International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 23(2), November, 2023, pp 87-106
DOI: https://doi.org/10.3329/ijss.v23i2.70131

Fitting Statistical Distributions to Rainfall Data with Different Estimation Techniques: An Empirical Study from Pabna and Dinajpur Districts

Afsana Afruz, Md. Mostafizur Rahman*, Md. Rafiqul Islam, Md. Abdul Khalek and M. Sayedur Rahman

Data Mining and Environmental Research Group, Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh.

*Correspondence should be addressed to Md. Mostafizur Rahman
(Email: mostafiz_bd21@yahoo.com)

[Received March 19, 2023; Accepted October 20, 2023]

Abstract

Bangladesh is a country of diverse climatic conditions because of its rainfall and other geographical conditions which have a complex impact on economic and social aspects. The statistical distributions are used in many real life data for modeling and predicting. Knowing the real distribution of rainfall rather than depending on basic summary statistics would improve many uses of rainfall data. The aim of this paper is twofold: first, the performance of different statistical distributions such as Normal, Log-Normal, Gamma Weibull, and Gumbel are compared for modeling the monthly rainfall data from Pabna and Dinajpur districts from January 1971 to December 2015; second, the performance of the Maximum Likelihood Estimation (MLE), Quantile Matching Estimation (QME), and the Maximum Spacing Estimation are also compared for fitting these statistical distributions. The empirical study showed that Gamma distribution performs better for fitting the monthly rainfall data for both Pabna and Dinajpur districts of the three methods like Maximum Likelihood Estimation (MLE) method, the Quantile Matching Estimation (QME), and the Maximum Spacing Estimation (MSE) method. The Normal distribution performs worse of these study areas. By the comparison of these three methods we found that Maximum Likelihood Estimation (MLE) gives better results. This study provides the actual distribution of rainfall data of these study areas.

Keywords and Phrases: Statistical distributions, Rainfall data, Estimation techniques and Empirical study.

AMS Classification: 62P12, 62M10.

1. Introduction

Bangladesh is an agrarian country. About 48.1% of Bangladesh's labor force is directly employed in agriculture, which is the foundation of the country's economy. One of the most essential natural resources is water and it is crucial to the sustainability of human life in all respects, including household life, industry, and agriculture. Statistical distributions are utilized in a wide range of research domains for a variety of goals. These are frequently serving as helpful tools for describing both natural and social events, offering appropriate models that can aid in solving practical issues, such as those involving the forecast of an important occurrence.

Information on rainfall is necessary for a variety of theoretical and practical reasons. In recent years, Bangladesh's climate has seen tremendous alteration (Shahid, 2009). It has been noted that Bangladesh's annual rainfall and daily mean temperature have grown by 5.2 millimeters per year and 0.9 degrees Celsius per decade, respectively (Shahid, 2010; Shahid et al., 2016 and Kirby et al., 2016). By 2030, rainfall in Bangladesh would increase by 5% to 6%, and the temperature will rise by 1.9 °C (IPCC, 2007). The probability of severe occurrences can be significantly altered by very small changes in the mean and standard deviation values (Easterling et al., 2000, Rodrigo, 2002, Chiew, 2006, Shahid, 2011). Climate change has become a top concern for scientists and mankind as a whole in light of recent patterns in the climate and future forecasts for the 21st century. Research on the fluctuations in rainfall and temperature is receiving more and more attention across the world. The most crucial parts of any climate change are rainfall and temperature since they have the greatest impact on ecosystems and society reactions. Rainfall trends over the past three decades (1991-2018) show a somewhat distinct pattern. This trend is noteworthy since monsoon seasons account for 70% of the yearly rainfall (Ahasan et al., 2010). Increases in extreme climate events, such as extended heat spells and days with heavy rain, have more detrimental effects on human civilization and the environment than do climate changes. Numerous studies have shown how a single catastrophic precipitation event may quickly undo long-term achievements of human culture (Zong and Chen, 2000).

Natural water shortages are most prevalent in arid and semiarid regions, which by definition have limited water supplies. Prior knowledge about the statistical distribution helps to improve the many uses of rainfall data. Many authors used different statistical distributions for various geographical locations. Sen and Eljadid (1999) examined the performance of Gamma distribution for modeling rainfall data of Libya and found that Gamma distribution is the suitable distribution for monthly rainfall data of Libya. Al-Mansory (2005) investigated the performance of Normal, Log-Normal, Log-Normal type III, Pearson type III, Log-Pearson type III and Gumbel distributions for the monthly rainfall data of Basra station of Iraq and found that Pearson type III and Gumbel distributions were the most appropriate distribution of this study area. Suhaila and Jemain (2007) proposed mixed Gamma, mixed Weibull and mixed Exponential and tested together with their single distributions to identify the optimal model for daily rainfall amount in several rain gauge stations in Peninsular Malaysia and found that mixed Weibull model showed better performance. Olumide et al. (2013) fitted a range of probability distribution models to various rainfall and runoff to assess which model was best appropriate for the prediction of their values at the Tagwai dam site in Minna, Niger State, Nigeria. Alghazali and Alawadi (2014) fitted three statistical distributions to monthly rainfall measurements from thirteen locations in Iraq and found that Gamma distribution was most appropriate distributions for five stations. Mohamed and Ibrahim (2016) examined Sudan's annual rainfall and fitted Normal, Log-Normal, Gamma, Weibull, and Exponential distributions and found that Normal and Gamma distribution performs better. Maliva and Missimer (2012) studied different statistical distributions in case of rainfall data from dry and semi-arid locations. Sreedhar (2019) studied to determine the best fit of probability distribution in the case of frequency of daily rainfall in past 35 years (1982-2017) from 24 districts of the state of Andhra Pradesh, India and found Double Exponential distribution showed best fit.

From the previous study it is confirmed that different distribution performs better for different weather stations. So, it is always interesting for researchers to find out the best statistical distribution for particular area. And it is also interesting for researchers to find out best estimation technique for estimating these distributions. The Pabna and Dinajpur district of Bangladesh will be taken into account for the time series meteorological data. The results of this study will be used as a foundation for the higher authorities' decisions about the governmental purchase of food grains from farmers whenever the harvest is favorable. The rest of this paper is organized as follows: section two presents the methodology, section three describes the characteristics of study area, section four presents the result and discussion and section five presents the conclusion.

2. Methodology

2.1 Probability Distribution

The probability density function of Normal, Log-Normal, Gamma, Weibull and Gumbel distribution are given below.

2.1.1 The Normal Distribution

A random variable X is said to have a normal distribution with parameters mean μ and variance σ^2, if its density function is given by the probability law:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2},\ -\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma^2 > 0$$

where:

- x = value of the variable or data being examined and f(x) the probability function
- μ = the mean
- σ = the standard deviation

2.1.2 The Log-Normal Distribution

If Y is a normal variate with mean μ and variance σ^2 and if Y = $\log_e$ X, then X is said to have a lognormal variate if its probability density function is defined by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\log x-\mu}{\sigma})^2}\ ,x > 0,\ \sigma > 0$$

Like normal distribution, μ and σ^2 are the two parameters of the distribution (Roy, 2004).

2.1.3 The Gamma Distribution

A continuous random variable x is said to have a generalized gamma distribution with parameters α and β if its probability density function is given by

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma\alpha} & ; 0 < x < \infty \\ 0 \quad otherwise & \end{cases}$$

Where α > 0 & β > 0 (Roy, 2004, Husak, et.al 2006).

2.1.4 The Weibull Distribution

A continuous random variable X is said to have a Weibull distribution if its pdf is given by

$f(x;\alpha,\beta) = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}; 0 < x < \infty$, where $\alpha > 0, and\ \beta > 0$ (Roy, 2004).

2.1.5 The Gumbel Distribution

A continuous random variable X is said to have a Gumbel distribution if its cumulative distribution function is defined as

$F(x;\alpha,\beta) = exp\left(-e^{-\frac{(x-\alpha)}{\beta}}\right)$, where -∞<α<∞ and β>0 are the two parameters of the distribution (Roy, 2004).

2.2 Estimation Methods

2.2.1 Maximum Likelihood Estimation (MLE)

The maximum probability principle is a quite simple concept. We start with a sample of random variables, $X = (X_1, X_2, \dots, X_n)$ selected based on one of a family of probabilities,P_θ. In addition, the density function for the data when is the real state of nature will be denoted as $f(x|\theta)$, where $x = (x_1, x_2, \dots, x_n)$.

Then, the principle of maximum likelihood yields a choice of the estimator $\hat{\theta}$ as the value for the parameter that makes the observed data most probable. The likelihood function is the density function. It is considered as a function of θ. $L(\theta|x) = f(x|\theta),\ \theta \in \Theta$. The maximum likelihood estimator (MLE), $\hat{\theta} = \underset{\theta}{Arg \max} L(\theta|x)$.

There is a crucial characteristic of this class of estimators. If $\hat{\theta}(x)$ is a maximum likelihood estimate for θ, then $g(\hat{\theta}(x))$ is a maximum likelihood estimate for $g(\theta)$. For example, if θ is a parameter for the variance and $\hat{\theta}$ is the maximum likelihood estimator, then $\sqrt{\hat{\theta}}$ is the maximum likelihood estimator for the standard deviation. Unbiased estimators lack the flexibility in estimating criteria that is shown here. Typically, maximizing the score function, $lnL(\theta|x)$, the logarithm of the likelihood, will be easier. We will look at numerous examples of the probability function because it is rare to have the parameter values be the variable of interest (Watkins, 2011). We shall see that the maximum likelihood estimators offer a number of advantageous qualities, particularly for large samples. The likelihood can have several local maxima, especially for high dimensional data. As a result, locating the global maximum can be quite difficult computationally.

2.2.2 Quantile Matching Estimation (QME)

Finding a linear combination of a collection of random variables that fits the distribution of a target random variable is made possible by the handy approach known as quantiles matching estimation (QME). It may be sensitive to outlier observations of the target random variable since it is based on ordinary least-squares (OLS). Thus, a general matching quantiles M-estimation technique is put forth that is immune to target random variable outlier data. Given that there may be many variables in most applications, a "sparse" form is highly preferred. In order to choose useful variables, the QMME is paired with the adaptive Lasso penalty. To calculate QMME, an iterative technique based on M-estimation is created. Similar to the QME, the suggested matching quantiles M-estimate is reliable. There are several simulations available that show the new method's effective finite-sample performance. Additionally, a real-life case study is offered for illustration. Sgouropoulos et al. (2015) originally placed out the quantiles matching estimation (QME) approach as a solution to the issue of estimating representative portfolios for backtesting counterparty credit risks.

2.2.3 Maximum Spacing Estimation (MSE)

Let $\{x_{(1)},\dots, x_{(n)}\}$ be the corresponding ordered sample, which is the outcome of sorting all observations from smallest to largest, given a iid random sample $\{x_1,\dots, x_n\}$ of size n from a univariate distribution with continuous cumulative distribution function $F(x;\theta_0)$, where θ_0 is an unknown parameter to be estimated. Also indicate $x_{(0)} = -\infty$ and $x_{(n+1)} = +\infty$ for convenience.

As the "gaps" between the distribution function values at adjacent ordered points, define the spacings:

$$D_i(\theta) = F(x_i;\theta) - F(x_{i-1};\theta), i = 1, \dots, n+1$$

The value that maximizes the logarithm of the geometric mean of the sample spacings is then described as the maximum spacing estimator of θ_0:

$$\hat{\theta} = arg \max_{\theta \in \Theta} S_n(\theta), where\ S_n(\theta) = {}^{ln^{n+1}}\sqrt{D_1 D_2 \dots D_{n+1}} = \frac{1}{n+1} \sum_{i=1}^{n+1} ln\, D_i(\theta)$$

The maximum must exist at least in the supremum sense since function $S_n(\theta)$ is constrained from above by -ln(n+1) due to the inequality of arithmetic and geometric means. Be note that various authors define $S_n(\theta)$ slightly differently. Ranneby (1984) multiplies each D_i by a factor of (n+1), but Cheng and Stephens (1989) remove the 1/n+1 factor before the total and add the "-" sign to convert maximizing to minimization. The alterations have no effect on where the maximum of the function S_n is located since they are constants with regard to θ.

2.3 Model Evaluation Statistics

For researchers, choosing an acceptable model is a difficult task. The following test statistics are well-known for vetting goodness-of-fit tests, according to researchers. These include the Log-Likelihood (LL) statistic, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). The table for calculating the goodness of fit for several empirical distributions was first introduced by Smirnov (Smirnov, 1984).

3. Study Area

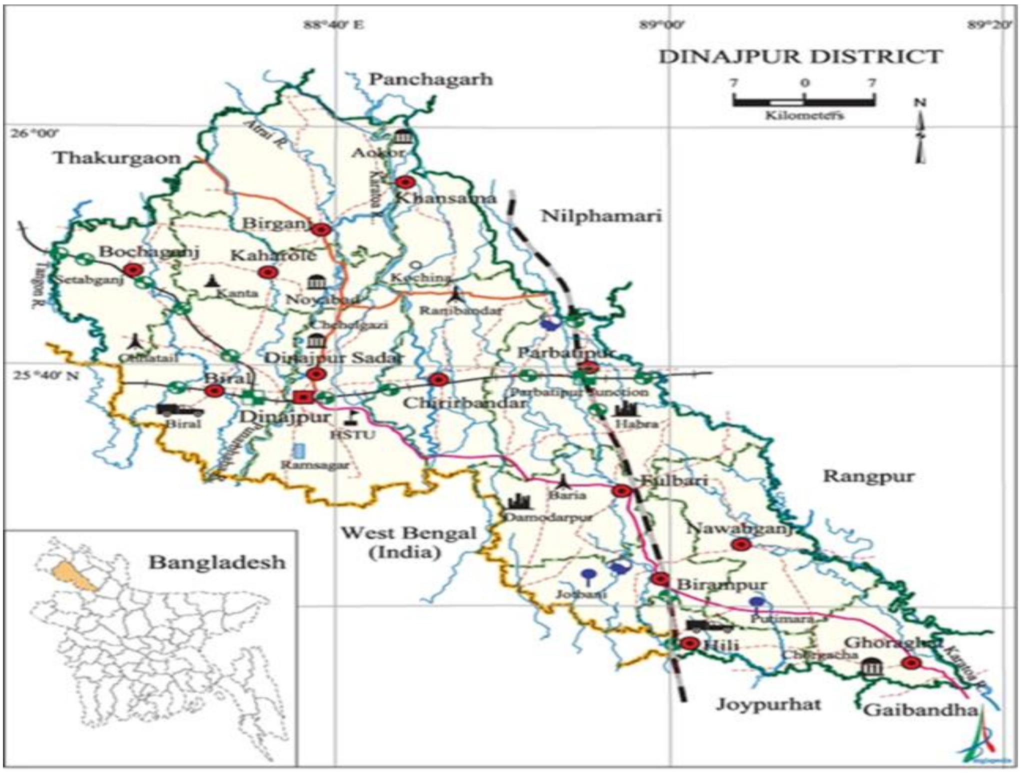
3.1 Pabna District

The total area of the Pabna district is 2371.50 sq km, and its latitude and longitude range from 23°48' to 24°21' north and 89°00' to 89°44' east, respectively. Pabna is the southern border part of Bangladesh and it is treated as the central district of Bangladesh. It is a region with significant economic impact. The Padma river, in the south divides Rajbari district and Kushtia district from Sirajganj district, which is to the northeast. On the east, the Jamuna river separates it from the Manikgonj district, while on the north and west, it has a border with the Natore district. Its annual rainfall is 1872 millimeters, and its average maximum and lowest temperatures are 36.8 °C and 9.6 °C, respectively. Due to the flood plains of the Ganges, Karatoya, Jamuna, and Barind Tracts, the district's soil is characteristically separated into four sections. The Ganges, Ichamati, Gumani, Boral, and Hurasagar rivers are some of its principal waterways. (Wikipedia, Pabna District). The land of this study area consists mainly of flat plains which produce a large variety of crops and vegetables like rice, wheat, pulses, potatoes, carrots, onions, sugarcane etc. The administrative map of the Pabna district is given in Map 1. The main occupation in Pabna District is agriculture. Agriculture accounts for 34% of all jobs in this region, followed by agricultural farm workers (22.77%), wage laborers (4.46%), transportation (2.18%), weaving (2.85%), commerce (13.27%), service (7.26%), and other jobs (13.21%).

Map 1: Map of Pabna District (Source: Banglapedia, Pabna District)

3.2 Dinajpur District

Dinajpur district is located in northern Bangladesh's Rangpur Division (Former Rajshahi division). Among these sixteen northern districts, Dinajpur is the biggest one. It had a total population of 2,990,128 (Census, 2011). Total area 3437.98 sq. km, located between 25°10' and 26°04' north latitudes and in between 88°23' and 89°18' east longitudes. It is bordered to the north by the districts of Thakurgaon and Panchagarh, to the south by Gaibandha and Joypurhat, to the east by Nilphamari and Rangpur, and to the west by the Indian state of West Bengal. The Dhepa, Punarbhaba, and Atrai rivers make up the majority of the district's waterways. A hot, wet, and humid tropical climate prevails in Dinajpur. Dinajpur is classified as having a tropical wet and dry climate by the Köppen system. With an annual average temperature of 25 °C (77 °F) and monthly means that range from 18 °C (64 °F) in January to 29 °C (84 °F) in August, the district has a distinct monsoonal season. Dinajpur typically receives about 158.53 millimeters (6.24 inches) of precipitation and has 172.24 rainy days (47.19% of the time) annually (Weather and Climate). Agriculture accounts for 63.90% of the district of Dinajpur's income, followed by non-agricultural laborers at 3.29%, industry at 0.90%, commerce at 12.89%, transport and communication at 3.35%, service at 6.58%, construction at 3.37%, religious service at 0.17%, rent and remittance at 0.23%, and others at 5.32% (Banglapedia, Dinajpur District). The geographical location of the Dinajpur district is given in Map 2.



Map 2: Map of Dinajpur district (Source: Banglapedia, Dinajpur district)

3.3 Data Source

Two metrological stations Pabna and Dinajpur districts metrological were chosen for empirical study. The data used in this thesis was collected from the Bangladesh Meteorological Department (BMD). The daily data covered the period from January 1971 to December 2015 which created a total number of 16016 observations from the Pabna and Dinajpur districts of the Rajshahi and Rangpur divisions respectively. We converted this daily rainfall data into monthly rainfall data using Microsoft Excel.

3.4 Data Smoothing

In environmental studies missing data is common. The World Meteorological Organization (WMO) states that although the data collected for each station contains several missing values, it is still possible to estimate them successfully because less than 10% of the missing data must be estimated using the Statistical Package for Social Sciences (SPSS) program. These missing numbers were random, and some years also had continuous missing data for one to many months. We used SPSS software to fill up the missing value and the 0 value is replaced by constant value 0.001. Then we used R for modeling.

4 Empirical Studies

4.1 Characteristics of Data

To find out best fitting distribution from a set of predefined distributions we need some measuring characteristics. This may be done by the knowledge of stochastic processes prevailing the modeled variable, or by the observation of its empirical distribution. Skewed distributions are measured using the asymmetry levels. The distribution is said to be right or favorably skewed if the tail is longer on the right side. A distribution is considered to be left or negatively skewed if the tail of the distribution is longer on the left side. Kurtosis is a measure of the "tailedness" of a distribution. Kurtosis can be classified as leptokurtic, platykurtic, or mesokurtic. When kurtosis is equal to 3, the distribution is mesokurtic. When kurtosis is less than 3, the distribution is platykurtic. When kurtosis is greater than 3, the distribution is leptokurtic. We present the summary statistics such as mean, standard deviation, minimum, maximum, skewness and kurtosis value of these two districts. The summary statistics is reported in Table 1.

Table 1: Summary statistics for Pabna and Dinajpur districts

District	Min.	Mean	Standard deviation	First quartile	Max.	Third quartile	Skewness	Kurtosis
Pabna	0.001	6.881	6.945	0.824	31.774	11.250	1.144	4.004
Dinajpur	0.001	6.238	5.012	1.433	19.590	9.796	0.581	3.043

The calculated result from Table 1 indicated that the average monthly rainfall for Pabna and Dinajpur districts are 6.881 and 6.238 mm respectively. The rainfall amount was maximum in Pabna district. The standard deviation for Pabna district was 6.945 and for Dianajpur was 5.012. Both of these rainfall series produced positive skewness and kurtosis was greater than 3 which indicated three common right skewed distributions Weibull, Gamma and Log Normal may gave better result. Besides these other distributions such as Normal and Gumbel distribution also showed better fitting result for rainfall data (Maliva and Missimer, 2012).

4.2 Empirical results

In this study, we consider rainfall data from Pabna and Dinajpur districts from Rajshahi and Rangpur divisions, Bangladesh. Papalexious (2012) suggested three steps for selecting a probability distribution for rainfall data, including (1) selecting some parametric families of distributions a priori, (2) estimating the parameters with a suitable fitting method, and (3) identifying the model that performs the best based on some model evaluation criteria. In this study, the distributions Normal, Log-Normal, Gamma, Weibull, and Gumbel are taken into consideration for analysis. In the study we also consider different estimation techniques such as Maximum Likelihood Estimation (MLE), Quantile Matching Estimation (QME) and Maximum Spacing Estimation (MSE) methods for rainfall data of these study areas.

4.2.1 Parameter estimation with MLE

To assess the amount of water availability, the study of the distribution of rainfall in time and space is very important for the development of a country. Applications of rainfall data were improved by the prior knowledge of actual distribution of rainfall. The parameter estimation result of different statistical distributions with Maximum Likelihood Estimation (MLE) method for Pabna and Dinajpur districts are given in Table 2.

The estimated result from Table 2 indicated that all the parameters from different distributions with Maximum Likelihood Estimation method were significant at 5 % significance level. The model evaluation criteria AIC and BIC indicate that the Gamma distribution provide lowest value whereas Normal distribution provide highest value in case of both districts. On the other hand Gamma distribution shows highest log likelihood value and Normal distribution showed lowest value. The estimation result of goodness of fit test statistic Anderson Darling test statistic (AD) and Kolmogorov-Smirnov (KS) test statistic shows that Gamma and Weibull distribution

Table 2: Parameter estimation results with Maximum Likelihood Estimation method

Pabna					**Model evaluation statistic**					
Model	Parameter estimation				Model evaluation			Test statistic		
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	AIC	BIC	-LL	AD	CM	KS
Normal	6.881			6.936	4017.961	4026.748	2006.981	23.223	3.668	0.161
Log-Normal	1.019			1.715	3587.637	3596.44	1791.818	16.603	2.875	0.141
Gamma		0.668	0.097		3431.212	3439.999	1713.606	6.189	1.340	0.050
Weibull		0.782	6.088		3448.68	3457.467	1722.34	8.278	1.658	0.101
Gumbel		3.823	4.825		3840.227	3849.014	1918.113	20.025	3.095	0.150
Dinajpur					**Model evaluation statistic**					
Model	Parameter estimation				Model evaluation			Test statistic		
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	AIC	BIC	-LL	AD	CM	KS
Normal	6.238			5.008	3627.982	3636.769	1811.991	17.862	3.072	0.121
Log-Normal	1.279			1.293	3538.042	3546.829	1767.021	12.283	1.864	0.111
Gamma		1.041	0.167		3384.667	3394.654	1690.433	8.558	1.405	0.100
Weibull		1.076	6.40		3388.867	3397.654	1692.433	9.066	1.466	0.109
Gumbel		3.894	3.909		3547.896	3556.683	1771.948	13.952	2.126	0.114

present the smallest value of test statistics. For comparing models the Gamma distribution produces the lowest value for the Anderson and Darling test statistic and the Normal distribution produces the highest value. The Cramer-von Mises statistic (CM) also shows a similar finding. The Kolmogorov-Smirnov (K-S) test statistics also indicates that Gamma and Weibull distribution present the significant result. Cullen and Frey (1999) provide four classes of classical goodness of fit plots and these are density plot which present the density function of the fitted distribution along with histogram of the empirical distribution, a CDF plot of both the empirical distribution and the fitted distribution, a Q-Q plot which represent the empirical quantiles against the theoretical quantiles and a P-P plot representing the empirical distribution function evaluated at each data point against the fitted distribution function. The fitted goodness of fit plots of Normal, Log Normal, Gamma, Weibull and Gumbel distribution with Maximum Likelihood Estimation method in case of Pabna and Dinajpur districts are given in Figure 1. In order to reduce the paper page we present only Q-Q and P-P plot.

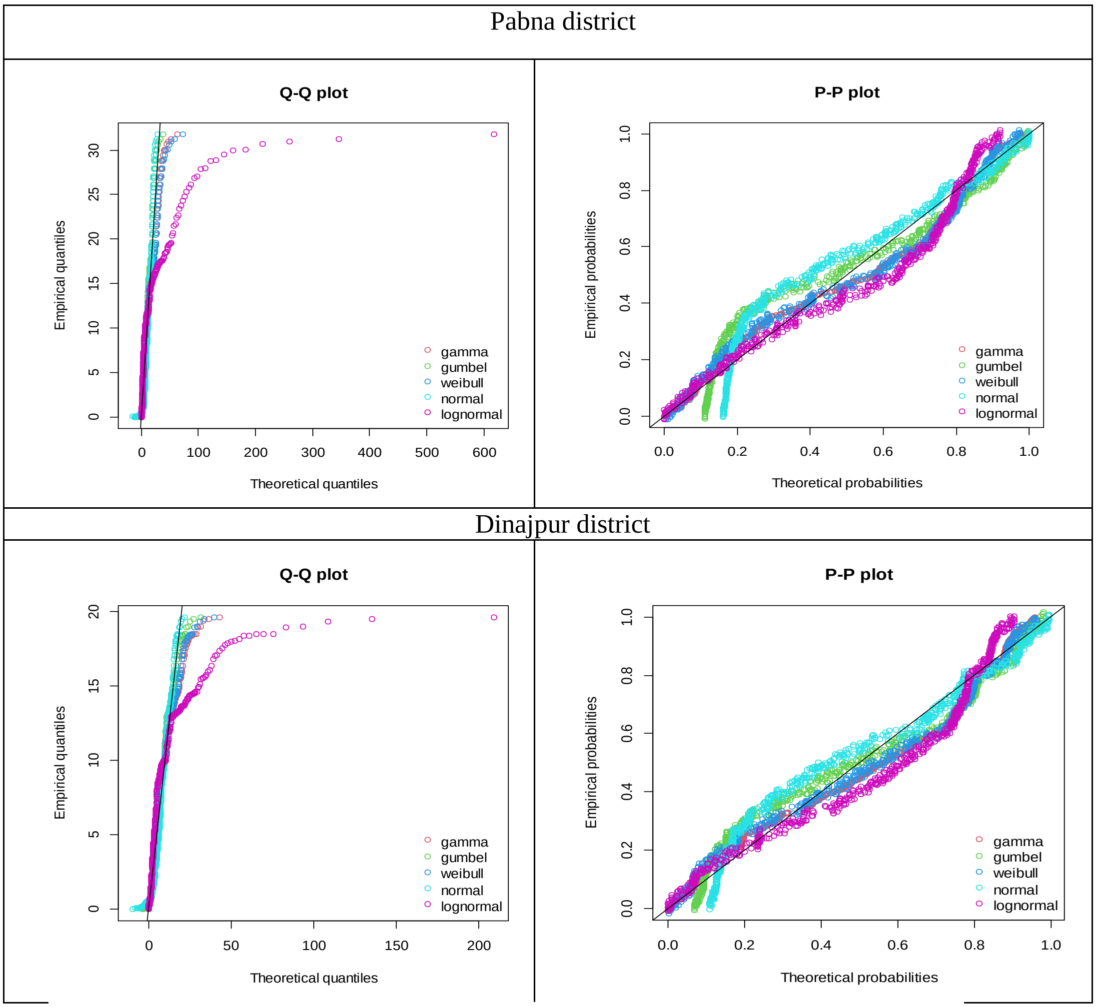


Figure 1: Fitted goodness of fit plots with Maximum Likelihood Estimation method

The density plot and the CDF plot may be considered the basic classical goodness-of-fit plots. The two other plots are complementary and can be very informative in some cases. The Q-Q plot emphasizes the lack of fit at the distribution tails while the P-P plot emphasizes the lack of fit at the distribution center. Figure 1 showed that Gamma, Weibull, and Gumbel distributions fitted well compare to Normal and Log-Normal distributions. The Q-Q and P-P plot also indicates that Gamma and Weibull distribution provided better results compared to Normal and Log Normal distribution in the case of the both districts with the Maximum Likelihood Estimation method.

4.2.2 Parameter estimation with QME

The parameter estimation result of different statistical distributions with Quantile Matching Estimation (QME) method is given in Table 3.

Table 3: Parameter estimation results with Quantile Matching Estimation method

Pabna					**Model evaluation statistic**					
Model	Parameter estimation				Model evaluation			Test statistic		
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	AIC	BIC	-LL	AD	CM	KS
Normal	6.069			7.728	4039.068	4047.856	2017.534	21.623	2.999	0.218
Log-Normal	1.114			1.936	3601.794	3610.581	1798.897	15.042	2.238	0.126
Gamma		0.486	0.057		3483.368	3492.155	1739.684	8.576	1.014	0.089
Weibull		0.599	6.562		3512.728	3521.515	1754.364	10.208	1.323	0.102
Gumbel		2.999	6.677		3942.072	3950.859	1969.036	19.033	2.483	0.209
Dinajpur					**Model evaluation statistic**					
Model	Parameter estimation				Model evaluation			Test statistic		
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	AIC	BIC	-LL	AD	CM	KS
Normal	5.615			6.199	3681.47	3690.257	1838.735	16.495	2.591	0.183
Log-Normal	1.321			1.424	3549.111	3557.898	1772.555	14.754	1.599	0.123
Gamma		0.742	0.104		3436.231	3445.018	1716.116	9.857	1.304	0.089
Weibull		0.818	6.572		3448.578	3457.365	1722.289	10.6	1.456	0.094
Gumbel		3.171	5.318		3649.758	3658.546	1822.879	14.995	1.811	0.164

The estimated result from Table 3 indicated that all the parameters from different distributions with the Quantile Matching Estimation method are significant at a 5% significance level. The AIC and BIC value of the Gamma distribution is minimum and the Normal distribution is maximum. The log-likelihood value of the Gamma distribution is maximum and the Normal distribution is minimum. The estimation result of the goodness of fit test statistic Anderson Darling test statistic (AD) and Kolmogorov-Smirnov (KS) test statistic shows that Gamma and Weibull distribution present the smallest of test statistics. The Cramer-von Mises statistic (CM) also shows a similar conclusion. The Kolmogorov-Smirnov (K-S) test statistics also indicate that Gamma and Weibull distribution present a significant result. So, Table 3 indicates that Gamma distribution provides better fitting performance and Normal distribution provides worse fitting performance with the Quantile Matching Estimation method in the case of the both districts. The fitted goodness of fit plots of Normal, Log Normal, Gamma, Weibull, and Gumbel distribution with the Quantile Matching Estimation method in the case of the Pabna and Dinajpur districts are given in Figure 2.

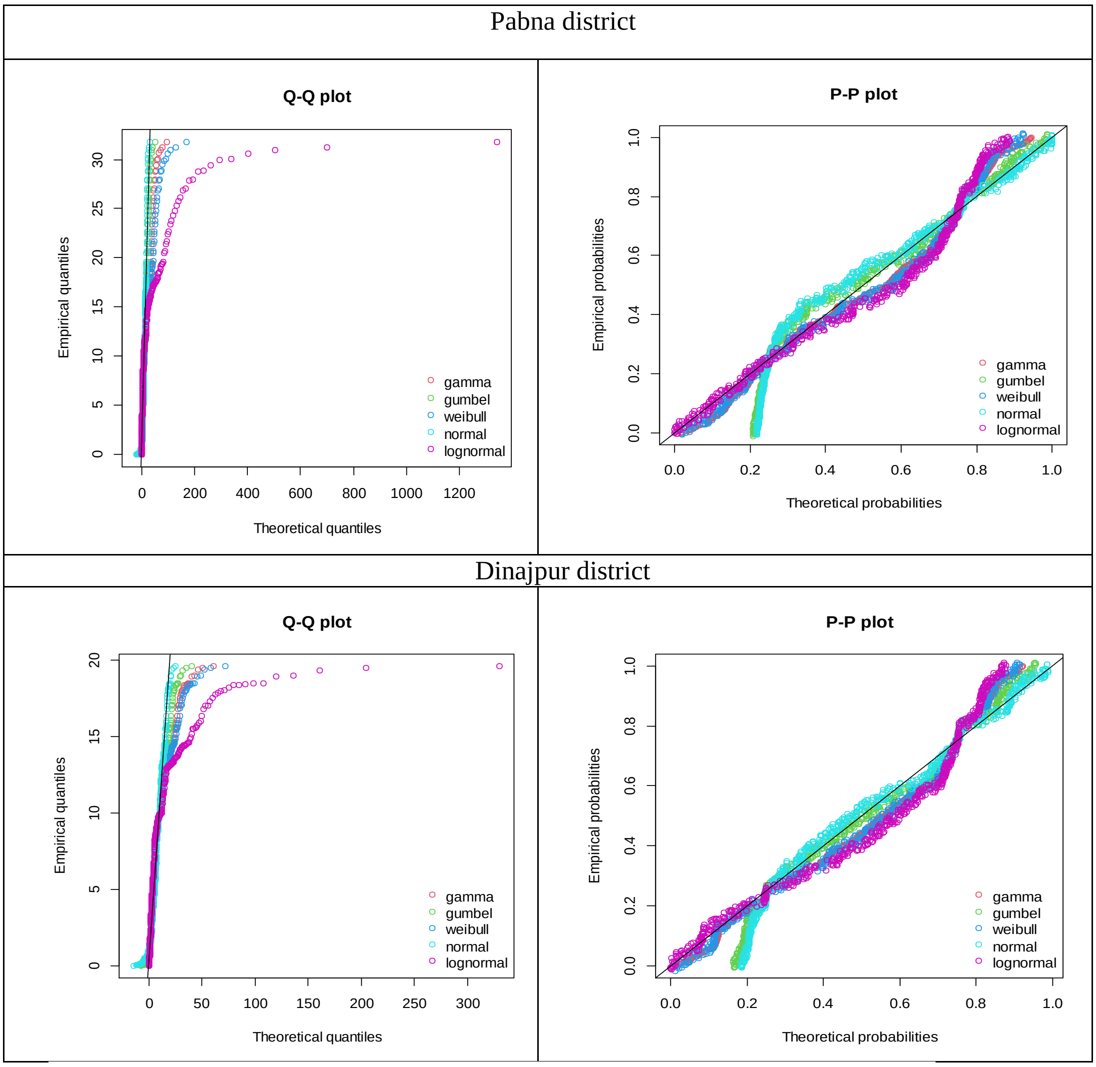


Figure 2: Fitted goodness of fit plots with Quantile Matching Estimation method

Figure 2 present that Gamma, Weibull, Normal, and Gumbel distributions fitted well compared to the Log Normal distribution. The Q-Q and P-P plot also indicates that Gamma and Weibull and Normal distribution provide better results compared to Log Normal distribution in the case of Pabna district with Quantile Matching Estimation method. So, Figure 2 also confirms the better fitting performance of the Gamma distribution.

4.2.3 Parameter estimation with MSE

The parameter estimation result of different statistical distributions with the Maximum Spacing Estimation (MSE) method is given in Table 4.

Table 4: Parameter estimation results with Maximum Spacing Estimation method

Pabna					**Model evaluation statistic**					
Model	Parameter estimation				Model evaluation			Test statistic		
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	AIC	BIC	-LL	AD	CM	KS
Normal	7.742				4027.221	4036.008	2011.61	33.729	6.058	0.190
Log-Normal	1.297				3613.783	3622.571	1804.892	18.57	2.414	0.131
Gamma		0.788	1.102		3449.622	3458.409	1722.811	10.452	1.886	0.101
Weibull		0.889	7.392		3471.676	3480.463	1733.838	16.456	2.115	0.117
Gumbel		4.586	5.137		3853.467	3862.255	1924.734	29.099	4.789	0.189
Dinajpur					**Model evaluation statistic**					
Model	Parameter estimation				Model evaluation			Test statistic		
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	AIC	BIC	-LL	AD	CM	KS
Normal	6.675			5.039	3632.532	3641.319	1814.266	18.763	3.223	0.134
Log-Normal	1.389			1.284	3542.557	3551.344	1769.279	15.235	2.140	0.109
Gamma		1.113	0.166		3391.455	3400.242	1693.727	8.656	1.136	0.096
Weibull		1.143	6.977		3393.721	3402.509	1694.861	11.121	1.392	0.103
Gumbel		4.302	4.052		3353.695	3562.482	1774.848	15.729	2.323	0.132

The estimated result from Table 4 indicates that all the parameters from Normal, Log Normal, Gamma, Weibull, and Gumbel distribution with the Maximum Spacing Estimation method are significant. The Gamma distribution shows the highest value of log-likelihood value whereas the Normal distribution provides the lowest log-likelihood value. The AIC and BIC value of the Gamma distribution is minimum and the Normal distribution is maximum. The estimation result of the goodness of fit test statistic Anderson Darling test statistic (AD) and Kolmogorov-Smirnov (KS) test statistic shows that Gamma distribution presents the smallest value of test statistics. The Cramer-von Mises statistic (CM) also shows a similar conclusion. The Kolmogorov-Smirnov (K-S) test statistics also indicate that Gamma presents a significant result. So, Table 4 indicates that Gamma distribution provides the best fitting performance with Maximum Spacing Estimation in the case of both Pabna and Dinajpur districts. The fitted goodness of fit plots of Normal, Log Normal, Gamma, Weibull, and Gumbel distribution with the Maximum Spacing Estimation method in the case of the Pabna district is given in Figure 3.

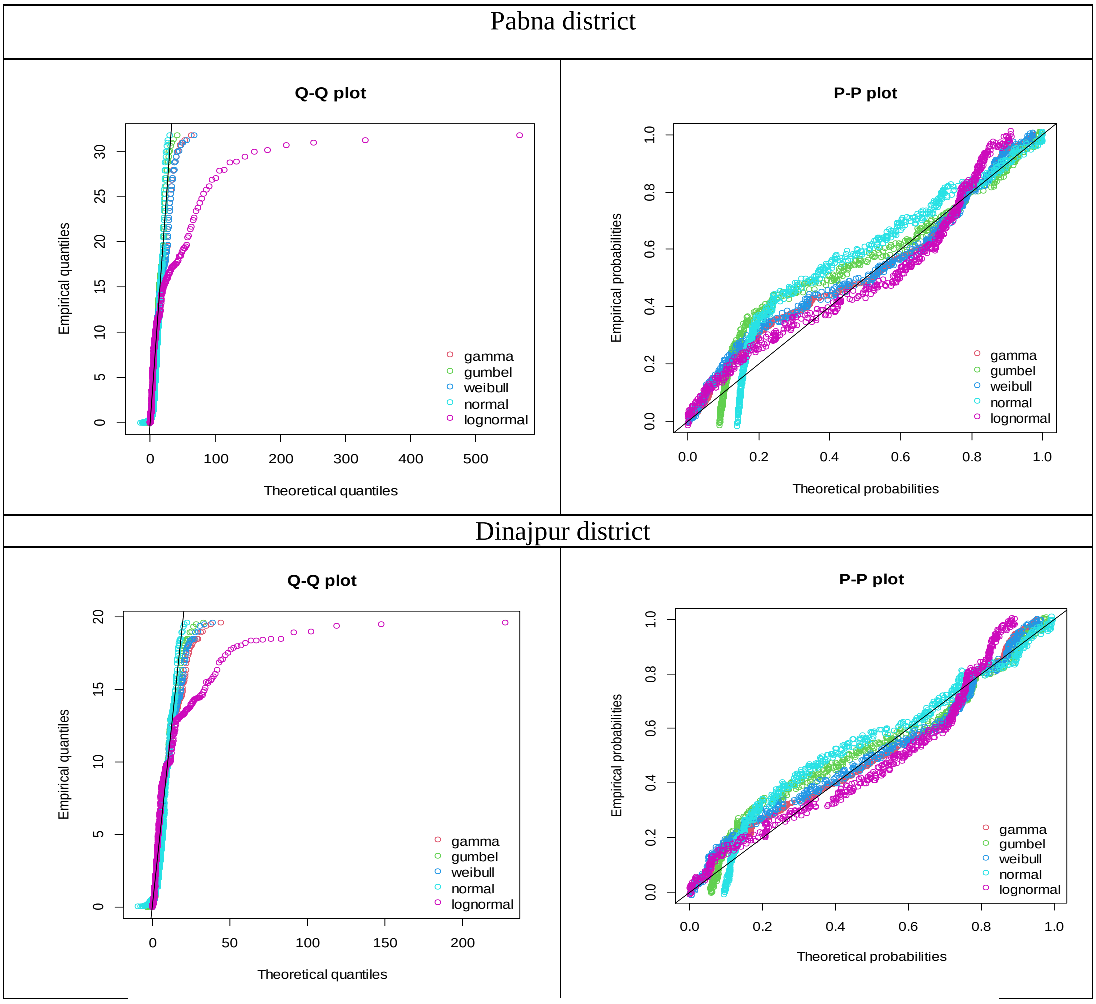


Figure 3: Fitted goodness of fit plots with Maximum Spacing Estimation method

Figure 3 presents that all of the distribution shows very close results to the fitted line except the Log-Normal distribution. Among these distributions, the Gamma distribution provides better fitting performance. So, Figure 3 also confirms that the best distribution with the Maximum Spacing Estimation method is the Gamma distribution in the case of both districts during this study period.

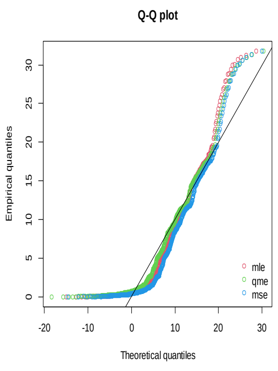
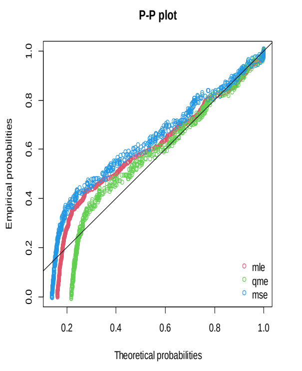
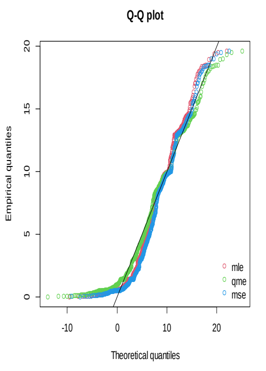
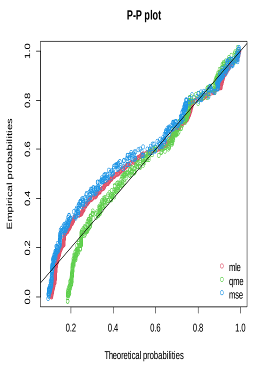
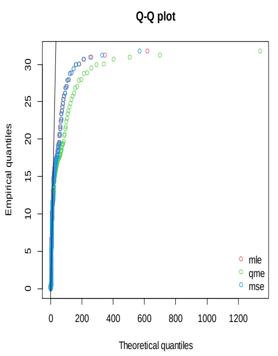
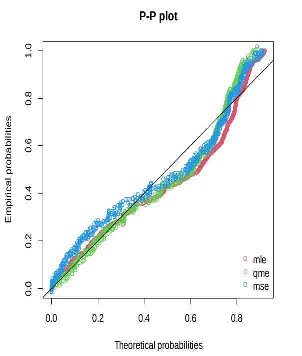
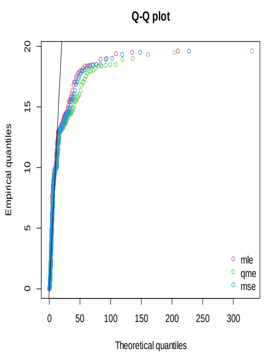
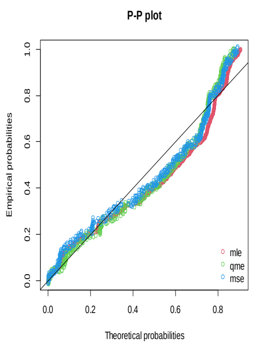
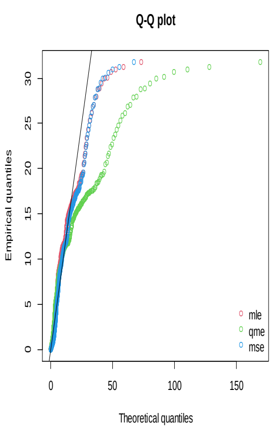
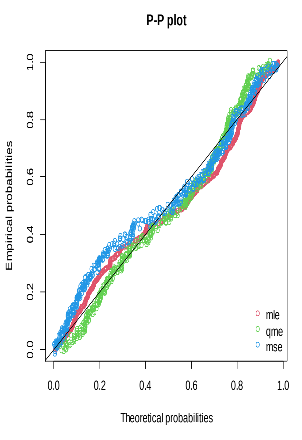
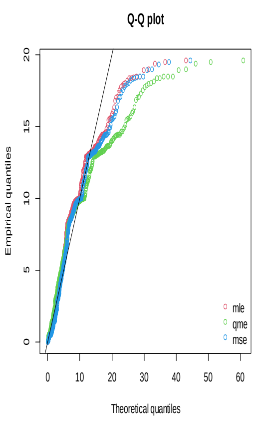
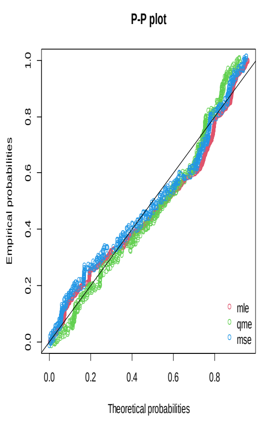
4.3 Estimation methods comparison

In statistics, fitting different distributions to data is a very common task. This requires judgment and expertise and generally needs an iterative process of distribution choice, parameter estimation, and quality of fit assessment. Usually, the most commonly used estimation method is the Maximum Likelihood Estimation method (MLE). In this study, we also consider another two methods Quantile Matching Estimation (QME) and Maximum Spacing Estimation (MSE) estimation methods. Here we compare the estimation performance of these different estimation methods in these study areas. The estimation results with different estimation methods with different statistical distributions of these study areas are given in Table 5.

Table 5: Estimation methods evaluation criteria for different distributions

Pabna	**Estimation methods comparison with different distributions in case of Pabna**								
Method	**MLE**			**QME**			**MSE**		
Distribution	**AIC**	**BIC**	**-LL**	**AIC**	**BIC**	**-LL**	**AIC**	**BIC**	**-LL**
Normal	4017.96	4026.74	2006.98	4039.06	4047.85	2017.53	4027.22	4036.00	2011.61
Log-Normal	3587.63	3596.44	1791.81	3601.79	3610.58	1798.89	3613.78	3622.57	1804.89
Gamma	3431.21	3439.99	1713.60	3483.36	3492.15	1739.68	3449.62	3458.40	1722.81
Weibull	3448.68	3457.46	1722.34	3512.72	3521.51	1754.36	3471.67	3480.46	1733.83
Gumbel	3840.22	3849.01	1918.11	3942.07	3950.85	1969.03	3853.46	3862.25	1924.73
Dinajpur	**Estimation methods comparison with different distributions in case of Dinajpur**								
Normal	3627.98	3636.76	1811.99	3681.47	3690.25	1838.73	3632.53	3641.31	1814.26
Log-Normal	3538.02	3546.82	1767.01	3549.11	3557.89	1772.55	3542.55	3551.34	1769.27
Gamma	3384.67	3397.65	1692.33	3436.31	3445.01	1716.11	3391.45	3400.24	1693.72
Weibull	3388.67	3397.65	1692.33	3448.78	3457.36	1722.28	3393.72	3402.50	1694.86
Gumbel	3547.96	3556.68	1771.48	3649.58	3658.54	1822.87	3353.69	3562.48	1774.84

Since earlier we already presented parameter estimation results, model evaluation criteria such as AIC, BIC, LL, and goodness of fit test statistics AD, CM, and K-S test statistics so, here for comparison estimation method we only present the AIC, BIC, and LL values of different models and estimation methods. The estimated result from Table 5 indicates that for Normal distribution the AIC value is minimum in the case of MLE and Maximum in the case of QME. In the case of Log-Normal distribution, the AIC value is minimum in MLE whereas is maximum in MSE. For Gamma distribution, the MLE method provides minimum AIC and QME provides maximum AIC value. In the case of Weibull distribution, the MLE method gives minimum AIC whereas QME gives maximum AIC and finally for Gumbel distribution the MLE method shows the lowest AIC value and the QME method shows the highest AIC value. The BIC values also present a similar pattern for all of these distributions and estimation methods. The log-likelihood value from Normal, Gamma, Weibull, and Gumbel with MLE provides the highest value whereas QME provides the lowest value. In the case of Log Normal distribution, the log-likelihood value is minimum in the MLE method but maximum in the MSE method. So, from this Table 5, we conclude that all of the distribution with different estimation methods confirms that the MLE method provides better estimation performance. The fitted goodness of fit plots of these distributions for Pabna and Dinajpur districts are given in Figure 4. To reduce the page here we only present Q-Q and P-P plot for both districts.

<table>
<tr><th colspan="2">Pabna</th><th>Distribution</th><th colspan="2">Dinajpur</th></tr>
<tr>
<td>
</td>
<td>
</td>
<td>Normal</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>
</td>
<td>
</td>
<td>Log-Normal</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>
</td>
<td>
</td>
<td>Gamma</td>
<td>
</td>
<td>
</td>
</tr>
</table>

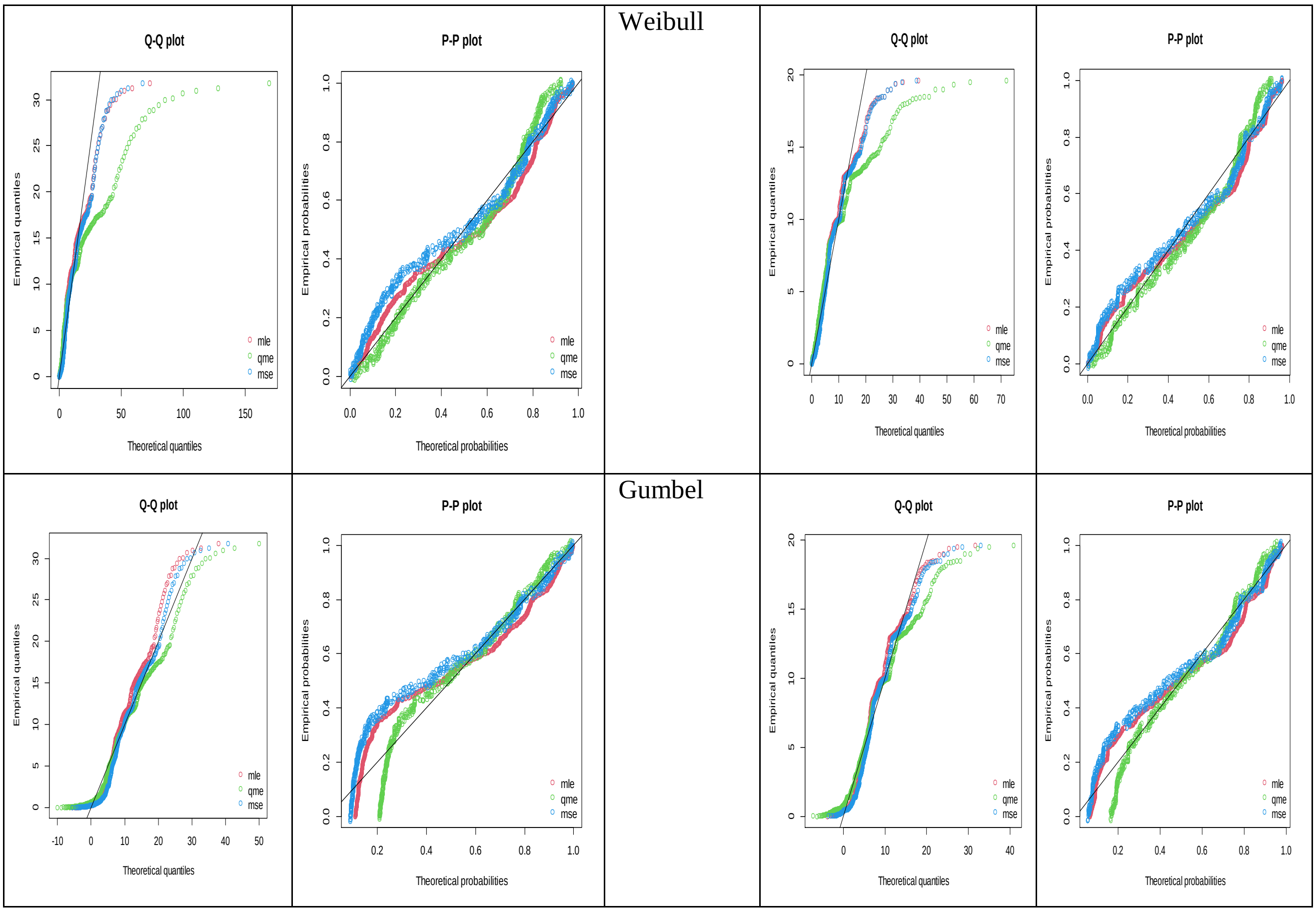


Figure 4: Fitted goodness of fit plots of different estimation methods with different distributions

The fitted goodness of fit plots of different estimation methods with different distributions reported in Figure 4.6 indicates that the Q-Q and P-P confirm that the MLE shows best performance compared to QME and MSE methods for all of these distribution for both districts.

5. Conclusion

The adequate amount of rainfall is essential for agricultural production, industrial development, and other human activities in particular area Sufficient rain is a blessing for agriculture but it becomes dangerous if it's scarce or in excess. The aim of this study is twofold. First, we fit different statistical distributions such as Normal, Log Normal, Gamma, Weibull, and Gumbel distribution on the monthly rainfall data of two districts Pabna (Ishurdi weather stations) from Rajshahi division and Dinajpur weather station from Rangpur division. Secondly, we fit the statistical distributions with different estimation methods such as Maximum Likelihood Estimation

(MLE), Quantile Matching Estimation (QME), and Maximum Spacing Estimation (MSE) methods and find out the best estimation method. This study considers the monthly rainfall data from January 1971 to December 2015. For model evaluation, we use AIC, BIC, and LL values for model comparison and A-D, CM, and K-S test statistics for measuring the goodness of fit test. Besides these, we also use some goodness-of-fit plots such as density plots, CDF plots Q-Q plots, and P-P plots. This finding concludes as follows:

- The distribution is considered right or positively skewed. Also, the value indicates a highly skewed distribution and the distribution is leptokurtic.
- By using the Maximum likelihood estimation (MLE) method, we can found that Gamma distribution performs better based on model evaluation criteria as well as some specific statistics as well as graphical plots for fitting the monthly rainfall data in case of Pabna and Dinajpur districts.
- By using the Quantile Matching estimation (QME) method, we found that the Gamma distribution performs better based on model evaluation criteria whereas Normal distribution shows worse fitting performance in case of both districts.
- By using the Maximum spacing estimation (MSE) method, we can find Gamma distribution performs better.
- By the comparison of these three methods (Maximum likelihood estimation, Quantile Matching estimation, and Maximum spacing estimation) we can find Maximum likelihood estimation (MLE) gives better estimation performance compared to MSE and QME methods for both districts.

Therefore, the study concludes that the Gamma distribution shows better fitting performance to the monthly rainfall data in the case of the Dinajpur district and the Maximum Likelihood Estimation (MLE) method is the best estimation method compared to QME and MSE methods. The fitting statistical distributions to rainfall data with different estimation techniques gives an interpretation of how stable the weather condition of the country is, as fitting statistical distributions in rainfall data is obviously of immense meteorological importance.

References

[1] Ahasan, M. N., Chowdhary, M. A., and Quadir, D. A. (2010). Variability and trends of summer monsoon rainfall over Bangladesh. Journal of Hydrology and Meteorology. 7(1), 1-17.

[2] Alghazali, N.O. and Alawadi, D.A., (2014). Fitting statistical distributions of monthly rainfall for some Iraqi stations. Civil and Environmental Research. 6(6), pp.40-46.

[3] Al-Mansory H. B. (2005). Statistical Analysis of Extreme Monthly Rainfall in Basrah City, South of Iraq, Marina Mesopotamica. Basrah, Iraq, Vol. 20, 2, PP.283-296.

[4] Cheng, R.C.H; Stephens, M. A. (1989). A goodness-of-fit test using Moran's statistic with estimated parameters. Biometrika. 76 (2): 386–392.

[5] Chiew, F. H. (2006). Estimation of rainfall elasticity of streamflow in Australia. Hydrological Sciences Journal. 51(4), 613-625.

[6] Cullen, A. and Frey, H. (1999). Probabilistics techniques in exposure assessment: A handbook for dealing with variability and uncertainty in models and inputs. https://www.semanticscholar.org/paper/Probabilistic-Techniques-in-ExposureAssessment%3A-A-Cullen-Frey/e66b37fd356f47f593755ad33d9ebffd69e22b2b#citing-papers

[7] Easterling, D, R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., and Mearns, L. O. (2000). Climate extremes: observations, modeling, and impacts. Science, 289(5487), 2068-2074.

[8] Husak, G. J., Michaelsen, J. and Funk, C. (2006). Use of the Gamma Distribution to Represent Monthly Rainfall in Africa for Drought Monitoring Applications, International Journal of Climatology, Published online in Wiley Inter. Science, Royal Meteorological Society, Santa Barbara, U. S. A., DOI: 10.1002/joc.

[9] IPCC. (2007). Climate Change 2007, Synthesis report, https://www.ipcc.ch/site/assets/uploads/2018/02/ar4_syr_full_report.pdf

[10] Kirby, J. M. Mainuddin, M., Mpelasoka, F., Ahmad, M. D., Palash, W., Quadir, M. E. and Hossain, M. M. (2016). The impact of climate change on regional water balances in Bangladesh. Climatic change. 135(3), 481-491.

[11] Maliva, R. and Missimer, T. (2012). Arid Lands and Water Evaluation and Management, Springer-Verlag Berlin Heidelberg. ISBN 978-3-642-29104-3.

[12] Mohamed, T.M. and Ibrahim, A.A.A., (2015). Fitting probability distributions of annual rainfall in Sudan. SUST Journal of Engineering and Computer Sciences (JECS). Vol. 17, No.2.

[13] Olumide B. A., Saidu M., Oluwasesan A. (2013). Evaluation of Best Fit Probability Distribution Models for the Prediction of Rainfall and Runoff Volume (Case Study Tagwai Dam, Minna-Nigeria). International Journal of Engineering and Technology. U. K., Vol. 3, No. 2, PP. 94-98.

[14] Papalexiou, S. M., Koutsoyiannis, D. and Makropoulos, C. (2012). How Extreme is Extreme? An Assessment of Daily rainfall Distribution Tails, Discussion Paper. Journal of Hydrology and Earth System Sciences. 9: 5757–5778.

[15] Ranneby, B. (1984). The maximum spacing method. An estimation method related to the maximum likelihood method. Scand. J. Statist. 11, 93-112.

[16] Rodrigo, F. S. (2002). Changes in climate variability and seasonal rainfall extremes: a case study from San Fernando (Spain), 1821-2000, Theoretical and Applied Climatology. 72(3), 193-207.

[17] Roth D. (2006). Unified Surface Analysis Manual, Hydro meteorological prediction Center, Retrieved 2006-10-22.

[18] Roy, M.K., (2004). Fundamental of Probability & Probability Distributions, 3rd edition, Romax Publications, Chittagong.

[19] Sen, Z. and Eljadid, A.G.,(1999). Rainfall distribution function for Libya and rainfall prediction. Hydrological sciences journal. 44(5), pp.665-680.

[20] Sgouropoulos, N., Yao, Q., and Yastremiz, C. (2015). Matching a distribution by matching quantiles estimation. Journal of the American Statistical Association. 110(510):742–759. PMID: 26692592.

[21] Shahid, S. (2009). Spatio-temporal variability of rainfall over Bangladesh during the time period 1969-2003. Asia-Pacific Journal of Atmospheric Science. 45(3), 375-389.

[22] Shahid, S. (2010). Recent trends in the climate of Bangladesh. Climate Research. 42(3), 185-193.

[23] Shahid, S. (2011). Trends in extreme rainfall events of Bangladesh. Theoretical and applied climatology. 104(3), 489-499.

[24] Shahid, S., Wang, X. J., Harun, S. B., Shamsudin, S. B., Ismail, T., and Minhans, A. (2016). Climate variability and changes in the major cities of Bangladesh: observations, possible impacts and adaptation. Regional Environmental Change. 16(2), 459-471.

[25] Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. Ann. Math. Stat. 19: 279–281.

[26] Sreedhar, B.R. (2019). Fitting of Probability distribution for analyzing the rainfall data in the state of Andhra Pradesh, India. International Journal of Applied Engineering Research. Vol. 14, 835-839.

[27] Suhaila, J. and Jemain, A.A., (2007). Fitting daily rainfall amount in Peninsular Malaysia using several types of exponential distributions. Journal of applied sciences research. 3(10), pp.1027-1036.

[28] Watkins, J.C. (2011). Topic 15: Maximum likelihood estimation. https://www.math.arizona.edu/~jwatkins/o-mle.pdf

[29] Zong, Y., and X. Chen, (2000). The 1998 flood on Yangtze, China. Nat. Hazards. **22**, 165–184.