ALTERNATIVE PROBABILITY DISTRIBUTION FUNCTIONS FOR DROUGHT CALCULATION USING THE STANDARDIZED PRECIPITATION INDEX

Md Fazzle Rabby*[1] and Sajal Kumar Adhikary[2]

[1] *Department of Civil Engineering, Bangladesh Army University of Science and Technology (BAUST), Saidpur, Bangladesh, e-mail: fazzlece@baust.edu.bd*

[2] *Department of Civil Engineering, Khulna University of Engineering & Technology (KUET), Khulna-9203, Bangladesh, e-mail: sajal@ce.kuet.ac.bd*

***Corresponding Author**

ABSTRACT

Drought is a major threat to food and water security, especially in regions that are prone to low and erratic precipitation. In order to monitor and assess drought conditions, various drought indices have been developed, among which the standardized precipitation index (SPI) is the most widely used. However, the choice of probability distribution for calculating SPI may affect its accuracy and reliability. The current study is an attempt to evaluate the performance of four probability density functions: gamma, lognormal (LN), Weibull, and generalized extreme value (GEV) distributions, for fitting the precipitation data from three meteorological stations, namely Bogra, Dinajpur, and Rangpur, in the northwestern region of Bangladesh. The distributions are fitted to the data using maximum likelihood estimators and compared using Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Square (Chi-Sq) tests. The best-fitted distribution function is selected on the basis of the lowest combined ranking score for each case. The results show that out of 108 cases, GEV distributions provide the best fit for 64% of cases, while the LN, Weibull, and Gamma distributions are only suitable for 9%, 8%, and 4% of the cases, respectively. The analysis highlights that the GEV distribution has a significant superiority over other distributions, regardless of the locations and time periods under consideration. The fact that GEV is identified as the most appropriate fit across all time scales implies that it captures the distribution of extreme values in the precipitation data well. The stability of SPI values with the GEV distribution as time scales increase suggests that these distributions may be considered more reliable in representing longer-term drought conditions without overemphasizing extreme conditions. Furthermore, it has also been seen that the lognormal and Weibull distributions can't accurately show extreme events that happen over shorter periods of time because their SPI values are much higher than those of the gamma and generalised extreme value distributions. However, when considering longer time periods, both distributions have shown the potential to accurately represent severe events, indicating a decrease in the tendency to overestimate or underestimate dry and wet periods. Nevertheless, the current study ultimately recommends using the GEV distribution to compute the Standardized Precipitation Index (SPI) in the study area. This approach will enhance drought prediction and early warning systems and provide more precise data for planning, designing and executing drought mitigation strategies.

Keywords: *Drought, SPI, probability distribution function, maximum likelihood,*

1. INTRODUCTION

The insufficiency of rainfall within a certain geographic region may be described as a drought, a topic of considerable importance in the present era marked by an increased occurrence of climate change phenomena, including in Bangladesh. Since independence, the nation has experienced high droughts, impacting agricultural output, human welfare, animal populations, land resources, public health, and employment. Effective management strategies rely on comprehensive data on drought magnitude, intensity, and duration, which can be obtained using drought indices (Mahmoudi et al., 2019).

Drought indices are of paramount importance in the observation and evaluation of drought conditions owing to their capacity to facilitate the complicated relationship among numerous climatic variables. (Angelidis et al., 2012). Various indices are frequently employed for the monitoring of different types of droughts. For instance, while the palmer drought severity index (PDSI) (Palmer, 1965) and the standardized precipitation index (SPI) (McKee et al., 1993) are used to assess precipitation-derived meteorological droughts, the standardized runoff index (SRI) is utilized to study hydrological droughts that are based on runoff or streamflow (Shukla & Wood, 2008). On the other hand, the standardized soil moisture index (SSI) and agricultural standardized precipitation index (aSPI) are widely employed for the monitoring of agricultural droughts (Hao & AghaKouchak, 2013; Tigkas et al., 2019). Furthermore, the standardized supply and demand water index (SSDWI) is employed to assess socio-economic droughts (Zhou et al., 2022), among other applications.

The standardized precipitation index (SPI) is a popular choice among researchers due to its simplicity, spatial consistency, probabilistic nature, and adaptability to user interests (Edossa et al., 2010), especially for low data requirements. SPI is calculated using various distribution functions, with the gamma distribution being the most commonly used. However, the suitability of theoretical distributions varies across geographical regions (Cindric et al., 2012; Hong et al., 2013; Vergni et al., 2017). For example, Pearson type III in America is more suitable (Guttman, 1999), while Weibull-type distributions offer better fits in Europe (Sienz et al., 2012). In Guadiana (Portugal), the log-normal distribution yields nearly identical results to gamma distributions (Angelidis et al., 2012). The generalized normal distribution also performs better in Brazil (Blain & Meschiatti, 2015). According to the study conducted by Bhakar et al. (2008), it was determined that the Gumbel distribution provided the most accurate match for modelling monthly maximum rainfall in India. In another study, Amin et al. (2016) used annual maximum precipitation data obtained from a daily sample. Their analysis revealed that the Log-Pearson type-III distribution exhibited the most optimal fit for the northern parts of Pakistan.

A study conducted in Bangladesh (Khudri & Sadia, 2013) discovered that the generalized extreme value and generalized gamma four parameter distributions exhibited the greatest degree of compatibility for around 50% of the assessed stations. In contrast, none of the other distributions have consistently shown compatibility with the other stations. Based on the study conducted by Mandal & Choudhury (2015), it was observed that normal distributions yielded the most precise alignment with the data for the annual, post-monsoon, and summer seasons. On the other hand, it was noted that the pre-monsoon, monsoon, and winter seasons had the highest degree of conformity with the lognormal, Weibull, and Pearson type V probability distributions, respectively. A recent study conducted by Rabby & Adhikary (2022) in the country's northwestern part concluded that log-normal distribution might be a viable alternative to traditional gamma distribution for drought assessment. Therefore, the careful selection of distribution functions is of the utmost importance in ensuring a precise estimation of the SPI.

Up to the present time, several studies in different parts of Bangladesh have assessed precipitation data using various probability distributions. However, none of these studies have considered including these distributions in SPI-based drought assessment. In order to fill this gap, the objective of the current study is to identify the most suitable probability density functions for accurately characterizing the precipitation data as well as computing drought events in the study region. The identified probability density functions will then be incorporated into the calculation of the SPI to enhance the representation of the data across different time series and scales. The study finally aims to assess the viability of using alternative probability distribution functions in addition to the standard gamma distributions to calculate SPI-based drought events.

2. MATERIALS AND METHODS

2.1 Study Area and Precipitation Data

Bangladesh, being a South Asian country, has approximately 170 million inhabitants and 148,460 km^2 of land area (Hosain & Amin, 2023). Droughts are more common in the northwestern areas (Dey et al., 2012; Mondol et al., 2016). The northwest region of Bangladesh receives 1,400 mm of rainfall annually, while the northeast part receives around 4,400 mm. The monsoon season accounts for 75% of total rainfall (Alamgir et al., 2015). The northwestern region of Bangladesh is equipped with six meteorological stations, namely Bogra, Dinajpur, Ishurdi, Rajshahi, Rangpur, and Saidpur, which are operated by the Meteorological Department of Bangladesh. For the current research, three stations in the region, namely Bogra, Dinajpur, and Rangpur, have been chosen. Fig. 1 shows the study area with the locations of selected stations.

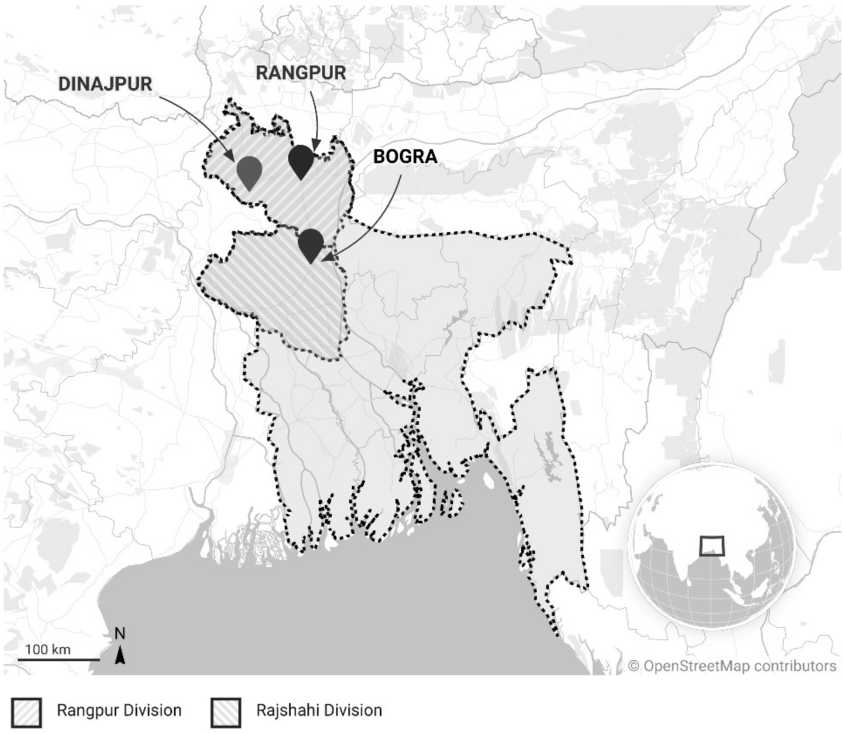


Figure 1: Meteorological stations considered for this study

Monthly rainfall data for the selected stations were collected from the Bangladesh Meteorological Department (BMD). Table 1 presents the details of the selected meteorological stations used in this study.

Table 1: Details of the selected meteorological stations

Station No.	Name of the Stations	Area (km^2)	Latitude	Longitude	Data Type	Period
1	Bogra	2898.68	24° 51′ N	89° 22′ E	Rainfall	1975-2019
2	Dinajpur	3444.30	25° 39′ N	88° 41′ E	Rainfall	1975-2019
3	Rangpur	2400.56	25° 44′ N	89° 14′ E	Rainfall	1975-2019

2.2 Calculation of Standardized Precipitation Index (SPI)

SPI was first proposed by McKee et al. (1993) and is the most widely adopted index for drought estimation among the researchers today. It is basically based on the probability of precipitation for multiple time scales, e.g., one-, three-, six-, nine-, and twelve-months, etc., and is usually calculated by fitting the cumulative precipitation with an appropriate probability density function to characterize the lack of precipitation. The calculation procedure of SPI is detailed in the following:

For a given time series of precipitation, the cumulative precipitation for a certain time scale at the n-*th* month during different years is calculated by:

$$x_a^n(t) = \sum_{i=j-a+1}^{j} x(i),\ j = 12(t-1)+n \quad (1)$$

Where, x is the monthly precipitation vaule of N years, t is the annual index (from 1 to N), and m is a given month (January, February,..., December)

A probability density function is then used to fit the cumulative precipitation series of a particular distribution. Table 2 provides a list of the types of distributions in which cumulative precipitation data is fitted to their corresponding probability density functions.

Table 2: Distributions used to compute Standardized Precipitation Index (SPI)

Distribution	**Probability Density Function (PDF)**	**Cumulative Distribution Function (CDF)**
Gamma	$f(x)=\frac{x^{k-1}e^{-x/\theta}}{\theta^k \Gamma(k)}$	$F(x)=\frac{\gamma(k,\frac{x}{\theta})}{\Gamma(k)}$
Generalized Extreme Value (GEV)	$f(x)=\frac{1}{\sigma}\left\{\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\left(\frac{1+\xi}{\xi}\right)} e^{-\left\{\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\left(\frac{1}{\xi}\right)}\right\}}\right\}$	$F(x)=\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right\}$
Lognormal	$f(x)=\frac{1}{\sqrt{2\pi}\sigma x}\exp\left\{\frac{-(lnx-\mu)^2}{2\sigma^2}\right\}$	$F(x)=\left(\frac{lnx-\mu}{\sigma}\right)$
Weibull	$f(x)=\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}\exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}$	$F(x)=1-\exp\{-(x)^k\}$

The cumulative probability of a precipitation event occurring over a specified time period and month is then calculated, and the SPI value is achieved by converting the CDF to a standard normal distribution using Eq. (2).

$$x_k = \varphi^{-1}(k) \quad (2)$$

Here, the quantile of the cumulative probability k is represented as x_k, whereas the inverse function of the cumulative distribution function (CDF) for a normal distribution is indicated as φ^{-1}. The drought category based on the Standardized Precipitation Index (SPI) is shown in Table 3.

Table 3: Drought category based on SPI values

SPI Value	**Classification**
-0.99 to 0.00	Mild drought
-1.49 to -1.00	Moderate drought

-1.99 to -1.50	Severe drought
≤2.00	Extreme drought

2.3 Parameters Estimation by Maximum Likelihood

Estimation of the distribution parameters are essential, and to obtain that a number of approaches, such as the method of moments, L-moments, and maximum likelihood (ML), have been established in the past. Researchers continue to choose the maximum likelihood estimator (MLE) owing to its superior performance across a range of probability distributions (Ashkar & Nwentsa, 2007; Park et al., 2009). The MLE is a statistical technique used to ascertain optimal parameter values for a given model. The parameter values are determined in a manner that optimizes the probability of the observed data being generated by the process represented by the model.

Let a random sample of size n, denoted as $X_1, ..., X_n$, drawn from the random variable X. The density function (or probability function) of X, denoted as $f(x|\theta)$, depends on a parameter θ that belongs to the parameter space Θ. The likelihood function of θ that corresponds to the observed random sample is expressed as:

$$L(\theta;x)=\prod_{i=1}^{n} f(x_i|\theta). \tag{3}$$

The ML estimator of θ is the value $\hat{\theta} \in \Theta$ that maximizes the likelihood function $L(\theta;x)$ (Bolfarine & Sandoval 2001). In practice, the logarithm of $L(\theta;x)$ is generally used.

$$l(\theta;x)=\log L(\theta;x)=\sum_{i=1}^{n} \ln f(x_i|\theta). \tag{4}$$

The ML estimates $\hat{\theta}$ are those that maximize $l(\theta;x)$, and $\hat{\theta}$ is called ML estimator (MLE) of θ.

2.4 Goodness-of-Fit Test

Goodness-of-fit is a statistical method that evaluates the compatibility between observed data and a theoretical distribution from a normal population. It helps in predicting future trends and patterns in decision-making processes. Though there are various methods available to carry out goodness-of-fit tests, the Chi-square test, the Kolmogorov-Smirnov test, and the Anderson-Darling test are quite famous (Arshad et al., 2003; Seier, 2002). These tests are used to determine the optimal probability distribution, with a significance level of $\alpha=0.05$. In this research, the null hypothesis (H_0) assumes that the observed monthly rainfall data corresponds to a prescribed distribution, while the alternative hypothesis (H_1) assumes that the data does not adhere to the prescribed distribution.

2.4.1 Kolmogorov-Smirnov Test

Kolmogorov-Smirnov (K-S) test (Karson, 1968) is a non-parametric statistical method to check whether a sample is from a specific distribution (Melesse et al., 2010) within a population and uses a null and alternative hypothesis as well as an alpha level of significance. This test is based on the empirical distribution function (ECDF). Given N ordered data points $Z_1, Z_2, ..., Z_N$, the ECDF is defined as

$$E_N=\frac{n(k)}{N} \tag{5}$$

Where $n(k)$ is the number of points less than Z_i and the Z_i represents a set of values arranged in ascending order. This is a step function that causes an increase of $1/N$ at the value of each ordered data point. The K-S test statistic is defined as follows:

$$D = \max_{1 \le i \le N} \left[F(Z_i) - \frac{i-1}{N}, \frac{i}{N} - F(Z_i) \right] \quad (6)$$

Where F is the theoretical cumulative distribution function of the distribution being tested and D is the test statistic. If D is less than the critical value, it accepts the null hypothesis. This evidently demonstrates that the hypothesis regarding the distribution form is accepted.

2.4.2 Anderson-Darling Test

The Anderson-Darling (A-D) test (Stephens, 1974) is a modified version of the Kolmogorov-Smirnov (K-S) test that assigns more significance to the extreme values in the distribution. The Kolmogorov-Smirnov (K-S) test exhibits more sensitivity towards discrepancies that may arise in proximity to the central region of the distribution, while the Anderson-Darling (A-D) test demonstrates higher sensitivity towards deviations detected in the tails. The Anderson-Darling test statistic is formally defined as follows:

$$A^2 = -N - \frac{1}{N} \sum_{i=1}^{N} (2i-1) \left[\ln F(X_i) + \ln \left(1 - F(X_{N-i+1}) \right) \right] \quad (7)$$

F and X_i represents the cumulative distribution function of the given distribution and ordered data respectively. The determination of critical values for the Anderson-Darling test is conditional on the particular distribution under examination. The test being conducted is a one-sided test, where the hypothesis on the distribution conforming to a certain form is deemed invalid if the test statistic, A exceeds the critical value.

2.4.3 Chi-square Test

The Chi-square test (Ridgman, 1990) assumes a large sample size, allowing the chi-square distribution to accurately approximate the test statistic's distribution, and is defined as:

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} \quad (8)$$

Where,
O_j = observed frequency in 'j' observations (1, 2, ……., k)
E_j = expected frequency in 'j' observations (1, 2, ……., k)
Calculated by $E_j = F(X_2) - F(X_1)$
F = the CDF of the probability distribution being tested.

The observed number of observation (k) in interval 'j' is computed from equation given below:

$$K = 1 + \log_2 n \quad (9)$$

Where, n = sample size.

This equation is specific to continuous sample data and determines if a sample is from a population with a specific distribution.

3. RESULTS AND DISCUSSION

One of the main goals of this current study is to identify the best-fit probability distribution function for each station at the specified time scales for each month of the year using monthly precipitation data extracted from the selected three meteorological stations located in the northwest region of Bangladesh from 1975 to 2019.

Table 4: Best-fit results of the selected stations at multiple time scales

Stn	SL No.	Month	6 Month Time Scale				12 Month Time Scale				24 Month Time Scale			
			K-S	A-D	Chi-Sq	Sum of Ranks	K-S	A-D	Chi-Sq	Sum of Ranks	K-S	A-D	Chi-Sq	Sum of Ranks
Bogra	1	Oct	GEV (0.056)	GEV (0.122)	GEV (1.08)	**GEV (3)**	W (0.075)	GEV (0.283)	GEV (1.47)	**GEV (4)**	LN (0.084)	GEV (0.395)	LN (1.12)	**LN (5)**
	2	Nov	GEV (0.068)	GEV (0.216)	GEV (0.432)	**GEV (3)**	GEV (0.063)	GEV (0.215)	GEV (1.59)	**GEV (3)**	G (0.079)	GEV (0.373)	G (1.41)	**G (5)**
	3	Dec	GEV (0.072)	GEV (0.208)	GEV (1.13)	**GEV (3)**	W (0.075)	GEV (0.232)	LN (1.78)	**GEV (5)**	LN (0.080)	GEV (0.358)	GEV (1.33)	**GEV (5)**
	4	Jan	GEV (0.067)	GEV (0.377)	G (1.23)	**GEV & LN (6)**	W (0.076)	GEV (0.221)	W (1.19)	**W (4)**	LN (0.074)	GEV (0.335)	LN (0.592)	**LN (5)**
	5	Feb	GEV (0.063)	GEV (0.189)	W (0.417)	**GEV (5)**	GEV (0.078)	GEV (0.267)	G (0.685)	**GEV & G (6)**	LN (0.071)	GEV (0.341)	LN (0.463)	**LN (5)**
	6	Mar	W (0.071)	GEV (0.18)	G (0.798)	**GEV (5)**	GEV (0.076)	GEV (0.31)	G (1.78)	**GEV & G (5)**	LN (0.065)	GEV (0.282)	G (0.528)	**LN (5)**
	7	Apr	W (0.082)	GEV (0.237)	LN (0.775)	**GEV (6)**	GEV (0.054)	GEV (0.137)	GEV (0.932)	**GEV (3)**	GEV (0.054)	GEV (0.19)	W (0.709)	**GEV (5)**
	8	May	LN (0.071)	GEV (0.26)	G (1.41)	**LN (5)**	GEV (0.067)	GEV (0.271)	G (1.13)	**GEV (5)**	GEV (0.073)	GEV (0.227)	GEV (0.454)	**GEV (3)**
	9	Jun	LN (0.09)	GEV (0.337)	LN (1.98)	**LN (5)**	GEV (0.088)	GEV (0.441)	GEV (4.75)	**GEV (3)**	G (0.061)	GEV (0.254)	G (0.403)	**G (4)**
	10	Jul	W (0.083)	GEV (0.267)	G (1.93)	**W (5)**	GEV (0.053)	GEV (0.164)	G (0.178)	**GEV (5)**	W (0.056)	G (0.257)	GEV (1.13)	**GEV (5)**
	11	Aug	GEV (0.061)	GEV (0.216)	W (0.608)	**GEV (6)**	W (0.048)	GEV (0.231)	W (0.672)	**GEV & W (5)**	W (0.106)	W (0.527)	W (0.881)	**W (3)**
	12	Sep	GEV (0.08)	GEV (0.271)	GEV (0.972)	**GEV (3)**	W (0.081)	GEV (0.233)	G (0.719)	**GEV (5)**	GEV (0.082)	GEV (0.363)	LN (2.38)	**GEV (4)**
Dinajpur	1	Oct	W (0.072)	GEV (0.319)	W (1.56)	**W (4)**	W (0.069)	GEV (0.285)	G (1.25)	**GEV & G (6)**	GEV (0.072)	GEV (0.179)	GEV (0.76)	**GEV (3)**
	2	Nov	GEV (0.074)	GEV (0.295)	G (0.835)	**GEV (4)**	W (0.064)	GEV (0.253)	GEV (0.249)	**GEV (4)**	GEV (0.07)	GEV (0.174)	GEV (0.103)	**GEV (3)**
	3	Dec	W (0.116)	GEV (0.495)	GEV (2.42)	**GEV (4)**	W (0.068)	GEV (0.301)	G (1.3)	**W (5)**	GEV (0.074)	GEV (0.176)	GEV (0.513)	**GEV (3)**
	4	Jan	W (0.103)	GEV (0.423)	LN (4.04)	**LN (5)**	GEV (0.068)	GEV (0.28)	G (1.02)	**GEV & G (6)**	GEV (0.059)	GEV (0.168)	GEV (0.385)	**GEV (3)**
	5	Feb	W (0.076)	G (0.391)	W (0.228)	**G & W (6)**	W (0.069)	GEV (0.224)	G (0.567)	**W (6)**	W (0.079)	GEV (0.183)	G (0.771)	**GEV & G (6)**
	6	Mar	LN (0.069)	GEV (0.17)	G (1.57)	**GEV (5)**	W (0.061)	GEV (0.232)	G (0.238)	**W (6)**	GEV (0.072)	GEV (0.201)	LN (0.791)	**GEV (5)**
	7	Apr	GEV (0.053)	GEV (0.125)	G (0.453)	**GEV (4)**	GEV (0.085)	W (0.341)	G (2.41)	**GEV (5)**	GEV (0.06)	GEV (0.225)	G (0.484)	**GEV & G (5)**
	8	May	GEV (0.079)	GEV (0.263)	GEV (1.26)	**GEV (3)**	W (0.067)	GEV (0.244)	GEV (0.629)	**GEV (4)**	GEV (0.063)	GEV (0.176)	GEV (3.64)	**GEV (3)**
	9	Jun	GEV (0.075)	GEV (0.18)	W (0.662)	**GEV & W (5)**	GEV (0.064)	GEV (0.182)	W (1.80)	**GEV (4)**	GEV (0.053)	GEV (0.125)	LN (0.349)	**GEV (5)**
	10	Jul	GEV (0.075)	GEV (0.187)	LN (1.5)	**GEV (5)**	GEV (0.063)	GEV (0.182)	GEV (0.295)	**GEV (3)**	W (0.077)	GEV (0.358)	G (0.216)	**GEV (5)**
	11	Aug	GEV (0.079)	W (0.322)	W (0.727)	**W (4)**	W (0.106)	GEV (0.411)	G (0.251)	**G (6)**	GEV (0.053)	GEV (0.135)	LN (0.699)	**GEV (5)**
	12	Sep	GEV (0.068)	GEV (0.303)	GEV (4.59)	**GEV (3)**	GEV (0.08)	GEV (0.357)	GEV (1.93)	**GEV (3)**	GEV (0.065)	GEV (0.163)	GEV (1.69)	**GEV (3)**
Rangpur	1	Oct	GEV (0.072)	LN (0.274)	W (1.59)	**GEV (5)**	LN (0.095)	GEV (0.289)	GEV (1.96)	**GEV & LN (5)**	GEV (0.052)	GEV (0.158)	GEV (1.35)	**GEV (3)**
	2	Nov	W (0.078)	G (0.288)	W (1.34)	**W (6)**	W (0.097)	GEV (0.278)	GEV (1.09)	**GEV (5)**	GEV (0.052)	GEV (0.158)	GEV (0.48)	**GEV (3)**
	3	Dec	GEV (0.096)	GEV (0.527)	G (0.228)	**GEV & G (5)**	LN (0.103)	GEV (0.305)	GEV (1.36)	**LN (5)**	GEV (0.059)	GEV (0.166)	GEV (0.286)	**GEV (3)**
	4	Jan	W (0.072)	GEV (0.161)	LN (0.396)	**GEV (6)**	LN (0.097)	GEV (0.247)	GEV (1.36)	**GEV & LN (5)**	GEV (0.063)	GEV (0.181)	GEV (0.29)	**GEV (3)**
	5	Feb	GEV (0.077)	GEV (0.32)	LN (0.79)	**GEV (5)**	LN (0.096)	GEV (0.288)	GEV (1.34)	**GEV (4)**	GEV (0.056)	GEV (0.147)	GEV (0.288)	**GEV (3)**
	6	Mar	GEV (0.05)	G (0.152)	GEV (1.42)	**GEV (4)**	GEV (0.069)	GEV (0.236)	GEV (1.25)	**GEV (3)**	GEV (0.054)	GEV (0.15)	GEV (0.494)	**GEV (3)**
	7	Apr	GEV (0.109)	GEV (0.589)	GEV (5.1)	**GEV (3)**	LN (0.081)	GEV (0.337)	W (2.0)	**GEV & LN (6)**	GEV (0.074)	GEV (0.204)	GEV (1.8)	**GEV (3)**
	8	May	GEV (0.062)	GEV (0.164)	LN (1.22)	**GEV (4)**	W (0.082)	GEV (0.386)	G (1.43)	**G (6)**	GEV (0.053)	GEV (0.128)	GEV (0.762)	**GEV (3)**
	9	Jun	LN (0.095)	LN (0.437)	W (3.39)	**LN (6)**	GEV (0.094)	GEV (0.474)	LN (3.69)	**GEV (4)**	GEV (0.064)	GEV (0.217)	GEV (0.733)	**GEV (3)**
	10	Jul	GEV (0.077)	GEV (0.621)	GEV (1.08)	**GEV (3)**	GEV (0.06)	GEV (0.159)	LN (0.712)	**GEV (4)**	GEV (0.048)	GEV (0.135)	LN (1.23)	**GEV & LN (5)**
	11	Aug	GEV (0.102)	GEV (0.57)	W (4.9)	**GEV & LN (6)**	GEV (0.105)	GEV (0.579)	LN (2.88)	**GEV (4)**	GEV (0.088)	GEV (0.346)	GEV (2.57)	**GEV (3)**
	12	Sep	LN (0.122)	GEV (0.618)	GEV (0.645)	**LN (5)**	GEV (0.105)	GEV (0.66)	LN (1.5)	**GEV (4)**	GEV (0.099)	GEV (0.304)	GEV (6.04)	**GEV (3)**

The best-fit results of K-S, A-D, and Chi-Sq. tests, as well as the best-scored results of proposed distributions at multiple time scales of precipitation data, are presented in Table 4. The table displays the test statistic results of the best-fit distribution type in probabilities for each model selection tool (K-S, A-D, Chi-Sq.). For each selection tool, all developed probability distributions are ranked, with rank 1 indicating the best fit. The ranking results of the three tools are combined to calculate a ranking score. The best-fit distribution model for each station and month is chosen by identifying the distribution model with the smallest ranking score. There was a total of 108 (3×36) cases. As can be seen from the table, the GEV distribution gives the best fit in most of the cases, followed by LN and Weibull. The weak fitting of the LN, Weibull, and Gamma distributions is usually expected as they have less flexibility in modeling the shape of the tail of the data than the GEV distribution, which has three parameters in it. The GEV distribution has a shape parameter that captures the frequency of extreme events, unlike other distributions. The lognormal distribution implies more extreme values, while the Weibull and gamma distributions indicate a lower likelihood for positive shape parameters. The GEV distribution can accommodate both high and low frequencies of extreme events, depending on the shape parameter, thus yielding the best fit among all the stations and at various time scales.

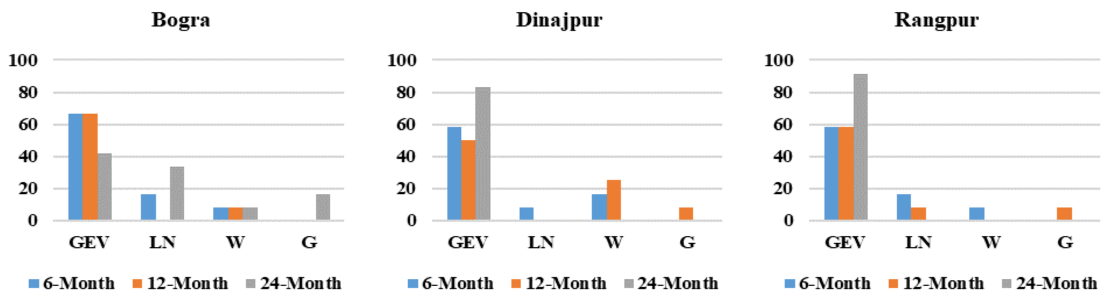


Figure 2: Percent best-fit of Generalized Extreme Value (GEV), Lognormal (LN), Weibull (W), and Gamma (G) distributions over 6,12-, and 24-month time scales at the selected stations

A comparative analysis of the percentage of best fit for the selected distributions at multiple time scales for each station is shown in Fig. 2. The GEV distribution appears to be the most significant across all stations and periods, particularly over 24 months. The LN, Weibull, and Gamma distributions also show increased significance over longer periods. These trends demonstrate that the GEV distribution is likely to be the most dominant for longer periods, followed by the LN, Weibull, and Gamma distributions.

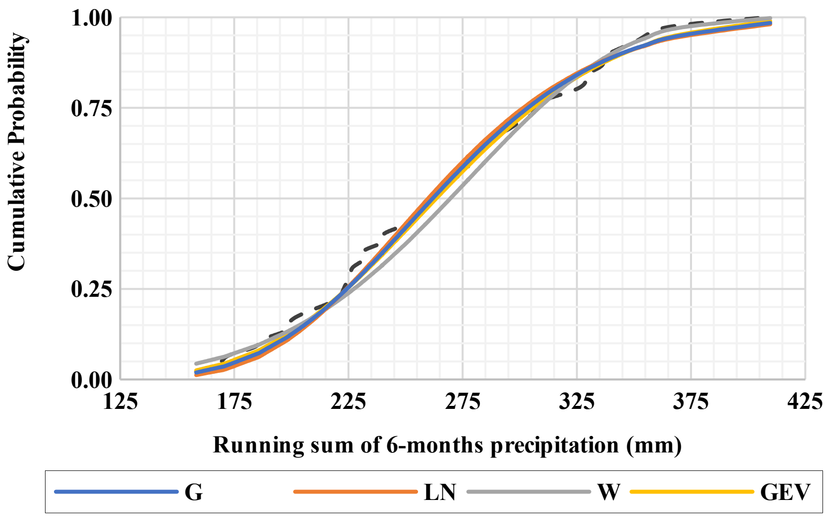
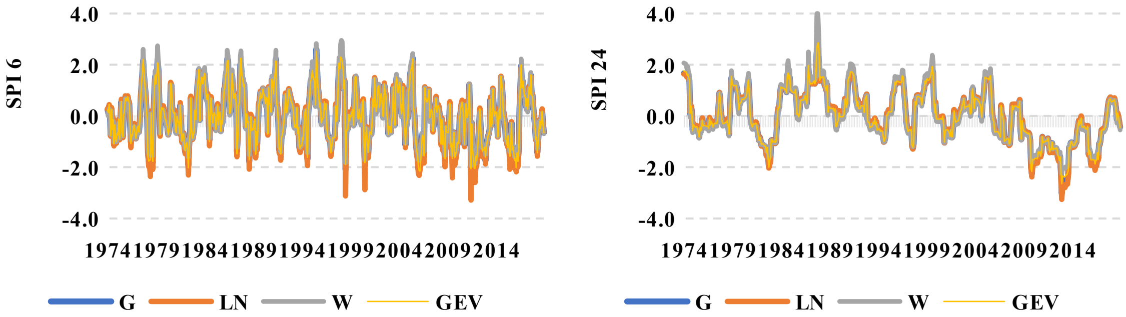
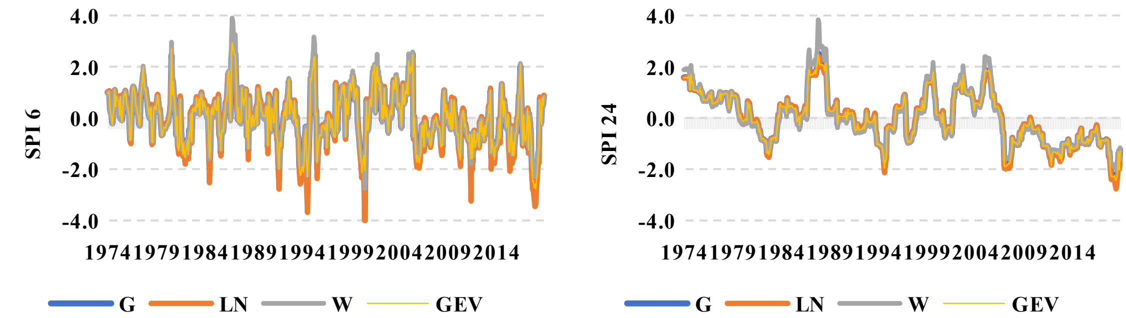


Figure 3: Comparison of empirical and theoretical probability distributions

Fig. 3 presents a comparison of the running sum of 6 months of precipitation data for the empirical and four theoretical cumulative probability distributions (Gamma, lognormal, Weibull, and generalized extreme value) for a single time series. All the theoretical distributions appear to provide a good fit for the empirical CDF. However, the generalized extreme value distribution is found to provide the best fit among them.



(a)



(b)

Figure 4: Drought index SPI for the time series (a) Bogra and (b) Dinajpur for the 1975-2019 period at 6-month and 24-month time scale

The gamma distribution is the most commonly used probability density function to model precipitation data for calculating the drought index SPI. However, it is worth examining if the SPI can be modelled equally well or better by the lognormal, Weibull, and generalized extreme value probability distributions, apart from the traditional gamma distribution across multiple time scales. In Fig. 4, SPI values are obtained using the above-mentioned distributions at 6- and 24-month time scales for Bogra and Dinajpur stations. It is observed at shorter time scales that lognormal and Weibull distributions both generate larger values of SPI quite often at very dry or very wet periods, while at larger scales it is hardly seen for both stations. It means that the Weibull and lognormal distributions are primarily liable for over-estimating and under-estimating the wet and dry periods at shorter scales in extreme cases. On the contrary, the generalized extreme value (GEV) distribution produces almost the same SPI values as the gamma distribution, providing a more stable representation of extreme precipitation events without overestimating the severity of drought or excessive wetness. The stability of SPI values based on the gamma and other distributions (LN, Weibull, and GEV in the current study) with increasing time scales demonstrates the fact that these distributions become more reliable in representing longer-term drought conditions without exaggerating extreme conditions.

4. CONCLUSIONS

The standardized precipitation index (SPI) is considered to be the most widely used index to assess drought conditions across many parts of the world due to its simplicity of use and lower data requirements. Only precipitation data is needed to obtain SPI values. In general, the gamma distribution is used to calculate the SPI values, and now it has been brought into question whether to use it in the calculation process as many other probability distribution functions have shown fitting the precipitation data well across various regions and multiple timescales. Therefore, the current study aims to explore the suitability of using alternative probability distribution functions such as LN, Weibull, and GEV to calculate SPI-based drought events. Three meteorological stations in the northwestern region of Bangladesh, namely Bogra, Dinajpur, and Rangpur, have been considered for this study, as these areas usually receive quite less precipitation than any other parts of the country over a longer period of time. Maximum likelihood estimators are used to obtain the parameters of the distributions. In order to observe the best fitting distribution for the precipitation data, the Kolmogorov-Smirnov (K-S) test, the Anderson-Darling (A-D) test, and the Chi-Square (Chi-Sq.) tests are performed. The tests are conducted on 6-, 12-, and 24-month time scales for each month at the three stations. Among 108 cases, 64% show the best fit for GEV distribution, while LN, Weibull, and Gamma distribution exhibit only 9%, 8%, and 4%, respectively. The other 15% signifies 16 cases, where 15 cases yield the best fit for the GEV distribution along with another distribution (LN, Weibull, or Gamma). It emphasizes that the GEV distribution has a clear dominance over other distributions, irrespective of any locations or time periods considered. LN, Weibull, and Gamma distributions are also seen to fit best to the precipitation data at higher time scales but yet seem to be insignificant when compared with the GEV distribution. Moreover, the lognormal and Weibull distributions are found to be incapable of capturing extremes at shorter time scales, as they yield comparatively larger SPI values than those of the gamma and generalized extreme value distributions. However, at higher time scales, both the distributions show improvement in capturing the extremes, which signifies the reduction of the overestimation and underestimation of the dry and wet periods. The findings of the study also show that the GEV distribution remains the best choice as it offers greater stability in presenting extreme precipitation events, which conclusively proves the viability of using alternative probability distribution functions.

REFERENCES

Alamgir, M., Shahid, S., Hazarika, M. K., Nashrrullah, S., Harun, S. Bin, & Shamsudin, S. (2015). Analysis of Meteorological Drought Pattern During Different Climatic and Cropping Seasons in Bangladesh. *JAWRA Journal of the American Water Resources Association*, *51*(3), 794–806. https://doi.org/10.1111/jawr.12276

Amin, M. T., Rizwan, M., & Alazba, A. A. (2016). A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan. *Open Life Sciences*, *11*(1), 432–440. https://doi.org/10.1515/biol-2016-0057

Angelidis, P., Maris, F., Kotsovinos, N., & Hrissanthou, V. (2012). Computation of Drought Index SPI with Alternative Distribution Functions. *Water Resources Management*, *26*(9), 2453–2473. https://doi.org/10.1007/s11269-012-0026-0

Ashkar, F., & Nwentsa Tatsambon, C. (2007). Revisiting some estimation methods for the generalized Pareto distribution. *Journal of Hydrology*, *346*(3–4), 136–143. https://doi.org/10.1016/j.jhydrol.2007.09.007

Bhakar, S., Iqbal, M., Devanda, M., Chhajed, N., & Bansal, A. (2008). Probability analysis of rainfall at Kota. *Indian Journal of Agricultural Research*, *42*.

Blain, G. C., & Meschiatti, M. C. (2015). Inadequacy of the gamma distribution to calculate the Standardized Precipitation Index. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *19*(12), 1129–1135. https://doi.org/10.1590/1807-1929/agriambi.v19n12p1129-1135

Cindric, K., Juras, J., & Pasaric, Z. (2012). Statistical distributions for the SPI computation. *EMS Annual Meeting AbstractsBerlin: EMS (Berlin). GF48: EMS2012-316*.

Dey, N., Alam, M., Sajjan, A., Bhuiyan, M., Ghose, L., Ibaraki, Y., & Karim, F. (2012). Assessing Environmental and Health Impact of Drought in the Northwest Bangladesh. *Journal of Environmental Science and Natural Resources*, *4*(2). https://doi.org/10.3329/jesnr.v4i2.10141

Edossa, D. C., Babel, M. S., & Gupta, A. Das. (2010). Drought analysis in the Awash River Basin, Ethiopia. *Water Resources Management*, *24*(7). https://doi.org/10.1007/s11269-009-9508-0

Guttman, N. B. (1999). ACCEPTING THE STANDARDIZED PRECIPITATION INDEX: A CALCULATION ALGORITHM [1]. *JAWRA Journal of the American Water Resources Association*, *35*(2), 311–322. https://doi.org/10.1111/j.1752-1688.1999.tb03592.x

Hao, Z., & AghaKouchak, A. (2013). Multivariate Standardized Drought Index: A parametric multi-index model. *Advances in Water Resources*, *57*, 12–18. https://doi.org/10.1016/j.advwatres.2013.03.009

Hong, X., Guo, S., & Zhou, Y. (2013). *Applicability of Standardized Precipitation Index with Alternative Distribution Functions*. https://api.semanticscholar.org/CorpusID:124861386

Hosain, N., & Amin, F. (2023). Four Decades of Cardiac Surgery in Bangladesh: A Noble Journey That Started with the Help of Japan. *JMA Journal*, *6*(1), 1–8. https://doi.org/10.31662/jmaj.2022-0148

Karson, M. J. (1968). Handbook of Methods of Applied Statistics. Volume I: Techniques of Computation Descriptive Methods, and Statistical Inference. Volume II: Planning of Surveys and Experiments. I. M. Chakravarti, R. G. Laha, and J. Roy, New York, John Wiley; 1967, $9.00. *Journal of the American Statistical Association*, *63*, 1047–1049. https://api.semanticscholar.org/CorpusID:122131958

Khudri, Md. M., & Sadia, F. (2013). Determination of the Best Fit Probability Distribution for Annual Extreme Precipitation in Bangladesh. *European Journal of Scientific Research*, *103*, 391–404.

Mahmoudi, P., Rigi, A., & Miri Kamak, M. (2019). A comparative study of precipitation-based drought indices with the aim of selecting the best index for drought monitoring in Iran. *Theoretical and Applied Climatology*, *137*(3–4), 3123–3138. https://doi.org/10.1007/s00704-019-02778-z

Mandal, S., & Choudhury, B. U. (2015). Estimation and prediction of maximum daily rainfall at Sagar Island using best fit probability models. *Theoretical and Applied Climatology*, *121*(1–2), 87–97. https://doi.org/10.1007/s00704-014-1212-1

M.Arshad, ., M.T.Rasool, ., & M.I.Ahmad, . (2003). Anderson Darling and Modified Anderson Darling Tests for Generalized Pareto Distribution. *Journal of Applied Sciences*, *3*, 85–88. https://api.semanticscholar.org/CorpusID:121942201

McKee, T. B., Nolan, J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. *Preprints, Eighth Conf. on Applied Climatology, Amer. Meteor, Soc.*, *January*.

Melesse, A., Abtew, W., Dessalegne, T., & Wang, X. (2010). Low and high flow analyses and wavelet application for characterization of the Blue Nile River system. *Hydrological Processes*, *24*(3), 241–252. https://doi.org/10.1002/hyp.7312

Mondol, Md. A. H., Das, S. C., & Islam, Md. N. (2016). Application of Standardized Precipitation Index to assess meteorological drought in Bangladesh. *Jàmbá: Journal of Disaster Risk Studies*, *8*(1). https://doi.org/10.4102/jamba.v8i1.280

Palmer. (1965). *USWB_Meteorological_Drought_1965*.

Park, J.-S., Seo, S.-C., & Kim, T. Y. (2009). A kappa distribution with a hydrological application. *Stochastic Environmental Research and Risk Assessment*, *23*(5), 579–586. https://doi.org/10.1007/s00477-008-0243-5

Rabby, M.F., Adhikary, S.K., "Uncertainty Based Assessment of Drought Using Standardized Precipitation Index – A Case Study", *6th International Conference on Advances in Civil Engineering (ICACE 2022)*, CUET, Chittagong, Bangladesh, pp.1145-1152, 21-23 December, 2022.

Ridgman, W. J. (1990). Statistical Methods , 8th edn, by G. W. Snedecor & W. G. Cochran. xx + 503 pp. Ames: Iowa State University Press (1989). $44.95 (hard covers). ISBN 0 8138 1561 6. *The Journal of Agricultural Science*, *115*, 153–153. https://api.semanticscholar.org/CorpusID:86714360

Seier, E. (2002). Comparison of tests of univariate normality. *Interstat*, *1*.

Shukla, S., & Wood, A. W. (2008). Use of a standardized runoff index for characterizing hydrologic drought. *Geophysical Research Letters*, *35*(2), L02405. https://doi.org/10.1029/2007GL032487

Sienz, F., Bothe, O., & Fraedrich, K. (2012). Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias. *Hydrology and Earth System Sciences*, *16*(7), 2143–2157. https://doi.org/10.5194/hess-16-2143-2012

Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, *69*(347), 730. https://doi.org/10.2307/2286009

Tigkas, D., Vangelis, H., & Tsakiris, G. (2019). Drought characterisation based on an agriculture-oriented standardised precipitation index. *Theoretical and Applied Climatology*, *135*(3–4), 1435–1447. https://doi.org/10.1007/s00704-018-2451-3

Vergni, L., Di Lena, B., Todisco, F., & Mannocchi, F. (2017). Uncertainty in drought monitoring by the Standardized Precipitation Index: the case study of the Abruzzo region (central Italy). *Theoretical and Applied Climatology*, *128*(1–2), 13–26. https://doi.org/10.1007/s00704-015-1685-6

Zhou, J., Chen, X., Xu, C., & Wu, P. (2022). Assessing Socioeconomic Drought Based on a Standardized Supply and Demand Water Index. *Water Resources Management*, *36*(6). https://doi.org/10.1007/s11269-022-03117-0