

Article

Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh

Md Ashraful Alam *, Kazuo Emura, Craig Farnham and Jihui Yuan 

Department of Housing and Environmental Design, Graduate School of Human Life Science, Osaka City University, Osaka 558-8585, Japan; emura@life.osaka-cu.ac.jp (K.E.); farnham@life.osaka-cu.ac.jp (C.F.); yuanjihui@hotmail.co.jp (J.Y.)

* Correspondence: alam13ocu@gmail.com; Tel.: +81-6-6605-2820

Received: 6 December 2017; Accepted: 28 January 2018; Published: 31 January 2018

Abstract: The study of frequency analysis is important to find the most suitable model that could anticipate extreme events of certain natural phenomena e.g., rainfall, floods, etc. The goal of this study is to determine the best-fit probability distributions in the case of maximum monthly rainfall using 30 years of data (1984–2013) from 35 locations in Bangladesh by using different statistical analysis and distribution types. Commonly used frequency distributions were applied. Parameters of these distributions were estimated by the method of moments and *L*-moments estimators. Three goodness-of-fit test statistics were applied. The best-fit result of each station was taken as the distribution with the lowest sum of the rank scores from each of the three test statistics. Generalized Extreme Value, Pearson type 3 and Log-Pearson type 3 distributions showed the largest number of best-fit results. Among the best score results, Generalized Extreme Value yielded the best-fit for 36% of the stations and Pearson type 3 and Log-Pearson type 3 each yielded the best-fit for 26% of the stations. The more practical result of this paper was that the 10-year, 25-year, 50-year and 100-year return periods of maximum monthly rainfall were calculated for all locations. The result of this study can be used to develop more accurate models of flooding risk and damage.

Keywords: probability distributions; maximum monthly rainfall; goodness-of-fit test; return period

1. Introduction

The rainfall pattern and amount at a certain location are important parameters affecting various natural and socio-economic systems e.g., flood protection, water resources management, agriculture and forestry and tourism. Extreme precipitation events are one of the primary natural causes of flooding. The ability to anticipate extremes of rainfall would aid in improved flood planning. Frequency analysis is used to determine the probability of occurrence of events. One goal of frequency analysis is to relate the magnitude of extreme events to their frequency of occurrence by using probability distributions.

Statistical theory for extremes is used to express that the frequency of such events is relatively more dependent on any changes in the variability (more generally, the scale parameter) than on the mean (more generally, the location parameter) of the climate [1]. Rainfall patterns vary from country to country as well as from weather station to station. However, homogeneity can also be found. Taking annual extreme precipitation as the extreme event, Khudri et al. [2] found that in Bangladesh, the generalized extreme value and generalized gamma four parameter distributions provide the best fit for 50% of the stations in the survey, while no other distribution was found consistently suitable for the other stations. Mandal & Choudhury [3] studied the annual, seasonal and monthly maximum daily rainfall data for Sagar Island, located on the continental shelf of the Bay of Bengal—which is near our study area—and found that normal (N) distributions were fitted best to annual, post-monsoon

and summer seasons, while Lognormal (LN2), Weibull (W2) and Pearson type 5 were fitted best to pre-monsoon, monsoon and winter seasons respectively.

Kumar [4] and Singh [5] found that the LN2 distribution is the best-fit probability distribution for annual maximum daily rainfall in India. Amin et al. [6] used annual maximum rainfall based on a daily sample and found that the log-Pearson type 3 (LP3) distribution was the best-fit distribution in the northern regions of Pakistan. Eslamian et al. [7] used the maximum monthly rainfall as an extreme event and found that in Iran, the generalized extreme value (GEV) and Pearson type 3 (P3) distributions provided the best fit. Bhakar et al. [8] found that for monthly maximum rainfall, the Gumbel (GUM) distribution was the best fit in India.

Lee [9] showed that LP3 distributions fitted best to 50% of the stations for the rainfall distribution of the Chia-Nan plain area of Taiwan. Ogunlela [10] described that the LP3 provided the best fit to peak daily precipitation in Nigeria. Kwaku et al. [11] found that the LN2 distribution was the best-fit probability distribution for one to five consecutive days' maximum rainfall for Accra, Ghana. Olofintoye et al. [12] showed that for peak daily rainfall in Nigeria, 50% of stations follow LP3 distributions and 40% follow Pearson type 3 (P3) distributions. Sen et al. [13] found that the gamma probability distribution provided the best fit to monthly maxima rainfall in arid regions in Libya. Arora & Singh [14] found that the LP3 distribution recommended by the U.S. Water Resources Council (USWRC) in 1967 was the best method of flood frequency analysis in the United States.

The frequency of extreme events such as precipitation and floods can be expressed in terms of the "return period," also known as the "recurrence interval." In extreme value analysis, the return period is defined as the average length of time between events of the same magnitude or greater. It is derived from quantiles of a parametric probability distribution fitted to the extreme values of the recorded data [15]. A goal of the frequency analysis of extreme events is to estimate the value of the event magnitude corresponding to a given return period i.e., the maximum daily rainfall that can be expected to occur every 50 years.

Bangladesh is an agriculture based economy where the role of precipitation is important. The extreme events of rainfall may affect agriculture, ecosystems, biodiversities and livelihood patterns. It is important to extract the patterns of extreme rainfall events to determine risk factors which can be used to develop long-term measures to save lives and property. In the short term, precipitation plays an important role in flooding every year.

The present work focuses on two important issues. First, to find the best-fit models for determining the frequency of extreme rainfall events in Bangladesh. Second, to extract expected maximum monthly rainfall over return periods of 10, 25, 50 and 100 years. These can be used to develop plans and policies for reducing risk and damage from extreme rainfall events in both short and long-term planning.

2. Data and Study Area

Bangladesh is mainly a riverine country. 79% of the land is occupied by the delta plain, 12% is occupied by the Chittagong hill tracks located in the southeast and northeast part of the country and 9% is occupied by four uplifted blocks which are mainly situated in the northwest and central part. Almost the entire land is floodplains except the south-eastern and north-eastern hilly areas. The Bay of Bengal is situated at the south, being a source of water vapour that passes over the country and contributes to increased rainfall. During the monsoon period, excessive rainfall occurs. Rivers which originate from the Himalayas at the north flow through Bangladesh and often cause flooding during this time. The highest elevation in the delta plain is 105 m above sea level, which is found in the northern area. Elevations decrease in the coastal south. In the Chittagong hill tracks in the southeast altitudes vary from 600 to 900 m above sea level.

The climate of Bangladesh is characterized as a tropical monsoon climate [16]. From June to October heavy seasonal rainfall occurs. About 72% of all rainfall occurs during the summer and monsoon period [17].

Time series of rainfall records were collected from the Bangladesh Meteorological Department (BMD) for 35 different locations all over the country (Figure 1). The data was provided as the daily total rainfall in millimetres. This analysis uses 30 years of data from 1984–2013. However, there are 6 stations (Ambagan (15 years), Chuadanga (25 years), Mongla (23 years), Kutubdia (29 years), Sydpur (23 years), Tangail (27 years)) which did not have 30 years of available data. The total rainfall of each calendar month was calculated. Then the calendar month in each year with the highest total was taken as the maximum monthly rainfall. This yields 30 maxima (1 for each year) for each station (with the exception of the previously noted six stations). This maximum monthly rainfall is used as the parameter for analysis of extreme rainfall estimation. The maximum monthly rainfall for the entire country always occurs in the monsoon period of June to October. At individual stations, the maximum monthly rainfall occurs in July, August or September in all of the 30 years studied. The best-fit probability distributions of these meteorological locations in Bangladesh are determined.

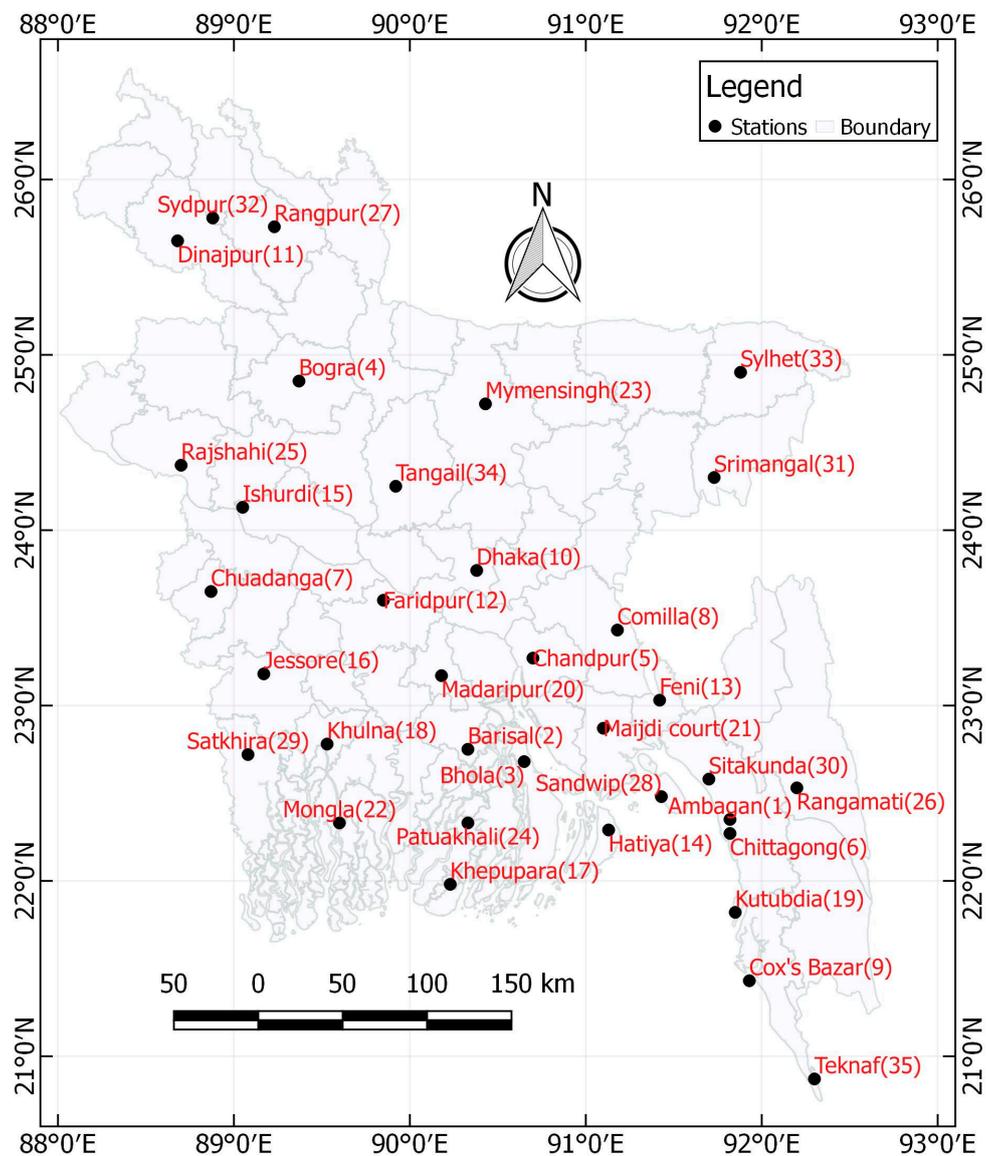


Figure 1. Meteorological stations of Bangladesh.

3. Methodology

For selecting the best-fit probability distribution for a certain location, the choice of probability distribution models is important. In this section, the selected distribution models which are commonly used in extreme rainfall analyses are presented. The method of parameter estimation, the goodness of fit tests for model selection, both numerical and graphical and the return period estimation procedure are presented.

3.1. Commonly Used Probability Distributions

This analysis adopts commonly used frequency distributions used in previous studies. The method of moments (MOM) estimators are used for parameter estimation of the distributions. They are often simple to derive but when MOM estimators are either too awkward or unavailable, the L -moments are used to estimate the parameters of the selected distributions. The L -moments are more robust than conventional moments. One disadvantage of the L -moment ratios for estimation is their smaller sensitivity. For calculating the parameters by using MOM; the sample mean \bar{X} , standard deviation σ and coefficient of skewness γ are determined by the following equations:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (2)$$

$$\gamma = \frac{n \sum_{i=1}^n (X_i - \bar{X})^3}{(n-1)(n-2)S^3} \quad (3)$$

where n is the number of observations in the data set, i is the observation number and X_i is the observation data.

The following distributions with the method of parameter estimation are used in our study. Note that the variable X represents observed data, while x represents any possible value of that data and is used as the basis to create the continuous distribution functions below. $F_x(x)$ is the value of the distribution function for the point where x corresponds to an observed data value of X .

3.1.1. Normal

The Gaussian or N distribution is often applied in annual precipitation and runoff analysis [18]. The two moments, mean μ and variance σ^2 , are the parameters of the normal distribution. The probability density function (pdf), $f(x)$ and cumulative distribution function (cdf), $F(x)$ for a normal random variable x are expressed as,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right] \quad (4)$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right]) dx \quad (5)$$

for the range of $-\infty < x < \infty$. The symmetric nature of the N distribution should tend to make it less suitable for fitting to heavy-tailed data, like some of the rainfall data samples of this study.

3.1.2. Log-Normal

The pdf and cdf of the 2-parameter Log-normal (LN2) are expressed as:

$$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\ln(x) - \mu_Y)^2\right] \tag{6}$$

$$F(x) = \frac{1}{\sigma_Y\sqrt{2\pi}} \int_0^x \left(\frac{1}{x} \exp\left[-\frac{1}{2\sigma_Y^2}(\ln(x) - \mu_Y)^2\right]\right) dx \tag{7}$$

where the range of random variable is $x > 0$. The logarithm of the x variable, $y = \ln(x)$, is well-described by a normal distribution. By using MOM estimators, the two parameters are expressed as follows:

$$\sigma_Y = \left[\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)\right]^{1/2} \tag{8}$$

$$\mu_Y = \ln(\mu_X) - \frac{1}{2}\sigma_Y^2 \tag{9}$$

which are the commonly used parameters of the LN2 distribution. Here, σ_Y is the standard deviation and μ_Y is the mean for the LN2 distribution. The LN2 should be better suited than the simple N distribution as it allows for a “heavy tail” on one side, which can better fit extreme rainfall data.

3.1.3. Pearson Type 3

One of the most commonly used distributions in hydrology is the P3 distribution which is a two-parameter Gamma distribution with a third parameter for the location. The pdf and cdf of P3 are:

$$f(x) = \frac{1}{|\alpha|\Gamma(\beta)} \left[\left(\frac{x - \xi}{\alpha}\right)^{\beta-1}\right] \exp\left[-\frac{(x - \xi)}{\alpha}\right] \tag{10}$$

$$F(x) = \frac{1}{|\alpha|\Gamma(\beta)} \int_{\xi}^x \left[\left(\frac{x - \xi}{\alpha}\right)^{\beta-1}\right] \exp\left[-\frac{(x - \xi)}{\alpha}\right] dx \tag{11}$$

The parameters are shape β , scale α and location ξ which are estimated by MOM estimators, which are:

$$\beta = 4/\gamma^2 \tag{12}$$

$$\alpha = \sigma\gamma/2 \tag{13}$$

$$\xi = \mu - 2\sigma/\gamma \tag{14}$$

3.1.4. Log-Pearson Type 3

The LP3, another gamma family distribution, describes a random variable whose logarithm follows the P3 distribution. The pdf and cdf of LP3 are as follows:

$$f(x) = \frac{1}{|\alpha|x\Gamma(\beta)} \left[\left(\frac{\ln(x) - \xi}{\alpha}\right)^{\beta-1}\right] \exp\left[-\frac{(\ln(x) - \xi)}{\alpha}\right] \tag{15}$$

$$F(x) = \frac{1}{|\alpha|\Gamma(\beta)} \int_0^x \frac{1}{x} \left[\left(\frac{\ln(x) - \xi}{\alpha}\right)^{\beta-1}\right] \exp\left[-\frac{(\ln(x) - \xi)}{\alpha}\right] dx \tag{16}$$

3.1.5. Exponential

The exponential distribution (EXP) is a special case of the Gamma family characterized by skewness. The pdf and cdf are:

$$f(x) = \alpha^{-1} \exp\left\{-\frac{x - \zeta}{\alpha}\right\} \tag{17}$$

$$F(x) = 1 - \exp\left\{-\frac{x - \zeta}{\alpha}\right\} \tag{18}$$

with a range of $\zeta \leq x < \infty$. The location, ζ and shape parameters, α , of the EXP distribution are estimated by *L*-moment estimators and expressed as:

$$\zeta = \lambda_1 - \alpha \tag{19}$$

$$\alpha = 2\lambda_2 \tag{20}$$

where λ_1 is the *L*-location or mean and λ_2 is the *L*-scale of the distribution (for more details see [19]).

3.1.6. Gumbel

The extreme value type 1 distribution, also called the GUM distribution, is often used to represent a maximum process, for example the maximum rainfall or flood discharge or the lowest stream flow or pollutant concentration. The GUM distribution is used in Finland and Spain as the best-fit distribution for flood data [20]. The pdf and cdf:

$$f(x) = \left(\frac{1}{\alpha}\right) \exp\left[-\left(\frac{x - \beta}{\alpha}\right) - \exp\left(-\left(\frac{x - \beta}{\alpha}\right)\right)\right] \tag{21}$$

$$F(x) = \exp[-\exp(-\frac{x - \beta}{\alpha})] \tag{22}$$

$$\alpha = 0.7797\sigma \tag{23}$$

$$\beta = \mu - 0.5772\alpha \tag{24}$$

where α is the scale parameter and β is the location parameter for the GUM distribution. We can calculate the parameters by any of the estimators easily. Here, the MOM was used to estimate the parameters.

3.1.7. Generalized Extreme Value

A well-known three-parameter distribution for maxima is the GEV. In many European countries, such as Austria, Germany, Italy and Spain, the GEV distribution is recommended to provide the best-fit to flood data [20]. It includes a shape, κ , a scale, α and a location, ζ , parameter. The parameters are estimated by *L*-moment estimators. The pdf and cdf are expressed as:

$$\begin{aligned} f(x) &= \alpha^{-1} \exp[-(1 - k)y - \exp(-y)], \\ y &= -k^{-1} \log\left\{1 - \frac{k(x - \zeta)}{\alpha}\right\}, k \neq 0 \\ y &= \frac{x - \zeta}{\alpha}, k = 0 \end{aligned} \tag{25}$$

$$F(x) = \exp[-\exp - y] \tag{26}$$

$$k = 7.8590c + 2.9554c^2,$$

where $\tag{27}$

$$c = \frac{2}{3 + \tau_3} - \frac{\log 2}{\log 3}$$

$$\alpha = \frac{\lambda_2 k}{(1 - 2^{-k})\Gamma(1 + k)} \tag{28}$$

$$\xi = \lambda_1 - \alpha\{1 - \Gamma(1 + k)\} / k \tag{29}$$

with a range of

$$-\infty < x \leq \xi + \alpha/k; \text{ if } k > 0;$$

$$-\infty < x < \infty; \text{ if } k = 0;$$

$$\xi + \alpha/k \leq x < \infty; \text{ if } k = 0$$

where τ_3 is the *L*-skewness of the distribution [19].

3.1.8. Weibull

The extreme value type 3 distribution or W2 is often used for minimum stream flows and related fields. As this is an analysis of maximum rainfall it is expected that the W2 will not be as suitable as other distributions. Stendinger et al. [21] listed the W2 as a commonly used frequency distribution in hydrology. The pdf and cdf are:

$$f(x) = \left(\frac{k}{\alpha}\right) \left(\frac{x}{\alpha}\right)^{k-1} \exp\left[-\left(\frac{x}{\alpha}\right)^k\right] \tag{30}$$

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^k\right] \tag{31}$$

with the range of $x > 0; \alpha, k > 0$. We estimate the parameters using the MOM estimators.

$$\mu = \alpha\Gamma\left(1 + \frac{1}{k}\right) \tag{32}$$

$$\sigma^2 = \alpha^2 \left\{ \Gamma\left(1 + \frac{2}{k}\right) - \left[\Gamma\left(1 + \frac{1}{k}\right) \right]^2 \right\} \tag{33}$$

where α is the scale parameter and β is the shape parameter for the W2 distribution.

3.1.9. Generalized Pareto

The Generalized Pareto distribution (GP) is applied in precipitation analysis. It is useful for describing events which exceed a specified lower bound, such as rainfall events above a given threshold. The three-parameters are location ξ , scale α and shape k . The parameters are estimated by *L*-moment estimators. The pdf and cdf are:

$$\begin{aligned} f(x) &= \alpha^{-1} \exp[-(1 - k)y], \\ y &= -k^{-1} \log\left\{1 - \frac{k(x - \xi)}{\alpha}\right\}, k \neq 0 \\ y &= \frac{x - \xi}{\alpha}, k = 0 \end{aligned} \tag{34}$$

$$F(x) = 1 - \exp(-y) \tag{35}$$

$$k = (1 - 3\tau_3) / (1 + 3\tau_3) \tag{36}$$

$$\alpha = (1 + k)(2 + k)\lambda_2 \tag{37}$$

$$\xi = \lambda_1 - (2 + k)\lambda_2 \tag{38}$$

with a range of

$$\zeta \leq x \leq \zeta + \alpha/k; \text{ if } k > 0;$$

$$\zeta \leq x < \infty; \text{ if } k \leq 0;$$

3.2. Goodness-of-Fit Test

Goodness-of-fit test statistics are used for checking the validity of a specified or assumed probability distribution model. Three procedures are used for checking the normality: graphical methods (histograms, boxplots and Quantile-Quantile plots (Q-Q plots)), numerical methods (skewness and kurtosis indices) and formal normality tests. There are dozens of tests of normality existing in statistical testing procedures. Among these, Empirical Distribution Function (EDF) tests provide a measure of the discrepancy between the empirical and theoretical distributions [22]. The most well-known EDF tests are the Kolmogorov-Smirnov (K-S) test, the Anderson-Darling (A-D) and the Cramer Von Mises test [23,24]. The K-S and the A-D tests were applied here. As a graphical test, a Q-Q plot was applied. Also, the root mean square error (RMSE) was used to test the best model. A graphical test was also applied to compare the observed and estimated values.

3.2.1. Kolmogorov-Smirnov (K-S) Test

The K-S test statistic, the *supreme* class of EDF statistics, is based on the maximum vertical difference between the theoretical and empirical distributions [25]. The main goal of this test is to compare the empirical cumulative frequency ($S_n(x)$) with the cdf of an assumed theoretical distribution ($F_X(x)$). The maximum difference between $S_n(x)$ and $F_X(x)$ is the K-S test statistic. For a sample size n , the data is rearranged in increasing order, $X_1 < X_2 < \dots < X_n$ and the K-S statistic is assessed for each ordered value:

$$\begin{aligned} S_n(x) &= 0; \text{ if } X < X_1 \\ &= k/n; \text{ if } X_k \leq X < X_{k+1} \\ &= 1; \text{ if } X > X_n \end{aligned} \tag{39}$$

$$D_n = \max|F_X(x) - S_n(x)| \tag{40}$$

$$P(D_n \leq D_n^\alpha) = 1 - \alpha \tag{41}$$

where D_n^α is the critical value, α is the significance level and k is the rank order of the data set.

3.2.2. Anderson-Darling (A-D) Test

The A-D test, the most powerful EDF test, was first introduced by Anderson & Darling [26] to place more weight at the tails of the distribution [27]. In cases with relatively large extremes, it may be expected the A-D test to be more suitable to select the best-fitted model to data maxima. The A-D test statistic, the *quadratic* class of the EDF test statistic, is expressed as A^2 as follows:

$$A^2 = - \sum_{i=1}^n [(2i - 1) \ln F_X(x_i) + \ln[1 - F_X(x_{n+1-i})]] / n - n \tag{42}$$

where $F_X(x_i)$ is the cdf of the proposed distribution at x_i , for $i = 1, 2, \dots, n$. The observed data must be arranged in increasing order, as $x_1 < x_2 < \dots < x_n$.

In the K-S test, both the theoretical and empirical cdfs are relatively flat at the tails of the probability distributions. On the other hand, the A-D test gives more weight to the tails. This can be a more accurate test when the tails of the selected theoretical distribution are the focus of the analysis, as with extreme rainfall.

3.2.3. Root Mean Square Error (RMSE)

The smallest RMSE value indicates the best-fit model of the variate and gives the standard deviation of the model prediction error. It measures the differences between observed and estimated values. The RMSE can be expressed as the following equation:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - X)^2} \quad (43)$$

where x_j denotes the estimated value and X denotes the observed value. The RMSE gives a relatively larger weight to large errors by squaring them.

3.2.4. Graphical Test

A graphical test is one of the most simple and powerful techniques for selecting the best-fit model. Here the Q-Q plot of observed and estimated values is presented. To calculate the plotting position of the non-exceedance probability $p_{i:n}$ various plotting position formulas are used for the specific distributions. Most of the plotting position formulae currently in use have the following form:

$$p_{i:n} = \frac{n - a}{N + 1 - 2a} \quad (44)$$

where n is the rank of the observed value X ($X_{(i)}$ is ascending order), $n = 1, 2, 3, \dots, N$. N is the total number of observed values and a is a constant.

Blom's plotting position ($a = 3/8$), which gives unbiased quantiles, is used here for N, LN2, P3, LP3 distributions; Gringorten's plotting position ($a = 0.44$) is used for EXP, GUM, W2, GP distributions; Cunnane's plotting position ($a = 0.4$) is used for GEV [21,28,29]. To construct the Q-Q plot the observed data, $X_{(i)}$, versus the estimated values, $x(F)$, are plotted, where $x(F)$ is the quantile function, with F determined by the $p_{i:n}$ for the certain probability distribution.

3.3. Return Period

One of the important objectives of frequency analysis is to calculate the recurrence interval or return period. If the variable (x) equal to or greater than an event of magnitude x_T , occurs once in T years, then the probability of occurrence $P(X \geq x)$ in a given year of the variable is:

$$P(x \geq x_T) = \frac{1}{T} \quad (45)$$

$$T = \frac{1}{1 - P(x \leq x_T)} \quad (46)$$

The precipitation amounts associated with the 50- or 100-year average return periods cannot be directly estimated from a data set but must be extracted from the 98th and 99th percentiles, respectively, of a fitted distribution (i.e., $[1 - 0.98^{-\text{year}}]^{-1} = 50$ years and $[1 - 0.99^{-\text{year}}]^{-1} = 100$ years) [15].

4. Results and Discussion

The main goal of this paper is to identify the best-fit probability distribution for every station which yields the maximum monthly rainfall for return periods of 10, 25, 50 and 100 years. These estimates can provide useful guidance for policy making and decision purposes. Knowing the return period of extreme and catastrophic events can be used in determining the risk level of damage by extreme events e.g., rainfall, floods etc.

4.1. Selecting the Best-Fit Results

The compiled data in our study showed the rainfall range varied widely in all the stations. The site's geographical location e.g., longitude, latitude, elevation and the surrounding environmental factors are among the reasons for the variation of rainfall. The south-eastern part of Bangladesh has the highest amount of measured precipitation e.g., Sandwip station has recorded a high of 3001 mm monthly maximum rainfall over the past 35 years. On the other hand, the stations in the region of the north-western part of Bangladesh measured the lowest amount of precipitation (such as Ishardi station, with a 664 mm monthly maximum) compared to the other regions. Table 1 shows the station names, statistical summary results, the best-fit results of K-S, A-D, RMSE and the best-scored results. The best-fit result is taken as the smallest goodness-of-fit result among all probability distributions developed for each station. The table shows the test statistic result of the best-fit distribution type in probabilities for each model selection tool (K-S, A-D, RMSE). All developed probability distributions are ranked with each selection tool (rank 1 is the best-fit). The 3 ranking results are summed to yield a ranking score. The distribution model of each station with the smallest ranking score is selected as the best-fit distribution model.

The Q-Q plots of all stations were created and the distribution fit with observed data was found using the RMSE. As an example, the plots for station Dinajpur are illustrated in Figure 2. The main purpose of the probability distribution fitting here is to most accurately represent the low probability, extreme events on the extreme right tail. By using a Q-Q plot the level of fit on the extreme right tail can be studied [30]. Any perfect matches with observed data points would fall on the [1:1] line. In Figure 2, the GEV distribution matches the data well, with the right tail falling near the [1:1] line. In the case of all other stations there is no common representation pattern of all the distributions. The right tail was often over-estimated or under-estimated. For most of the stations, the GEV, P3, or LP3 distributions perform better in representing the right tail. However, to select the best fit probability model not only the graphical observation but the numerical test is also important.

Table 1. Statistical results and best fit results of the BMD stations.

Sl.no	Station Name	Mean (\bar{X})	SD (σ)	Coeff. of Skew. (γ)	Best-Fit Test Statistic Results			Highest Ranked Distribution (Sum of Ranks)
					By K-S Test	By A-D Test	By RMSE Test	
1	Ambagan	875.2	209.2	0.118	LN2(0.115)	P3(0.305)	N(38.25)	N(5)
2	Barisal	555.0	167.6	1.618	GEV(0.095)	GEV(0.382)	GEV(37.95)	GEV(3)
3	Bhola	600.6	174.8	0.675	GUM(0.081)	LP3(0.280)	P3(28.71)	P3(7)
4	Bogra	483.9	151.1	0.187	LP3(0.066)	LP3(0.213)	LP3(21.05)	LP3(3)
5	Chandpur	549.8	215.3	2.102	GEV(0.088)	GEV(0.136)	LP3(27.51)	GEV(4)
6	Chittagong	897.4	238.0	0.335	GP(0.068)	GP(0.249)	GP(28.14)	GP(3)
7	Chuadanga	433.2	147.8	0.923	P3(0.07)	LN2(0.18)	GUM(16.18)	GEV(7)
8	Comilla	531.2	152.0	0.656	GP(0.082)	P3(0.243)	W2(18.53)	P3(5)
9	Cox's Bazar	1116.9	243.1	0.840	P3(0.108)	P3(0.731)	LN2(67.98)	P3 & LP3(7)
10	Dhaka	511.3	147.1	0.296	LN2(0.098)	LP3(0.326)	W2(20.04)	LP3(6)
11	Dinajpur	574.1	194.6	1.592	GEV(0.102)	GEV(0.432)	GEV(28.82)	GEV(3)
12	Faridpur	450.3	156.2	1.106	GEV(0.083)	GEV(0.233)	P3(27.34)	GEV(7)
13	Feni	791.6	182.8	0.383	GUM(0.113)	P3(0.453)	P3(36.07)	P3(6)
14	Hatiya	858.5	249.7	-0.112	GUM(0.062)	P3(0.226)	N(36.49)	GEV(6)
15	Ishurdi	402.1	134.2	0.433	LN2(0.07)	LP3(0.19)	GP(16.19)	P3 & LP3(9)
16	Jessore	460.4	170.9	1.103	GEV(0.071)	LP3(0.181)	P3(22.27)	LP3(7)
17	Khepupara	766.9	150.7	0.158	GEV(0.075)	GEV(0.263)	GP(19.3)	GEV(5)
18	Khulna	483.2	159.2	0.778	P3(0.066)	P3(0.193)	P3(23.71)	P3(3)
19	Kutubdia	930.9	293.4	1.514	GEV(0.08)	GEV(0.382)	LP3(48.937)	GEV(4)
20	Madaripur	518.6	154.1	0.580	P3(0.07)	LP3(0.18)	W2(18.32)	P3(5)
21	Maijdi Court	807.5	202.2	1.139	GEV(0.066)	GUM(0.25)	GEV(31.24)	GEV(4)
22	Mongla	484.9	143.7	1.929	LP3(0.08)	LP3(0.219)	GEV(32.543)	LP3 & GEV(5)
23	Mymensingh	558.8	164.3	0.142	GP(0.105)	GP(0.36)	GP(25.12)	GP(3)
24	Patuakhali	707.8	175.6	0.156	P3(0.082)	P3(0.236)	P3(22.59)	P3(3)
25	Rajshahi	388.0	123.6	1.107	LN2(0.084)	P3(0.198)	LP3(18.78)	LP3(7)
26	Rangamati	705.1	206.8	0.450	GUM(0.064)	GEV(0.262)	LP3(34.64)	GEV(7)
27	Rangpur	634.0	202.6	1.483	GP(0.067)	P3(0.192)	LP3(22.08)	LP3(7)
28	Sandwip	1096.9	496.5	2.399	GEV(0.054)	GEV(0.125)	LP3(66.31)	LP3(5)
29	Satkhira	449.6	117.6	0.453	W2(0.091)	P3(0.306)	W2(18.2)	P3(7)
30	Sitakunda	890.1	240.6	0.244	P3(0.076)	GEV(0.274)	GP(29.48)	GEV(6)
31	Srimangal	596.2	176.6	0.937	P3(0.05)	LP3(0.135)	GUM(17.37)	GEV & LP3(7)
32	Sydpur	625.8	171.4	-0.103	LN2(0.104)	N(0.212)	P3(21.5)	N(5)
33	Sylhet	979.8	206.6	0.246	GP(0.06)	GEV(0.215)	GP(18.68)	GEV(6)
34	Tangail	444.8	125.1	0.930	P3(0.066)	GEV(0.13)	P3(12.77)	P3(7)
35	Teknaf	1332.4	238.6	1.022	GUM(0.063)	LP3(0.161)	GUM(29.63)	GUM(6)

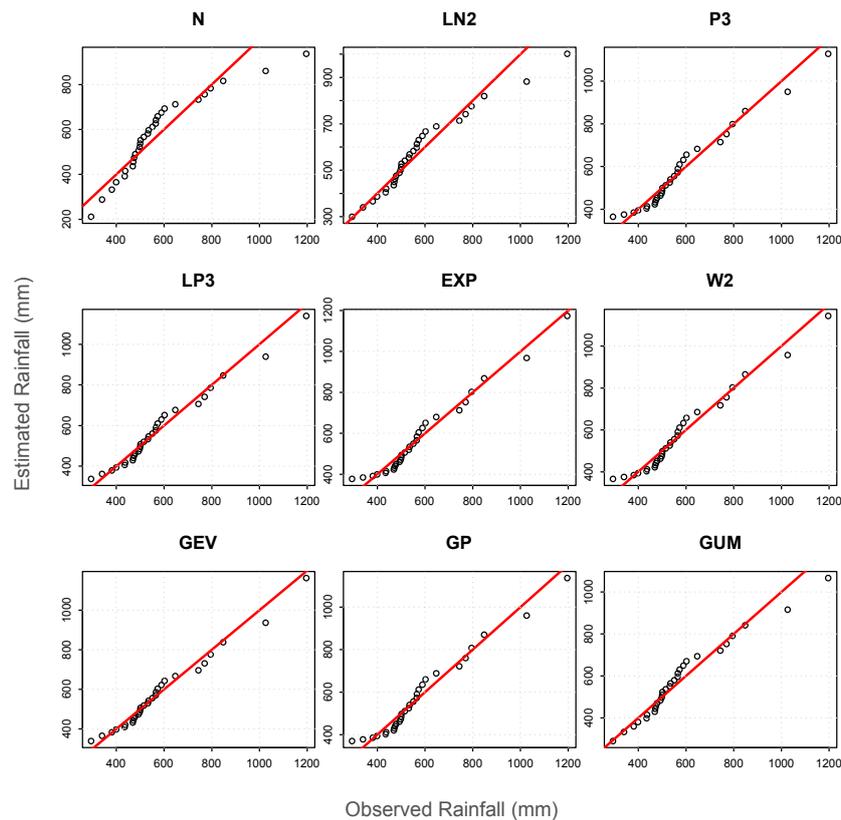


Figure 2. Q-Q plot of all distributions at Station: Dinajpur.

The result of statistical parameters, mean, standard deviation and coefficient of skewness of all stations are included in Table 1. Stations at Hatiya and Sydpur show negative skewness. With distributions tending to have a heavy tail on the right side (in this case more extreme rainfall) the skewness will be positive and relatively high. The absolute value of the skewness reveals the relative magnitude of the deviations from the mean. The coefficients of skewness of 12 stations are greater than 1, thus can be regarded as highly skewed. Eight stations lie within the range of 0.5 to 1, thus are moderately skewed. Fifteen stations are approximately symmetric, lying within the range of -0.5 to 0.5 .

The best-fit test statistic results of K-S, A-D and RMSE are tabulated in Table 1. Among all the stations, it was found that among all distributions, the GEV yielded the most cases of best-fit distributions, while the P3 and LP3 yielded the second and third amount of best-fits respectively. There were no best-fit results from the EXP distribution. The LN2, GUM, GP distributions gave some best fit results among all stations. The other two distributions, W2 and N gave only a few best fit results. The trend of weak fitting of the two-parameter distributions (W2, N, EXP) can be expected as they are less flexible than three-parameter distributions (GEV, P3, LP3). The GP was not among the best performing distributions although it has three-parameters but this distribution is commonly used with peaks over threshold of precipitation or flood values. The weak fit of the W2 is expected as it is often used for minima estimation. However, most stations were best fitted by the GEV, P3 and LP3, all of which have three parameters.

In most of the cases the three goodness of fit tests yielded different types of distribution. To conclude, for selecting the best-fit distribution for every station a ranking system was made, with rank 1 being the best, 2 the second best and so on. The distribution type with the lowest sum of the three rankings is selected as the best fit distribution for each of the stations. Four stations had two distributions with the same value of best score result (a tie between two types). In the best score result,

36% stations are best-fitted by the GEV distribution. Each of the P3 and LP3 yielded the best-fit in 26% of the stations.

4.2. Return Period Results

In Figure 3, 10-year, 25-year, 50-year and 100-year rainfall return levels of all stations are illustrated. The horizontal axis, ranging from 1 to 35, represents the station numbers which are given in Table 1. The vertical axis represents the maximum monthly rainfall. There are four parts in this figure. Return periods of 10 years, 25 years, 50 years and 100 years are represented vertically from top to bottom. The return levels of the best-fit distributions for each station are shown here. As an example, for the station Dinajpur, station number 11, which is best-fitted by the GEV distribution, the rainfall amounts of 10, 25, 50 and 100 year return periods are 809 mm, 997 mm, 1161 mm and 1347 mm respectively.

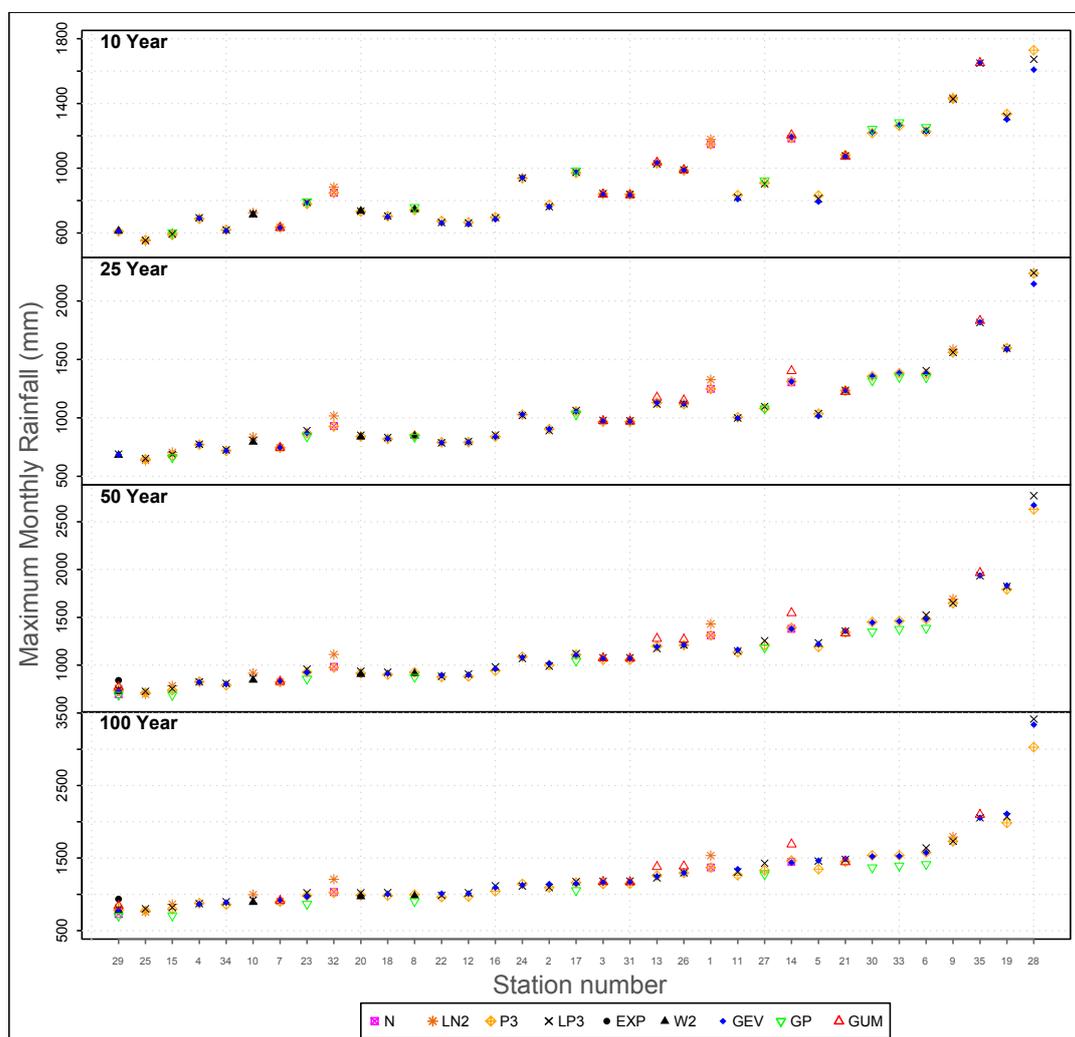


Figure 3. Return period results of all stations (serial number from x axis corresponds to the respective serial number of Table 1). The y axis represents maximum monthly rainfall in mm. From top to bottom the sections represent 10-year, 25-year, 50-year, 100-year return periods. Only the return period of the best-fit distributions is shown here.

The confidence interval, a statistical estimate of the range given a chosen level of probability within which a true value can be expected to fall, is essential in risk analysis as well as in the design process. The upper and lower boundary values of the confidence interval are called confidence limits.

In Figure 4, the rainfall return levels of 1 to 100 years with 95% confidence intervals are illustrated for 3 stations. Here, the sample data size is relatively small compared to the return levels. Anticipating up to 100-year return periods with only 30 years of data available unavoidably yields much lower confidence in the results as the return period increases. It is apparent for these cases that the GEV tends to yield the narrowest range of 95% confidence with the P3 nearly as good, while the LP3 yields a much wider range. If the sample size is small the confidence interval is quite wide. As the sample size increases, that is, as more weather data is recorded over time, the width of the confidence interval around the estimated rainfall magnitude will likely be reduced as improved distributions are found.

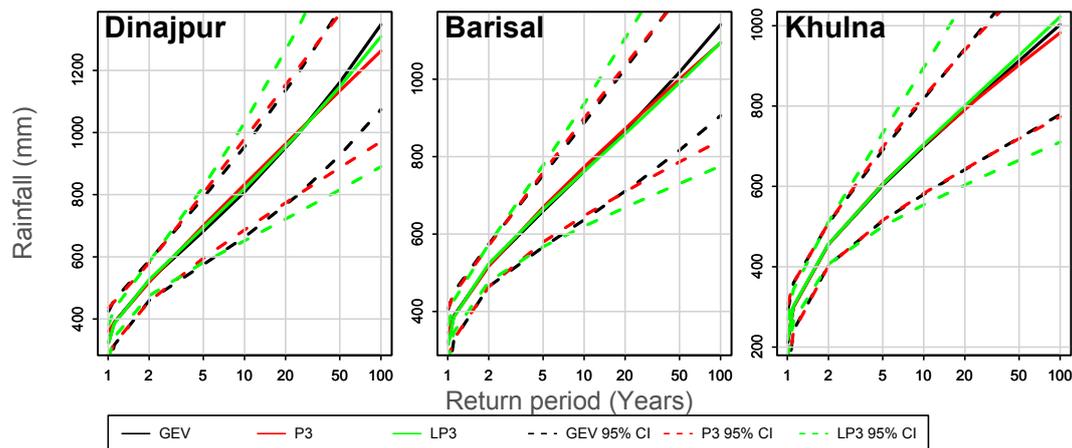


Figure 4. Rainfall return level estimations (solid lines) and 95% confidence intervals (dashed lines) of the top three fitted distributions—GEV, P3 and LP3 of three stations, as an example.

In the south-eastern region, for example near the area of the station name Kutubdia, Cittagong, Cox’s Bazar, Teknaf, more intense rainfall occurred than in the other regions. During the monsoon time landslides occur frequently in this area causing deaths and damage of assets. Extreme amounts of rainfall in the piedmont of hilly areas is the main source of flash floods. Landslides often occur in the areas composed of unconsolidated rocks [31]. The Daily Star, the leading daily newspaper, reported on 13 June 2017 that a landslide in the hilly area caused 133 deaths [32]. Sarker & Rashed [31] also mentioned that slope saturation by water is the main cause of landslides. Saturation occurs due to excessive rainfall.

The information of return periods can help guide policy makers and planners. They can more easily make their decisions for a given span of time using the expected maximum monthly rainfall as determined by the return period. Taking the example of Barisal as shown in Figure 4, a risk level plan using the GEV distribution for a 50-year span would anticipate that the expected maximum monthly rainfall is 1018 mm on average, with a 95% confidence that it will range from 816 mm to 1220 mm. This can then be used to develop more detailed risk and damage estimates (i.e., flooding levels, infrastructure damage etc.).

5. Conclusions

The rainfall patterns vary across all locations. The site’s geographical location and surrounding environment are important factors in the variation of the rainfall pattern. The south-eastern part measures the highest amount of precipitation while the north-western part has the lowest amount of rainfall.

In terms of selecting the best-fit distribution to model the maximum monthly rainfall for a certain location, the result varied significantly. Parameters of the distributions were estimated by the method of moments and *L*-moments estimators. The fit results were calculated by the K-S, A-D and RMSE as

well as a graphical procedure, the Q-Q plot. The return period, which is important in determining long-term risks, was also calculated.

In some cases, all three test statistics yielded the same best-fit distribution type. However, for most of the locations the testing statistics yielded different best-fit distribution types. Thus, a scoring system was used to determine the best fit distribution for each location. The best-fit result of each station was taken as the distribution with the lowest sum of the rank scores from each test statistic. GEV, P3 and LP3 showed the larger number of best-fit results compared to other distributions. In the best score results, GEV fitted 36% of the stations, while P3 and LP3 each fitted 26% of the stations. This result aligns well with results in the literature. The GEV was also the most common best fit in the research by Khudri et al. [2] on annual extreme values of precipitation in Bangladesh. Also, the GEV and P3 were the most common best fits for maximum monthly rainfall in Iran as found by Eslamin et al. [7].

The more practical result of this paper was the return period calculation for all locations and all distributions. The presentation of all distributions gives a clear inspection of the range of projected return periods, including the best fit (most likely) result. Here, 10-year, 25-year, 50-year and 100-year return periods of rainfall return levels were illustrated. A policy maker making a risk analysis for a 50 year plan could use the 50 year return period result of projected maximum monthly rainfall to determine flooding risks, damage projections, etc. The 100 year return period results are high intensity but low probability which has an inverse relationship with 10 year return period results. It depends on the policy makers to choose the duration and risk level.

The results of this study can be used to develop better models of risk and damage from extreme rainfall events and flooding. These estimates can help policy makers to create initiatives with the result of saving lives and property. In addition, Knowledge of the pattern of extreme rainfall will be useful in various disciplines such as the environment, agriculture, construction and building structures.

Acknowledgments: The authors acknowledge the financial support of the general research funding of the Osaka City University Department of Human Life Science.

Author Contributions: Md Ashraf Alam and Kazuo Emura conceived the research theme. Md Ashraf Alam performed the data analysis, the interpretation of results and wrote the paper. Craig Farnham contributed to writing the paper as well as evaluated the results, chose and advised on calculation methods with Kazuo Emura. Jihui Yuan contributed to documenting the results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Katz, R.W.; Brown, B.G. Extreme events in a changing climate: Variability is more important than averages. *Clim. Chang.* **1992**, *21*, 289–302. [[CrossRef](#)]
2. Khudri, M.M.; Sadia, F. Determination of the Best Fit Probability Distribution for Annual Extreme Precipitation in Bangladesh. *Eur. J. Sci. Res.* **2013**, *103*, 391–404.
3. Mandal, S.; Choudhury, B.U. Estimation and prediction of maximum daily rainfall at Sagar Island using best fit probability models. *Theor. Appl. Clim.* **2015**, *121*, 87–97. [[CrossRef](#)]
4. Kumar, A. Prediction of annual maximum daily rainfall of Ranichauri (Tehri Garhwal) based on probability analysis. *Ind. J. Soil Conserv.* **2000**, *28*, 178–180.
5. Singh, R.K. Probability analysis for prediction of annual maximum rainfall of Eastern Himalaya (Sikkim mid hills). *Ind. J. Soil Conserv.* **2001**, *29*, 263–265.
6. Amin, M.T.; Rizwan, M.; Alazba, A.A. A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan. *Open Life Sci.* **2016**, *11*, 432–440. [[CrossRef](#)]
7. Eslamian, S.S.; Feizi, H. Maximum Monthly Rainfall Analysis Using L-Moments for an Arid Region in Isfahan Province, Iran. *J. Appl. Meteorol. Climatol.* **2007**, *46*, 494–503. [[CrossRef](#)]
8. Bhakar, S.R.; Iqbal, M.; Devanda, M.; Chhajed, N.; Bansal, A.K. Probability analysis of rainfall at Kota. *Ind. J. Agric. Res.* **2008**, *42*, 201–206.
9. Lee, C. Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in Southern Taiwan. *Crop Environ. Bioinf.* **2005**, *2*, 31–38.

10. Ogunlela, A.O. Stochastic Analysis of Rainfall Evens in Ilorin, Nigeria. *J. Agric. Res. Dev.* **2001**, 39–50. [[CrossRef](#)]
11. Kwaku, X.S.; Duke, O. Characterization and frequency analysis of one day annual maximum and two to five consecutive days maximum rainfall of Accra, Ghana. *ARPN J. Eng. Appl. Sci.* **2007**, 2, 27–31.
12. Olofintoye, O.O.; Sule, B.F.; Salami, A.W. Best fit Probability distribution model for peak daily rainfall of selected Cities in Nigeria. *N. Y. Sci. J.* **2009**, 2, 1–12. [[CrossRef](#)]
13. Sen, Z.; Eljadid, A.G. Rainfall distribution functions for Libya and Rainfall Prediction. *Hydrol. Sci. J.* **1999**, 44, 665–680. [[CrossRef](#)]
14. Arora, K.; Singh, V.P. A comparative evaluation of the estimators of the log Pearson type 3 distribution. *J. Hydrol.* **1989**, 105, 19–37. [[CrossRef](#)]
15. Wilks, D.S. Comparison of three-parameter probability distributions for representing annual extreme and partial duration precipitation series. *Water Resour. Res.* **1993**, 29, 3543–3549. [[CrossRef](#)]
16. Ali, A. Climate change impacts and adaptation assessment in Bangladesh. *Clim. Res.* **1999**, 12, 109–116. [[CrossRef](#)]
17. Ahsan, M.N.; Chowdhary, M.A.M.; Quadir, D.A. Variability and trends of summer monsoon rainfall over Bangladesh. *J. Hydrol. Meteorol.* **2010**, 7, 1–17. [[CrossRef](#)]
18. Markovic, R.D. *Probability Functions of Best Fit to Distributions of Annual Precipitation and Runoff*; Colorado State University Hydrology Paper No. 8; Colorado State University: Fort Collins, CO, USA, 1965.
19. Hosking, J.R.M.; Wallis, J.R. *Regional Frequency Analysis: An Approach Based on L-Moments*, 1st ed.; Cambridge University Press: Cambridge, UK, 1997; pp. 14–43, ISBN 978-0-521-43045-6.
20. Salinas, J.L.; Castellarin, A.; Kohnová, S.; Kjeldsen, T.R. Regional parent flood frequency distributions in Europe—Part 2: Climate and scale controls. *Hydrol. Earth Syst. Sci.* **2014**, 18, 4391–4401. [[CrossRef](#)]
21. Stender, J.R.; Vogel, R.M.; Foufoula-Georgiou, E. Frequency Analysis of Extreme Events. In *Handbook of Hydrology*, 1st ed.; Maidment, D.A., Ed.; McGraw-Hill: New York, NY, USA, 1993; Chapter 18, ISBN 0070397325.
22. Dufour, J.M.; Farhat, A.; Gardiol, L.; Khalaf, L. Simulation-based Finite Sample Normality Tests in Linear Regressions. *Econ. J.* **1998**, 1, 154–173. [[CrossRef](#)]
23. Arshad, M.; Rasool, M.T.; Ahmad, M.I. Anderson Darling and Modified Anderson Darling Tests for Generalized Pareto Distribution. *Pak. J. Appl. Sci.* **2003**, 3, 85–88.
24. Seier, E. Comparison of Tests for Univariate Normality. *InterStat Stat. J.* **2002**, 1, 1–17.
25. Conover, W.J. *Practical Nonparametric Statistics*, 3rd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1999; pp. 428–433, ISBN 0471160687.
26. Anderson, T.W.; Darling, D.A. A Test of Goodness-of-fit. *J. Am. Stat. Assoc.* **1954**, 49, 765–769. [[CrossRef](#)]
27. Farrel, P.J.; Stewart, K.R. Comprehensive study of tests for normality and symmetry: Extending the Spiegelhalter test. *J. Stat. Comput. Simul.* **2006**, 76, 803–816. [[CrossRef](#)]
28. Vogel, R.M.; Kroll, C.N. Low-Flow Frequency Analysis Using Probability-Plot Correlation Coefficients. *J. Water Resour. Plan. Manag.* **1989**, 115, 338–357. [[CrossRef](#)]
29. Cunnane, C. Unbiased plotting positions—A review. *J. Hydrol.* **1978**, 37, 205–222. [[CrossRef](#)]
30. Graedel, T.E.; Kleiner, B. Exploratory analysis of atmospheric data. In *Probability, Statistics and Decision Making in the Atmospheric Sciences*, 1st ed.; Murphy, A.H., Katz, R.W., Eds.; Westview: Boulder, CO, USA, 1985; pp. 1–43, ISBN 0865311528.
31. Sarker, A.A.; Rashid, A.K.M.M. Landslide and Flashflood in Bangladesh. In *Disaster Risk Reduction Approaches in Bangladesh*; Shaw, R., Mallick, F., Islam, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 165–189.
32. Landslide in Hilly Areas: Death Toll Rises to 133—The Daily Star. Available online: <http://www.thedailystar.net/country/16-killed-bandarban-rangamati-landslides-1419571> (accessed on 19 June 2017).

