetry of a distribution around the mean. For all the districts, the values of skewness are positive indicating that AMRAM are positively skewed.

The maximum skewness (1.21) is obtained in Bangalore Rural district while in Dakshina Kannada district, the AMRAM distribution is least positively skewed (0.29). Kurtosis provides an idea about the flatness or peakedness of the frequency curve.

The values of kurtosis, are within a range of -0.82 to 2.08. Bangalore rural district has the highest value of kurtosis implying the possibility of a distribution having a distinct peak near to the mean with a heavy tail. Dakshina Kannada district has the smallest negative value of kurtosis which indicates that the distribution is probably characterized with a relatively flat peak near to the mean and which is too flat to be normal.

### **Distribution fitting**

The AMRAM (Average Monthly Rainfall for Active Monsoon Period) data for each of the 10 districts of Karnataka are fitted to the 20 continuous probability distributions. The three test statistics for each rainfall station data were calculated for all probability distributions. Based on Kolmogorov-Smirnov, Anderson-Darling and Chi-squared GOF test statistic values, 3 different rankings have been given to each of the distributions for all the districts. No rank is given to a distribution when the

concerned test fails to fit the data. Results of the GOF tests for all the districts are depicted in Table 1 to 3.

### **Identification of best fitted probability distribution**

As the ranks are given for each of GOF tests separately, it is difficult to identify the best fitted distribution for a district or the study region (Karnataka) as a whole. A single distribution is not ranking first in all GOFs for example, in case of Bangalore Rural district Weibull (3P) distribution ranks first based on Kolmogorov-Smirnov test. However, the same distribution ranks second and fourth based on Anderson-Darling and Chi-Squared test respectively. Hence, by employing the method of scoring as depicted in the methodology, score is given to all the distributions for each of the three GOF tests ranking separately and the final score is obtained by adding these three scores.

Based on the total score, the distribution which receives the highest score is the best fitted distribution to the corresponding district. For each district, the best fitted probability distribution with total score along with the estimates of its parameters is given in Table 3.

It is observed from Table 3 that for Log logistic (3P) and Generalized gamma distributions comes out as a most suitable for 2 districts each. Log logistic (3P) is found to be most suitable for Chitradurga and Davangere districts and Generalized gamma for Kolar and Mandya districts. While, Dagum (4P) explains best the Monthly monsoon rainfall of DNK district and Dagum (3P) for udapi district. Weibull (3P), Johnson SB(4P), Pearson 5(3P) and Gamma (2P) are found to be most fitted distribution for Bangalore(Rural), Chikkamagaluru, Hassan and Tumkur respectively.

In Conclusion, a systematic assessment procedure was applied to evaluate the performance of different probability distribution with view to identifying the best fit probability distribution for monthly rainfall data at 10 different districts of Karnataka. The data showed that the monthly minimum and maximum rainfall at monsoon period at 10 different districts ranged from 2.90 mm (Tumkur) to 2154 mm (Udapi) which is obviously indicating a huge range of fluctuation during the period of the study. It was observed that the Log-Logistic (3P) and Gen. Gamma (4P) distributions provides a good fit of the two selected districts each, Weibull (3P), Johnson SB (4P), Dagum (3P and 4P), Pearson 5 (3P), Gamma (2P) distributions explained best for 1 district each. Identifying the distribution amount of monthly rainfall data could have a wide range of applications in agriculture, hydrology, engineering design and climate research.

### **References**

- Anderson T W, Darling D A. 1952. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*. 23(2), 193-212.
- Lala I P Ray, Bora P K, Ram V, Singh A K, Singh N J, Singh R, Feroze S M. 2013. Estimation of Annual Maximum Rainfall for Central Meghalaya. *Indian Journal of Hill Farming*, 26(1), 47-51.
- Mandal K G, Padhi J, Kumar A, Ghosh S, Panda D K, Mohanty R K, Raychaudhuri M, 2015. Analyses of rainfall using probability distribution and Markov chain models for crop planning in Daspalla region in Odisha, India. *Theoretical and Applied Climatology*, 121(3-4), 517-528.
- Sanjib Ghosh, Manindra Kumar Roy, Soma Chowdhury Biswas. 2016. Determination of the Best Fit

Probability Distribution for Monthly Rainfall Data in Bangladesh. *American Journal of Mathematics and Statistics*, 6(4), 170-174.

Tariq Mahgoub Mohamed, Abbas Abd Allah Ibrahim. 2016. Fitting Probability Distributions of Annual Rainfall in Sudan. *SUST Journal of Engineering and Computer Sciences*, 17(2), 34-39.

**How to cite this article:**

Bhavyashree, S. and Banjul Bhattacharyya. 2018. Fitting Probability Distributions for Rainfall Analysis of Karnataka, India. *Int.J.Curr.Microbiol.App.Sci*. 7(03): 1498-1506.  
doi: <https://doi.org/10.20546/ijcmas.2018.703.178>