

## Analyzing the Best Fitted Probabilistic Model for the Seasonal Rainfall Data in Khulna Region of Bangladesh

K. Fatema<sup>ID</sup>, Muhammad Habibulla Alamin, M. Zahid Hasan<sup>\*ID</sup>, M. Murad Hossain<sup>ID</sup>

Department of Statistics, Faculty of Science, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj-8100, Bangladesh

khairunnasafatema@gmail.com, habibulla.stat@gmail.com, zahid1680@gmail.com, sadid.shimul@gmail.com

\*Correspondence: zahid1680@gmail.com

**ABSTRACT.** There are several pieces of research on the statistical modeling of rainfall data in Bangladesh. Since all the seasons of a year do not receive a similar amount of rainfall, hence one single statistical model might not be able to explain the pattern of rainfall at any season of a year. According to the climatologists, Bangladesh has four seasons which are Monsoon, Post-monsoon, Summer, and Winter based on the geographical characteristics of this country. This paper aims to determine the best-fitted probability distribution model for the monthly rainfall data of each particular season in the Khulna district of Bangladesh using the rainfall data of the Khulna region from 1951 to 2018. Very commonly used seven continuous distributions- Normal, Weibull, Gamma, Log-normal, Exponential, Cauchy, and Logistic distributions were used to model the data and to evaluate the performances of the distributions, three non-parametric goodness-of-fit tests were conducted, and AIC, BIC values were calculated. Parameters of the distributions were estimated by the maximum likelihood method. The best-fit result of each season was taken as the distribution with the lowest AIC and BIC values. Among the seven distributions, the Gamma distribution showed the best-fit results of the monthly rainfall data for the Monsoon, Post-Monsoon, and Winter Season, and the Weibull distribution showed the best-fit result for Summer Season.

### 1. Introduction

The amount of rainfall received by a region is an influencing factor in determining the amount of water available to meet the different demands of agriculture, industry, and other human activities. Therefore, for the welfare of the national economy, the study of the distribution pattern of rainfall in time and space is very important. The analysis of the rainfall data depends heavily on its distribution pattern.

---

Received: 21 Oct 2021.

Key words and phrases. gamma distribution; Weibull distribution; goodness of fit; distribution of rainfall; Khulna.

The establishment of a probability distribution that provides a good fit for monthly rainfall has long been a topic of interest in the fields of meteorology. In the context of Bangladesh, several studies can be found based on such probabilistic modeling of data in the field of meteorology especially in the earthquake, precipitation, humidity, and temperature (Fatema et al. 2021). In a study on the daily maximum rainfall of Pantnagar, India for 37 years; Sharma and Singh (2010) established the best-fitted probability distribution using the least square method among the 16 compared distributions (Sharma and Singh 2010). The log-Pearson type III distribution fitted for fifty percent of the total stations for the rainfall distribution characteristics of the Chia-Nan plain region in the study conducted by Lee (2005) (Lee 2005). In their papers, Nadarajah (2006) and Nadarajah and Withers (2000) provided the application of extreme value distributions to the rainfall data over fourteen locations spread throughout West Florida and sixteen locations in New Zealand, respectively (Nadarajah 2006) (Withers and Nadarajah 2000). Deka and Borah (2009) have derived the best-fitted probability distribution to correctly describe the annual series of maximum rainfall data for the period 1966 to 2007 of nine distantly located stations in northeast India, considering only five extreme values distributions (Deka, Borah, and Kakaty 2009). Annual rainfall data for fourteen rainfall stations in Sudan for the period 1971 to 2010 were analyzed where the best fitted probability distributions were considered to be the normal distribution and gamma distribution (Mahgoub and Abd Allah 2017). The spatial and temporal distribution analysis of rainfall intensity on a local scale was conducted by Redano and Lorente (1993) (Redaño and Lorente 1993). The empirical distribution pattern and hierarchical clustering technique were used to identify homogenous areas based on precipitation of different regions in Bangladesh in the study of Rahman et al. (2018) (Rahman, Matin, and Salma 2018). Ghosh et al. (2016) showed that the generalized extreme value distribution empirically provides the best-fitted distribution for the monthly rainfall data for Chittagong, Rajshahi, Sylhet and the gamma distribution for Dhaka (Ghosh, Roy, and Biswas 2016). But still now, there is no research like probabilistic modeling of monthly rainfall data of four different seasons in Khulna regions of Bangladesh. The motive of this study is to fit and select the best-fitted probability distribution for the monthly rainfall data for four different seasons of Khulna district for the period 1951–2018 in Bangladesh so that visualize and predict

about the nature of monthly rainfall data of this district. For this purpose the normal distribution, Weibull distribution, Gamma distribution, lognormal distribution, exponential distribution, Cauchy distribution and Logistic distribution have been fitted while the parameters of the distributions are estimated through iterative maximum likelihood method. In this study, the following steps were applied particularly for the fitting of probability distributions—Collecting of Data of interest and exploratory data analysis is such as visualization, detection of outlier, trimming etc. Several descriptive measures are used to visualize the nature of the data, Fit the probability distributions for monthly rainfall data of different seasons of Khulna district, check the accuracy of fitted probability distributions, and identify the best-fitted distribution based on Akaike information criterion (AIC), and Bayesian information criterion (BIC).

## 2. Study Area

Bangladesh has 64 districts and 8 major divisions among which Khulna is the third-largest divisional city and economic Centre in Bangladesh after Dhaka and Chittagong. It has an area of 4394.46 square kilometers. It is located in south-western Bangladesh coordinates at extends from  $22^{\circ}49'0''N$  latitude and from  $89^{\circ}33'0''E$  longitude on the besides of the Rupsha and Bhairab river. It is also situated at the north of the Port of Mongla and has a variety of light and heavy industries like jute, chemicals, fish and sea-food processing and packaging, sugar milling, shipbuilding, and so on. The amount of rainfall received by this region has a noticeable impact on its agricultural and industrial sectors.

## 3. Data collection and Methodology

### 3.1 Data collection

The monthly rainfall data of Khulna district for the period 1951–2018 were collected from the Bangladesh Meteorological Department (BMD). In Bangladesh, four distinct seasons can be recognized from a climatic point of view (Hasan, Akhter, and Islam 2014). These four seasons are as follows:

- Dry winter season (December to February)
- Pre-monsoon hot summer season (March to May)
- Rainy monsoon season (June to September)

- Post monsoon autumn season (October to November)

The study focuses on the statistical analysis and modeling of these above seasonal rainfalls in the Khulna region.

### 3.2 Fitting of the probability distributions

In the recent past, a lot of probability distribution functions have been proposed to explain the amount of monthly rainfall data. But in this study, seven continuous probability distributions viz. Normal, Weibull, Gamma, Lognormal, Exponential, Cauchy, and Logistic distribution have been considered to describe the characteristics of rainfall in the particular Khulna region. The description of various probability distribution functions viz. density function, range, and the parameter involved are presented in table 1.

**Table 1.** Probability density functions with parameters.

Distribution	Probability density function	Parameter	Range
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$	$\mu$ : Location parameter $\sigma$ : Scale parameter	$-\infty \leq x \leq \infty$ $-\infty \leq \mu \leq \infty$ $\sigma^2 > 0$
Weibull	$f(x) = \frac{\alpha}{\beta} \left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x-\gamma}{\beta}\right)^\alpha\right]$	$\alpha$ : Shape parameter $\beta$ : Scale parameter $\gamma$ : Location parameter	$x \geq 0$ $\alpha, \beta, \gamma > 0$
Gamma	$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x-\gamma}{\beta}} (x-\gamma)^{\alpha-1}$	$\alpha$ : Shape parameter $\beta$ : Scale parameter $\gamma$ : Location parameter	$0 \leq x \leq \infty$ $\alpha \geq 0, \beta \geq 0$
Lognormal	$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\ln(x)-\mu_Y)^2\right]$	$\mu$ : Location parameter $\sigma_Y$ : Scale parameter	$0 \leq x \leq \infty$ $-\infty \leq \mu \leq \infty$ $\sigma_Y > 0$ $x > \alpha$
Exponential	$f(x) = \lambda e^{-\lambda(x-\gamma)}$	$\gamma$ : Location parameter $\lambda$ : Inverse scale parameter	$\alpha > 0$ $\lambda > 0$
Cauchy	$f(x; \theta, \lambda) = \frac{1}{\pi\lambda \left[1 + \left(\frac{x-\theta}{\lambda}\right)^2\right]}$	$\theta$ : Location parameter $\lambda$ : Scale parameter	$-\infty < x < \infty$ $\lambda > 0$
Logistic	$f(x) = \frac{\exp\left[-\frac{(x-\alpha)}{\beta}\right]}{\beta \left[1 + \exp\left\{\frac{-(x-\alpha)}{\beta}\right\}\right]^2}$	$\alpha$ : Location parameter $\beta$ : Scale parameter	$-\infty < x < \infty$ $\beta > 0$

### 3.3 Goodness-of-Fit Tests (GOF)

The goodness of fit tests measures the compatibility of a random sample with a theoretical probability distribution function. The tests were performed to justify the following hypothesis:

$H_0$ : The amount of monthly rainfall data follows the specified distribution.

$H_1$ : The amount of monthly rainfall data does not follow the specified distribution.

The summarized information of Kolmogorov-smirnov test, Anderson-Darling test, Cramer-Von Mises Test is given in table 2. The following goodness of fit test was used  $\alpha$  (0.05) level of significance for the selection of the best-fitted probability distribution model.

**Table 2.** Goodness of fit Statistic.

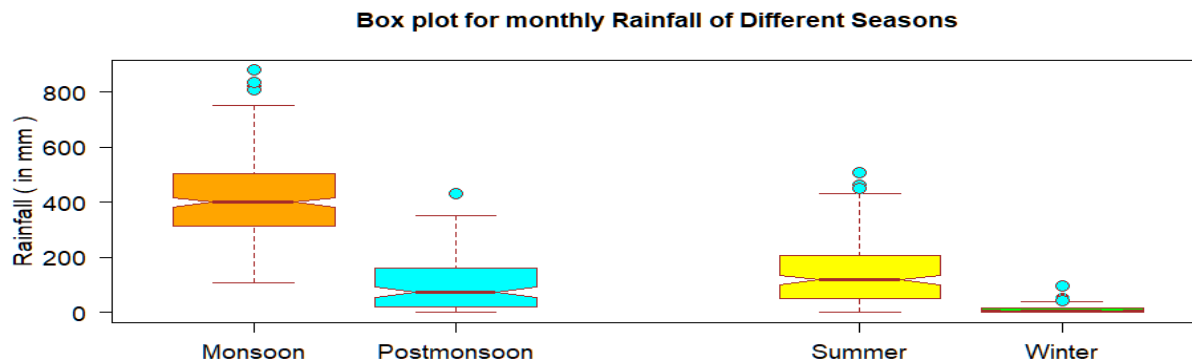
Statistic	General Formula	Computational Formula
Anderson-Darling (AD)	$n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]} dx$	$-n - \frac{1}{n} \sum_{i=1}^n (2i-1) \log [F_i(1 - F_{n+i-1})]$ ; where, $F_i \triangleq F(x_i)$
Cramer-von-Mises (CvM)	$n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dx$	$\frac{1}{12n} + \sum_{i=1}^n \left( F_i - \frac{2i-1}{n} \right)^2$
Kolmogorov-Smirnov(KS)	$\sup  F_n(x) - F(x) $	$\max(D^+, D^-)$ with $D^+ = \max_{i=1..n} \left( \frac{i}{n} - F_i \right)$ and $D^- = \max_{i=1..n} \left( F_i - \frac{i-1}{n} \right)$

## 4. Results and Discussions

### 4.1 Data Description

The rainfall (in mm) data of monsoon (Jun- September), post-monsoon (October-November), summer (March-May), and winter (December-February) season from 1951 to 2018 have been taken into consideration to fit the probability distribution. The rainfall (in mm) data were recorded at the Khulna station (latitude, longitude) "Khulna (22.78, 89.53)" by the Bangladesh Meteorological Department from 1951 to 2018. The

descriptive statistics for the data can be discussed in the first section, to have a better understanding of the rainfall data before entering the desired study.



**Figure 1.** Box-and-Whisker plot of the rainfall (in mm) of four seasons of Khulna District in Bangladesh.

**Table 3.** Summary statistics of the seasonal rainfall data for the Khulna district during the year 1951–2018.

Parameter	Season			
	Monsoon	Post-monsoon	Summer	winter
Mean	411.807	97.649	140.351	11.444
Median	399.293	72.767	119.044	5.7429
Mode	406	25	5.00	1.00
Standard Deviation	140.331	95.490	111.915	13.729
Kurtosis	3.444	3.496	3.234	10.157
Skewness	0.529	1.066	0.915	2.087
Range	772.927	433.251	506.806	96.998
Minimum	106	0.113	1.00	0.001
Maximum	878.927	433.365	507.81	97.000

The box-and-whisker plots in figure 1 demonstrate that outliers were present in the series of the monsoon, post-monsoon, summer, winter rainfall data of the Khulna district. From table 3, it can be seen that the maximum rainfall occurs in Monsoon (878.927 mm)

and minimum rainfall occurs in Winter (0.001 mm). The coefficient of skewness is used to verify the degree of asymmetric distribution around the mean. From the result of table 3, the distribution of monthly rainfall data was found to be positively skewed for each season. The coefficient of skewness of post-monsoon and winter season is greater than one, thus can be regarded as highly skewed, Monsoon and summer seasons lie within the range of 0.5 to 1, are moderately skewed. Kurtosis refers to the extent of peakedness or flatness of a probability distribution in comparison with the normal probability distribution. Most values of kurtosis are around 3 in the range of 3.23 in the summer season to 3.49 in the post-monsoon season. The Winter season obtain the highest value of kurtosis with the value of 10 which is highly leptokurtic.

#### 4.2 Fitted Parameters of Probability Distributions and Goodness of Fit Results

The estimated parameters of the distributions are presented in Table 4.

**Table 4.** Estimated values of the parameters of different PDFs for the seasonal rainfall data of the Khulna district, Bangladesh.

Season	Distribution	Parameter	Estimated value of the Parameter	Standard Error	
Monsoon	Normal	Mean	411.807	8.493	
		Standard deviation	140.073	6.005	
	Weibull	Shape	3.130	0.141	
		Scale	459.917	9.417	
	Gamma	Shape	8.284	0.678	
		Rate	0.020	0.0016	
	Lognormal	Log Mean	Log Mean	5.958	0.0219
			Log Standard Deviation	0.362	0.0155
		Exponential	Rate	0.002	0.00011
			Location	396.021	8.468
	Cauchy	Scale	84.333	6.444	
		Location	405.606	8.364	
	Logistic	Scale	79.047	3.975	
		Normal	Mean	97.649	8.158
Standard deviation	95.138		5.768		
Post-Monsoon	Weibull	Shape	0.8365	0.059	
		Scale	89.934	9.663	
	Gamma	Shape	0.7295	0.0747	
		Rate	0.0074	0.0010	

Summer	Lognormal	Log Mean	3.7578	0.1445	
		Log Standard Deviation	1.6859	0.1022	
	Exponential	Rate	0.0102	0.00086	
	Cauchy	Location	51.302	8.0882	
		Scale	50.321	6.216	
	Logistic	Location	85.554	8.056	
		Scale	53.493	3.813	
	Normal	Mean	140.352	7.8163	
		Standard deviation	111.641	5.5270	
	Weibull	Shape	1.1799	0.0667	
		Scale	148.125	9.2299	
	Gamma	Shape	1.23445	0.1077	
		Rate	0.00879	0.00092	
	Winter	Lognormal	Log Mean	4.4871	0.08074
			Log Standard Deviation	1.1532	0.05709
			Exponential	Rate	0.00712
		Cauchy	Location	102.932	7.78448
			Scale	65.1031	5.81935
Logistic		Location	128.460	7.77064	
		Scale	63.2478	3.67133	
Normal		Mean	11.444	0.9589	
		Standard deviation	13.696	0.6780	
Weibull		Shape	0.6724	0.0386	
		Scale	9.0435	0.9865	
Gamma		Shape	0.5425	0.0446	
		Rate	0.0473	0.0059	
Winter		Lognormal	Log Mean	1.2811	0.1489
			Log Standard Deviation	2.1274	0.1053
			Exponential	Rate	0.0873
		Cauchy	Location	4.1645	0.5855
			Scale	4.9787	0.5470
Logistic	Location	9.276	0.840		

### 4.3 Accuracy Measures – Goodness of Fit Statistic and Criteria

The Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling goodness of fit statistics and the Akaike information criterion (AIC) value and Bayesian information criterion (BIC) value are given in the following table 5.



**Table 5.** Goodness of Fit Statistics and goodness of fit criteria for different distributions for four seasons rainfall of Khulna district, Bangladesh.

Season	Distribution	Goodness-of-fit statistic			Goodness-of-fit criteria		
		Kolmogorov-Smirnov	Cramer-von Mises	Anderson-Darling	Log-likelihood	AIC	BIC
Monsoon	Normal	0.0468	0.1216	0.8642	-1730.22	3464.44	3471.65
	Weibull	0.0417	0.1201	0.9274	-1729.03	3462.06	3469.27
	*Gamma	0.0325	0.0514	0.3522	-1724.74	3453.49	3460.71
	Lognormal	0.0514	0.1956	1.2615	-1730.95	3465.89	3473.10
	Exponential	0.3554	11.2045	55.3365	-1909.59	3821.18	3824.78
	Cauchy	0.0969	0.4338	4.3494	-1775.23	3554.45	3561.66
	Logistic	0.0396	0.0806	0.6746	-1730.26	3464.52	3471.73
Post-Monsoon	Normal	0.1574	0.8838	5.4406	-812.50	1629.01	1634.83
	Weibull	0.0862	0.1891	1.2065	-755.63	1515.27	1521.09
	*Gamma	0.0823	0.1561	0.9120	-754.02	1512.04	1517.87
	Lognormal	0.1281	0.5885	3.6966	-775.08	1554.16	1559.99
	Exponential	0.1028	0.3961	3.1958	-759.07	1520.14	1523.05
	Cauchy	0.2472	1.6352	10.2591	-829.87	1663.75	1669.57
	Logistic	0.1683	0.6312	4.6985	-812.36	1628.72	1634.54
Summer	Normal	0.1059	0.7353	4.7322	-1251.38	2506.76	2513.39
	*Weibull	0.0540	0.0921	0.6099	-1208.63	2421.27	2427.90
	Gamma	0.0595	0.1261	0.7570	-1209.95	2423.91	2430.54
	Lognormal	0.1041	0.6428	3.9189	-1233.92	2471.85	2478.48
	Exponential	0.0766	0.3169	1.7652	-1212.61	2427.21	2430.53
	Cauchy	0.1809	1.0610	7.7326	-1283.33	2570.66	2577.3
	Logistic	0.1176	0.4552	3.9037	-1252.15	2508.29	2514.93
Winter	Normal	0.2017	2.1004	12.0312	-823.35	1650.70	1657.34
	Weibull	0.0714	0.2158	1.4281	-672.64	1349.29	1355.93
	*Gamma	0.0593	0.1215	0.7408	-668.29	1340.57	1347.21
	Lognormal	0.1157	0.8713	5.6400	-704.82	1413.63	1420.27
	Exponential	0.1792	1.6137	14.7077	-701.26	1404.52	1407.83
	Cauchy	0.2783	3.5032	21.8755	-809.34	1622.68	1629.32
	Logistic	0.2086	1.3584	9.6571	-808.11	1620.22	1626.86

(\* indicates the best-fitted model comparing the others)

Based on AIC and BIC values presented in table 5, it is observed that the Gamma distribution for Monsoon season (AIC=3453.49, BIC=3460.71), Post-monsoon season (AIC=1512.04, BIC=1517.87), Winter season (AIC=1340.57, BIC=1347.21), and Weibull distribution for Summer season (AIC=2421.27, BIC=2427.90) are the best-fitted probability distributions for the monthly rainfall data of Khulna district in

Bangladesh. The graphical comparisons of different probability distributions given in figure 2 for the Monsoon season, figure 3 for the Post-monsoon season, figure 4 for the Summer season and figure 5 for the Winter season also show support in the favor of the result.

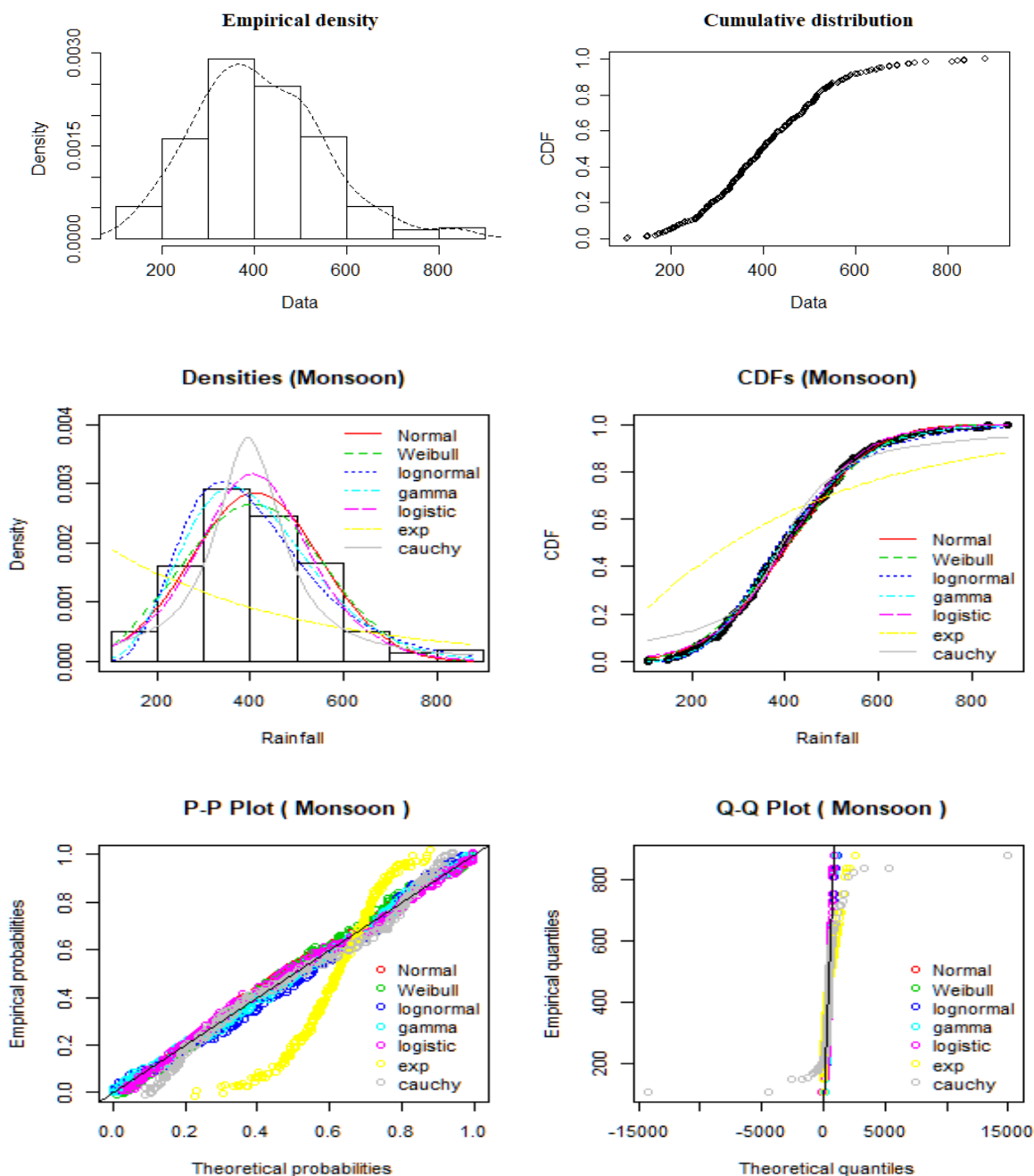


Figure 2. Histogram and density function, cumulative density function, P-Pplot, and Q-Q plot for the Monsoon season rainfall data of Khulna district.

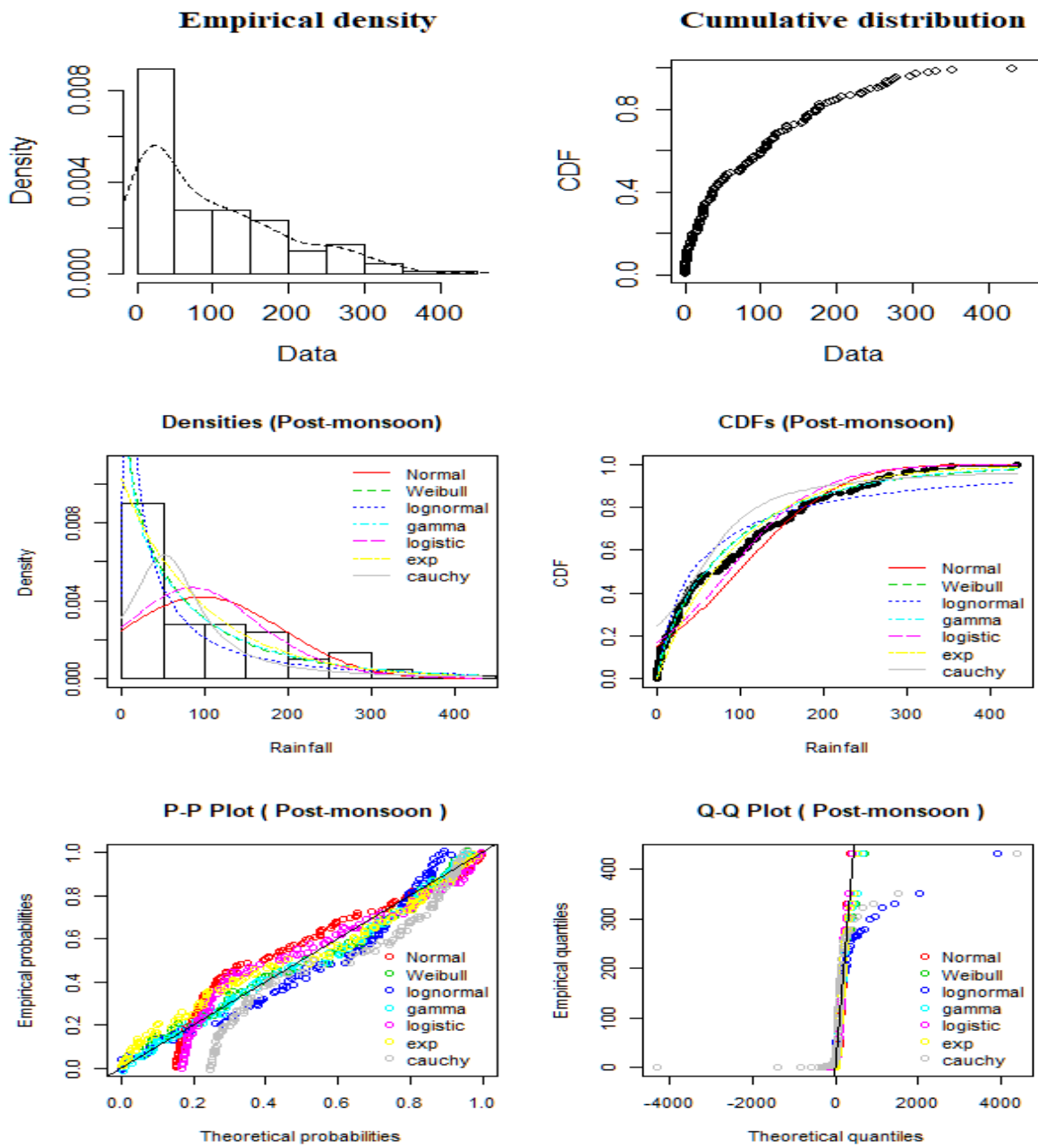


Figure 3. Histogram and density function, cumulative density function, P-P plot, and Q-Q plot for the Post-monsoon season rainfall data of Khulna district.

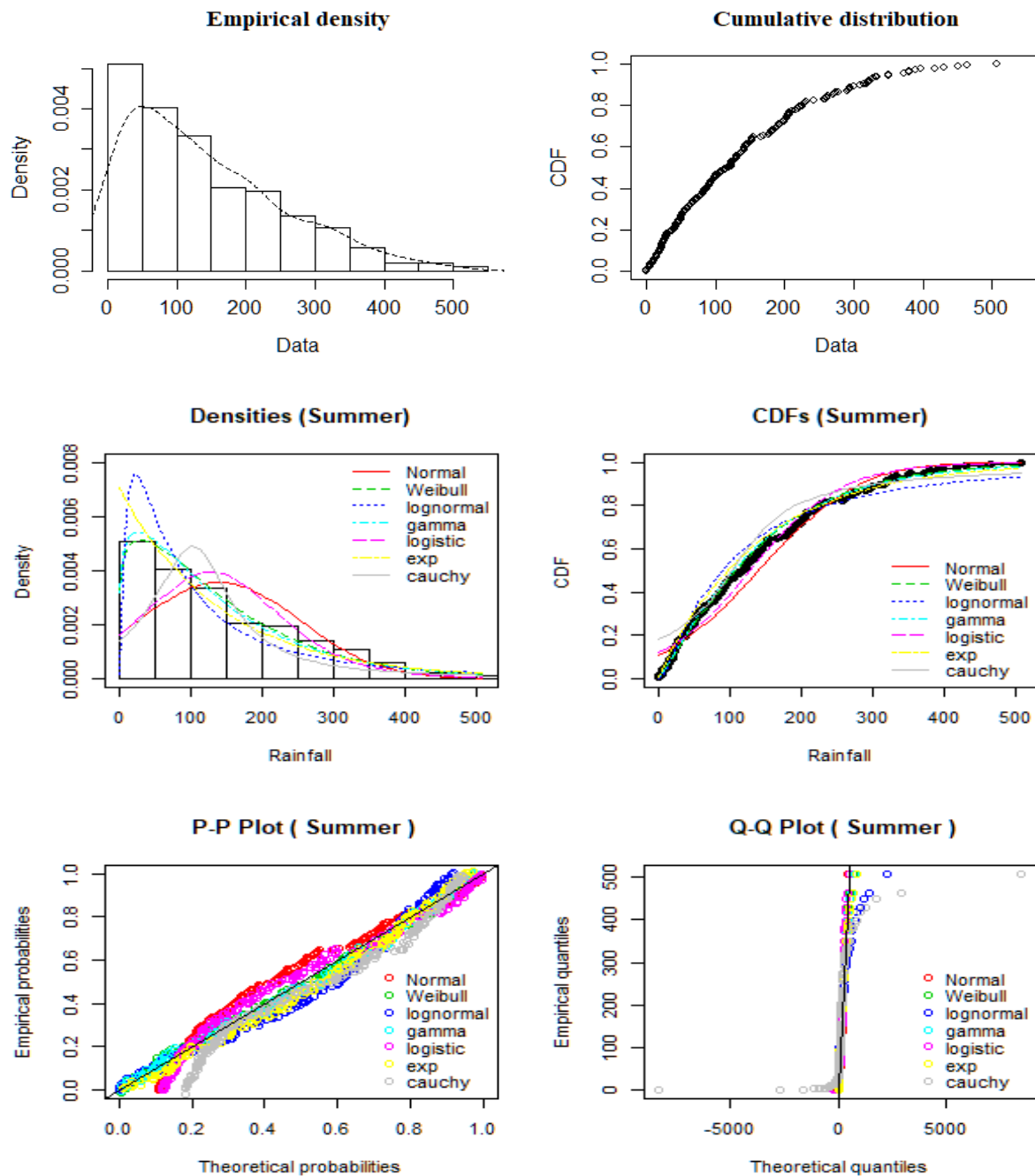


Figure 4. Histogram and density function, cumulative density function, P-P Plot, and Q-Q Plot for the Summerseason rainfall data of Khulna district.

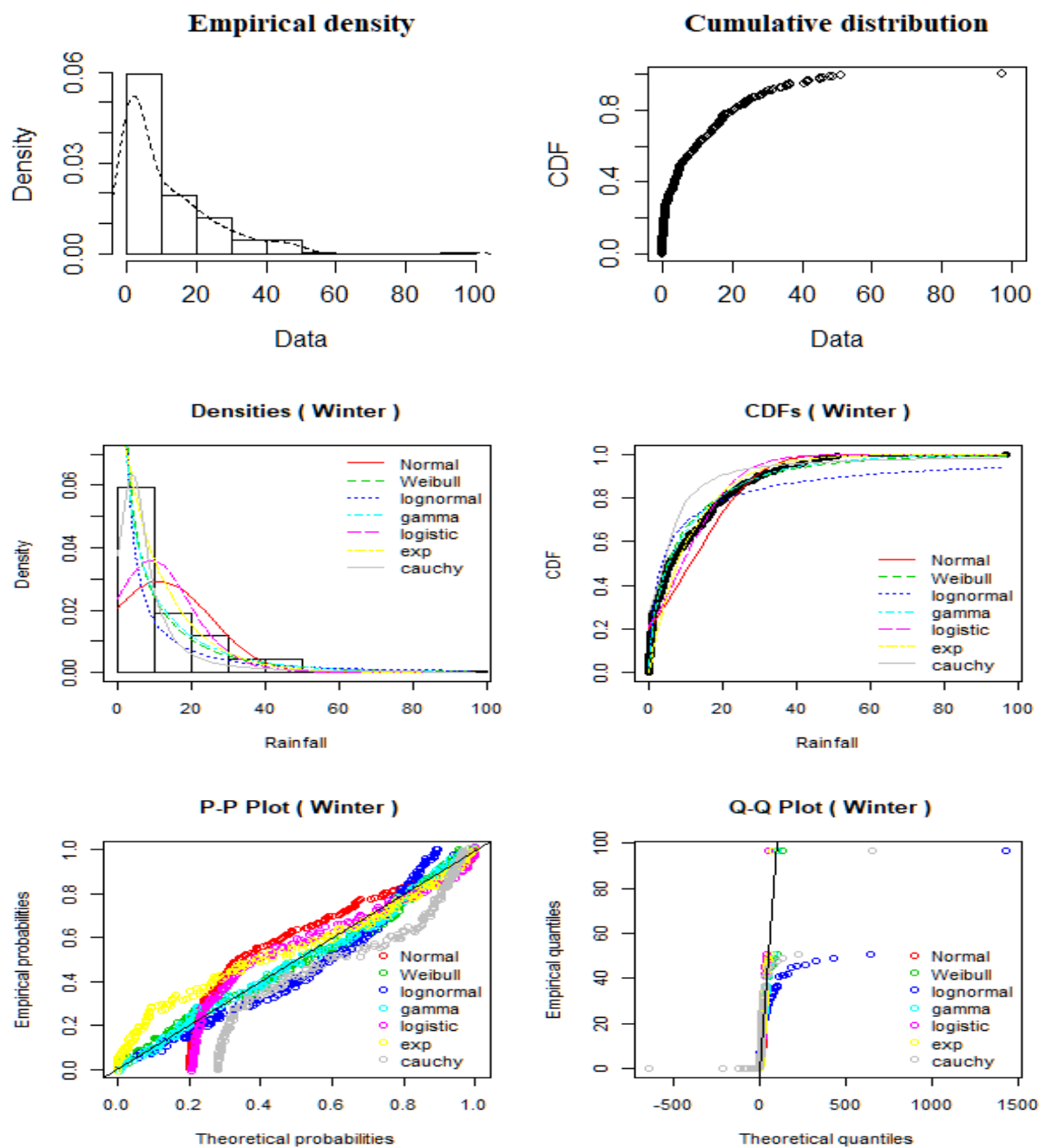


Figure 5. Histogram and density function, cumulative density function, P-P plot, and Q-Q plot for the Winterseason rainfall data of Khulna district.

## 5. Conclusion

The Seven Distributions are fitted for monthly rainfall data of Khulna station and the parameters are estimated using the maximum likelihood method. Based on the graphical assessment and goodness-of-fit criterion, the Gamma distribution for Monsoon, Post-monsoon, and Winter and the Weibull distribution for the Summer are found to be the best fit probability distributions for the monthly rainfall data considered for the Khulna district in Bangladesh. The study of frequency analysis can be used as a rough guide by engineers and hydrologists to prevent floods and droughts and applied to the planning and designing of water resources related to engineering such as reservoir design, flood control work. Analysis of rainfall and determination of annual maximum daily rainfall would enhance the management of water resources applications as well as the effective utilization of water resources. The findings can be used for the future plan, welfare to mankind of the country. Therefore, getting an idea about the distribution of rainfall may help a lot in policy-making decisions in different sectors.

## Conflict of Interest

No conflict of interest.

## REFERENCES

- [1] S. Deka, M. Borah, and S. C. Kakaty, Distributions of annual maximum rainfall series of North-East India, *Eur. Water* 27(28) (2009) 3–14. [http://www.ewra.net/ew/pdf/EW\\_2009\\_27-28\\_01.pdf](http://www.ewra.net/ew/pdf/EW_2009_27-28_01.pdf).
- [2] K. Fatema, M. Zahid Hasan, M. Murad Hossain, M. Forhad Hossain, Probabilistic modeling and predicting mean recurrence time of major earthquakes in Bangladesh, *Appl. Math. Sci.* 15 (2021) 239–247. <https://doi.org/10.12988/ams.2021.914482>.
- [3] S. Ghosh, M.K. Roy, and S.C. Biswas, Determination of the best fit probability distribution for monthly rainfall data in Bangladesh, *Amer. J. Math. Stat.* 6 (2016) 170–174.
- [4] Z. Hasan, S. Akhter, and M. Islam, Climate change and trend of rainfall in the south-east part of Coastal Bangladesh, *Eur. Sci. J.* 10(2) (2014) 25–39. <https://core.ac.uk/download/pdf/328024062.pdf>.
- [5] C. Lee, Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in Southern Taiwan, *Crop Environ. Bioinf.* 2 (2005) 31–38. <http://dx.doi.org/10.30061/2fCEB.200503.0003>.
- [6] T. Mahgoub, and A.A. Allah, Fitting Probability Distributions of Annual Rainfall in Sudan, *J. Eng. Computer Sci.* 17(2) (2017) 35–39. <http://journal.sustech.edu/index.php/JECS/article/view/173>.

- [7] S. Nadarajah, The exponentiated Gumbel distribution with climate application, *Environmetrics*. 17 (2006) 13–23. <https://doi.org/10.1002/env.739>.
- [8] Md.H. Rahman, M.A. Matin, U. Salma, Analysis of precipitation data in Bangladesh through hierarchical clustering and multidimensional scaling, *Theor. Appl. Climatol.* 134 (2018) 689–705. <https://doi.org/10.1007/s00704-017-2319-y>.
- [9] A. Redano, J. Lorente, Modelling the spatial and temporal distribution of rainfall intensity at local scale, *Theor. Appl. Climatol.* 47 (1993) 25–32. <https://doi.org/10.1007/BF00868893>.
- [10] M.A. Sharma, and J.B. Singh, Use of Probability Distribution in Rainfall Analysis, *N. Y. Sci. J.* 3 (9) (2010) 40–49. [http://danida.vnu.edu.vn/cpis/files/Refs/Drought/Use of Probability Distribution in Rainfall Analysis.pdf](http://danida.vnu.edu.vn/cpis/files/Refs/Drought/Use%20of%20Probability%20Distribution%20in%20Rainfall%20Analysis.pdf).
- [11] C.S. Withers, and S. Nadarajah, Evidence of trend in return levels for daily rainfall in New Zealand, *J. Hydrol. (New Zealand)* 39 (2) (2000) 155–166. <https://www.jstor.org/stable/43944839>.