



Suitability of distributions for standard precipitation and evapotranspiration index over meteorologically homogeneous zones of India

N T MONISH and S REHANA*

Lab for Spatial Informatics, International Institute of Information Technology, Hyderabad 500 032, India.

*Corresponding author. e-mail: rehana.s@iiit.ac.in

MS received 5 February 2019; revised 16 June 2019; accepted 25 July 2019; published online 20 December 2019

The Standardised Precipitation and Evapotranspiration Index (SPEI) became one of the popular drought indices in the context of increasing temperatures under global warming in recent periods. The SPEI is estimated by fitting a probability distribution for the difference between precipitation (P) and potential evapotranspiration (PET), which represents the climatic water balance. The choice of an inappropriate probability distribution may lead to bias in the index values leading to distorted drought severity. Till date, none of the studies have focused on the suitability of the probability distribution for SPEI over India. The objective of the present study is to compare and evaluate the performance of a group of candidate probability distributions over seven meteorologically homogeneous zones and all over India using high resolution (0.25°) gridded daily precipitation data from India Meteorological Department (IMD). The Kolmogorov–Smirnov (K–S) test was used to test the goodness-of-fit for (P–PET) and Akaike Information Criterion (AIC) was used to obtain the relative distribution rankings for each grid point. The results of the study suggest that Pearson type III distribution has performed better than other distributions, significantly for shorter time scales and slightly for longer time scales, for each meteorological homogeneous zone based on K–S test. Also, for shorter time scales, Pearson type III distribution has been observed to be significantly better based on AIC with 82.89% and 71.91% grid points for 3 and 6 months, respectively. However, the relative ranking by AIC revealed GEV distribution as the best fit for SPEI values all over India for longer time scales with total grid points as 50.26%, and 58.81% for 12- and 24-month time scales respectively. Pearson type III distribution for shorter time scales (3 and 6 months) and GEV distribution for longer time scales (12 and 24 months) have been identified as the best distributions for fitting SPEI for Indian case study. Comparison of GEV based SPEI with remote sensing-based drought severity index (DSI) for drought events indicated concordance for most of regions in India. Also, SPEI is evaluated to test its capability to represent seasonality and its performance has been compared with Standardised Precipitation Anomaly Index (SPAI) which is known to represent seasonality well.

Keywords. SPEI; India; GEV; log-logistic; K–S test; AIC.

1. Introduction

Drought is an extreme event which continues for a long duration affecting natural resources, environment, and society. The onset, persistence,

and termination of a drought are usually gradual processes; however, its overall impact may be far more disastrous than that of flood (Bhalme and Mooley 1980). Several drought indices have been developed, which evaluate the deviation of

climate variables in a given year from the normal conditions (Dai 2011; Liu *et al.* 2017). These drought indices also serve as monitoring tools and operational indicators for water resources management. The Standardised Precipitation Index (SPI) developed by McKee *et al.* (1993) measures the normalised rainfall anomalies. This was considered as simple and most widely used universal drought index and also recommended by the World Meteorological Organization (WMO 2012). Several studies adopted SPI for the drought characterisation globally (Bordi *et al.* 2004; Wu *et al.* 2007) and for Indian subcontinent (Guhathakurta 2003; Goswami *et al.* 2006; Ghosh and Mujumdar 2007; Pai *et al.* 2011). However, SPI can only estimate the drought under lack of precipitation, but will not be able to detect the drought conditions under higher than normal atmospheric evaporative demand (Vicente-Serrano *et al.* 2015). More specifically, a drought indicator should account for the atmospheric demand of moisture in the context of increasing temperatures in the recent periods of global warming. SPEI has been proposed by Vicente-Serrano *et al.* (2010), which considers the PET in addition to precipitation and it can be worked at several time scales. In the recent years, SPEI has been widely used to study the drought events worldwide (Allen *et al.* 2011; Aadhar and Mishra 2017). In the context of Indian subcontinent, several studies (Kumar *et al.* 2013; Mallya *et al.* 2016; Alam *et al.* 2017; Nath *et al.* 2017) used SPEI to analyse the Indian drought scenario.

The formulation of SPEI necessitates the choice and fitting of an appropriate parametric probability distribution for transformation of accumulated climatic water demand into standard normal distribution. The choice of an inappropriate probability distribution may lead to bias in the index values leading to inaccurate drought characterisation (Sienz *et al.* 2012; Stagge *et al.* 2015). Moreover, most of the earlier drought studies all over India were based on the original structure of SPEI formulation by Vicente-Serrano *et al.* (2010). Vicente-Serrano *et al.* (2010, 2015) formulated SPEI using three parameter log-logistic (LL) distribution. However, Stagge *et al.* (2015) recommended Generalised Extreme Value (GEV) distribution for formulating the SPEI. Few more studies (Homdee *et al.* 2016) also revealed that GEV distribution fits well for the climatic water balance. Pearson Type III (PT-III) distribution

was also noted to be a reliable distribution for the estimation of SPEI over China (Wang *et al.* 2019). Given the ambiguity and debate on the selection of appropriate probability distribution for SPEI drought index formulation (Stagge *et al.* 2016; Vicente-Serrano and Beguería 2016), it is important to study the performance of various probability distributions and their corresponding suitability over a particular region in the context of water resources management decision making. To the best of author's knowledge, none of the drought studies over India have been conducted to analyse the performance of a probability distribution in the drought characterisation using SPEI. So we considered GEV, 3-parameter log-logistic, Pearson Type III and Normal distributions for the performance evaluation based on the literature (Vicente-Serrano *et al.* 2010; Stagge *et al.* 2015; Homdee *et al.* 2016; Wang *et al.* 2019). Further, application of SPEI with a single probability distribution all over India may not be a meaningful index due to the large spatial heterogeneity in the rainfall patterns at various temporal scales of Indian monsoon rainfall (Rajeevan *et al.* 2010; Saikranthi *et al.* 2013).

The objective of the present analysis is to compare and evaluate the performance of a group of candidate probability distributions over seven meteorological homogeneous zones covering Indian land mass at various time scales. The probability distributions were tested using Kolomogrov–Smirnov test and relative distribution rankings by Akaike Information Criterion (AIC) for all zones with fine resolution gridded datasets at 0.25° provided by India Meteorological Department (IMD). Further, comparison is made between the SPEI drought index obtained using the best fitted distribution and remote sensing-based drought severity index (DSI) to evaluate the performance of SPEI in the event of drought. Given that India is a monsoon dominated country, the study tested for the seasonality of SPEI in capturing the variation of rainfall in dry and wet months.

The paper is structured as follows: The study area, data, and methodology involved in fitting the climatic water balance are explained in section 3. The results of each distribution for each meteorologically homogeneous zone and all over India are compared in section 4 with the concluding remarks in the section 5. All the figures in this paper are generated using MATLAB and QGIS.

2. Materials and methods

2.1 Study area and data

The Indian subcontinent extent which is defined as the region between 7.5°–37.5°N and 67.5°–97.5°E, covering main land region of India, was considered in the present study (figure 1). India is characterized by seven major meteorological homogeneous zones based on India Meteorological Department (IMD) (Parthasarathy *et al.* 1996), described herein as the North, South, West, Central, North-East, North-East hills, and Jammu and Kashmir (J&K). Out of these seven zones, complete data is available for five zones, while portions of J&K and North-East hills were not considered due to inconsistent data availability.

The present study made an effort to use a high resolution 0.25° daily rainfall data prepared by IMD for the period of 1901–2014 for a spatial domain of 7.5°–37.5°N and 67.5°–97.5°E. The daily rainfall records are generated from 6955 rain gauge stations with varying availability. The density of the stations is relatively high in the southern Peninsula and relatively low over northern-most areas of India, northwest India, northeast India, and eastern parts of central India (Pai *et al.* 2014). The new fine resolution datasets were tested by Pai *et al.* (2014) for the comparability of large scale climatological features of rainfall all over India with the existing IMD rainfall datasets at 0.5° and 1° resolutions (Rajeevan *et al.* 2008; Rajeevan and Bhate 2009). The daily maximum and minimum

temperature datasets at 1° resolution for the period of 1951–2014 from IMD are used in the present study. The data of 395 quality controlled stations are used with an angular distance weighting algorithm (Shepard 1968) and compared with monthly mean temperature dataset by Cort Willmott and Kenji Matsuura of University of Delaware (www.cdc.noaa.gov) with 0.5° resolution for the period of 1951–1999 with 0.8 correlation coefficient for most parts of the country (Srivastava *et al.* 2009). The average temperature datasets are brought to a common resolution of 0.25° at precipitation gridded dataset by bilinear interpolation method with the adequacy of interpolation of smoothly varying variables by National Centre for Atmospheric Research (NCAR). The precipitation and temperature datasets were aggregated to monthly scale to perform the analysis.

DSI is based on the current operational satellite-based terrestrial evapo-transpiration (ET/PET) obtained from a Moderate Resolution Imaging Spectroradiometer (MODIS) MOD16 ET product driven by the National Centers for Environmental Prediction–Department of Energy Atmospheric Model Intercomparison Project Reanalysis II (NCEP–DOE II). It also takes normalized difference vegetation index (NDVI) products based on MODI-13 as inputs for the calculation of drought index values. DSI data sets of 0.5° resolution for a period of 12 yrs (2000–2011) were obtained from Numerical Terradynamic Simulation Group (NTSG) website. The data is then interpolated using bilinear interpolation to 0.25° resolution for comparison with SPEI.

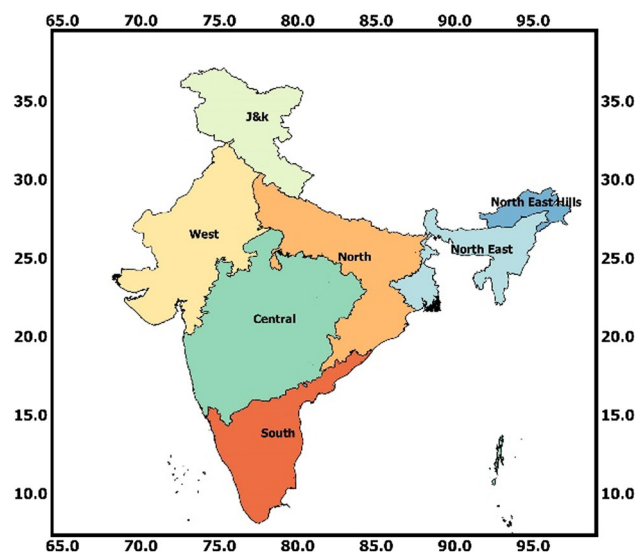


Figure 1. IMD meteorologically homogeneous regions of India.

2.2 Methodology

2.2.1 Calculation of PET by Hargreaves model

The drought index SPEI is estimated by fitting a probability distribution for the climatic water balance (D) which is calculated as follows:

$$D = P - PET \quad (1)$$

where P is the monthly precipitation (mm) and PET is the monthly potential evapotranspiration (mm). There are several methods to estimate the PET (Thornthwaite 1948; Penman 1948; Ivanov 1954; Doorenbos and Pruitt 1977) and applicability of each method depends on the availability of the meteorological data for the given region. Most of the drought assessment studies (Thornthwaite 1948; Hargreaves and Samani 1982, 1985) have

dealt with only temperature (maximum, minimum and mean air temperatures) based PET methods. Vicente-Serrano *et al.* (2010) used Thornthwaite equation to estimate the PET, which considers the monthly mean air temperature and geographical location of the region of interest as input variables. The use of Thornthwaite equation leads to an underestimation of PET in arid and semiarid regions (Jensen *et al.* 1990), and an overestimation of PET in humid and tropical regions (Van der Schrier *et al.* 2011). Stagge *et al.* (2014) suggest that, if data permits, Hargreaves equation could be a useful balance between consistency and minimal data requirements. Therefore, the present study used Hargreaves model to estimate PET, which considers the monthly minimum, maximum and mean air temperatures, and geographical location of the region of interest as input variables. Hargreaves and Samani (1982, 1985) proposed several improvements to the Hargreaves (1975) equation for estimating daily PET (mm d^{-1}), as given:

$$PET = 0.0023 * (T_{\max} - T_{\min})^{1/2} * (T_{\text{mean}} + 17.8) * R_a \quad (2)$$

where T_{\max} , T_{\min} and T_{mean} represent the monthly maximum, minimum and mean temperature, respectively. R_a is the extra-terrestrial radiation expressed in equivalent evaporation units which is computed from information on location of the site based on the latitude and time of the year. Therefore, the only data required for the calculation of PET is the maximum, minimum and mean temperatures which were obtained from IMD.

Once, the PET is estimated, the values of climatic water balance (D) were estimated for each month using (equation 1). The PET values estimated based on Hargreaves model and monthly rainfall are aggregated to the respective n -month scale as follows:

$$P_i^k = \sum_{i-k+1}^i P_i \quad (3)$$

where $k=n$, $i=\{1, \dots, \text{length of time series}\}$

$$PET_i^k = \sum_{i-k+1}^i PET_i \quad (4)$$

where $k=n$, $i=\{1, \dots, \text{length of time series}\}$ P_i^k and PET_i^k are the accumulated rainfall and PET in month i respectively. Then, the n -month scale

climatic water balance (D) of SPEI in month i was estimated as follows:

$$D_i^n = P_i^n - PET_i^n \quad (5)$$

The present study used 3-, 6-, 12- and 24-month time scales in the estimation of D_i^n values. Based on the original formulation of SPEI proposed by Vicente-Serrano *et al.* (2010), a three-parameter log-logistic distribution is used for fitting the D_i^n values. The present study used the GEV, PT-III and normal distributions in addition to LL distribution for the transformation of D_i^n values in the standard normal values.

2.2.2 Fitting different probability distributions for accumulated climatic water balance

The present study considered the most commonly used probability distributions in the literature (Vicente-Serrano *et al.* 2010, 2015; Stagge *et al.* 2015; Wang *et al.* 2019). The climatic water balance (D) may take negative values when the PET exceeds rainfall, and hence not bounded by zero (Stagge *et al.* 2015). Therefore, the present study considered (i) GEV distribution, (ii) Three-parameter LL distribution, (iii) PT-III distribution, and (iv) Normal distributions to fit the time series of accumulated climatic water balance values at various time scales. The parameter estimation of GEV, PT-III and Normal distributions were fitted by Maximum likelihood estimation and the LL distribution with L-moments methods as suggested by Vicente-Serrano *et al.* (2010).

Generalized extreme value (GEV) distribution: GEV distribution comes under the family of extreme value theory, which is the limiting distribution for an observed variable of maximum or minimum values that are independent and identically distributed. The probability density function ($f(x)$) of the three-parameter GEV distribution is given as follows:

$$f(x) = \begin{cases} \left(\frac{1}{\sigma}\right) \left[(1 + \xi z(x))^{-1/\xi} \right]^{\xi+1} e^{-[(1+\xi z(x))^{-1/\xi}], \xi \neq 0, 1 + \xi z(x) > 0} \\ \left(\frac{1}{\sigma}\right) e^{-z(x) - e^{-z(x)}}, & \xi = 0, -\infty < x < \infty \end{cases} \quad (6)$$

where

$$z(x) = \frac{x - \mu}{\sigma} \quad (7)$$

where μ , σ and ξ are the location, scale and shape parameters respectively which are estimated using the maximum likelihood method. The cumulative distribution function (CDF), $F(x)$, for GEV can be calculated as follows:

$$F(x) = e^{-(t(x))} \tag{8}$$

where

$$t(x) = \begin{cases} \left(1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right)^{\frac{-1}{\xi}} & \text{if } \xi \neq 0 \\ e^{-(x - \mu)/\sigma} & \text{if } \xi = 0 \end{cases} \tag{9}$$

Three-parameter log-logistic distribution: The probability density function ($f(x)$) of the three-parameter log-logistic distribution is given as follows:

$$f(x) = \frac{\beta}{\alpha} \left(\frac{X - \gamma}{\alpha}\right)^{\beta-1} \left[1 + \left(\frac{X - \gamma}{\alpha}\right)^\beta\right]^{-2} \tag{10}$$

where α , β and γ are the scale, shape and origin parameters respectively which are obtained by L-moments procedure as follows:

$$\beta = \frac{2w_1 - w_0}{6w_1 - w_0 - 6w_2} \tag{11}$$

$$\alpha = \frac{(w_0 - 2w_1)\beta}{\Gamma(1 + 1/\beta)\Gamma(1 - 1/\beta)} \tag{12}$$

$$\gamma = w_0 - \alpha\Gamma(1 + 1/\beta)\Gamma(1 - 1/\beta) \tag{13}$$

where w_0 , w_1 , and w_2 are the probability weighted moments calculated based on method by Sheng and Hashino (2007), as follows:

$$W_r = \frac{1}{n} \binom{n-1}{r}^{-1} \sum_{j=1}^{n-r} \binom{n-j}{r} x_j, \quad r = 0, 1, 2 \tag{14}$$

where n is the sample size and x_j is the ordered vector of observations in descending order.

Next, the cumulative distribution function of log-logistic distribution can be calculated with the estimated parameters as

$$F(x) = \left[1 + \left(\frac{X - \gamma}{\alpha}\right)^{-\beta}\right]^{-1} \tag{15}$$

Normal distribution: The probability density function ($f(x)$) of the two-parameter Normal distribution is given as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty \tag{16}$$

where μ and σ are the mean and variance parameters respectively which are estimated using the maximum likelihood method. The cumulative distribution function for Normal distribution is calculated as

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right] dx \tag{17}$$

Pearson Type III distribution: The Pearson Type III distribution has been frequently used in hydrology studies. It is also known as generalized gamma distribution as it can be obtained by introducing a location parameter to a two-parameter gamma distribution. The probability distribution function of the Pearson Type III distribution is given as

$$f(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x - c}{\beta}\right)^{\alpha-1} e^{-(x-c)/\beta} \tag{18}$$

where α , β and c represent the shape, scale and location parameters respectively which are estimated using maximum likelihood method. The cumulative distribution function of the Pearson Type III distribution is not available in closed form. The Pearson Type III distribution parameters and cumulative distribution are calculated by ‘PearsonDS’ package in *R*.

2.2.3 Calculating the values of SPEI

With the values of $F(x)$, the SPEI values can be calculated as follows:

$$SPEI = W - \frac{C_0 + C_1 W + C_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \tag{19}$$

where

$$W = \sqrt{-2\ln(p)} \quad \text{for } p \leq 0.5 \tag{20}$$

where P is the probability of exceeding a determined D value, $P = 1 - F(x)$. If $P > 0.5$, then P is replaced by $1 - P$ and the sign of the resultant SPEI is reversed. The constants are $C_0 = 2.5515517$, $C_1 = 0.802583$, $C_2 = 0.010328$, $d_1 = 1.432788$, $d_2 = 0.189269$, and $d_3 = 0.001308$. By substituting the C_0 , C_1 and C_2 values in

equation (19), we calculate the SPEI values at n -month scale.

2.2.4 Performance evaluation criteria of candidate distributions

The best probability distribution for fitting the climatic water balance values in the formulation of SPEI value can be selected based on (i) distance measure between candidate distributions and the empirical distributions of the data using goodness-of-fit tests (K–S test) (ii) likelihood ratio tests based on the information criteria for relative ranking of the distributions (AIC test). In the present study two performance evaluation criterion, K–S goodness-of-fit and statistical method, AIC (Akaike 1974) is used to select the most appropriate candidate distribution in the formulation of SPEI for meteorologically homogeneous zones and overall India at different time scales.

Kolmogorov–Smirnov (K–S) test: The goodness-of-fit test, K–S test was adopted to decide if the estimated climatic water balance time series belongs to a hypothesized continuous distribution. The K–S test measures the maximum difference between the theoretical cumulative distribution and empirical cumulative distribution of sampled points. If x_1, \dots, x_n are the random samples from a candidate distribution with CDF ($F(x)$), the empirical CDF can be estimated as

$$F_n(x) = \frac{\text{number of observations} \leq x}{n} \quad (21)$$

The K–S test statistic will be based on the largest vertical difference (Δ) between the theoretical and empirical CDFs of candidate distribution.

$$\Delta = \max|F_n(x) - F(x)|. \quad (22)$$

The hypothesis that the data follows the candidate distribution will be rejected if the test statistic, Δ is greater than the critical value obtained from the standard table for a given significance level (Chakravarty *et al.* 1967).

Akaike Information Criterion (AIC): AIC provides the best distribution in a particular set of distributions by relative comparisons with one another. Majorly, AIC captures the bias of fit and

the unreliability from the number of model parameters. Each probability distribution will be ranked according to the AIC value and the distribution function with the minimum AIC value can be considered as the best distribution for the estimation of SPEI for each grid point.

$$AIC = 2k - 2 \ln(L), \quad (23)$$

where k is the number of parameters of a probability distribution and L is the maximized value of the likelihood function for the distribution.

3. Results and discussion

The selected four distributions (LL, GEV, PT-III, and Normal) were tested and compared with K–S test for goodness-of-fit and relative comparison among the selected distributions using AIC information. A rejection frequency for each candidate distribution was defined based on K–S test to study the suitability of a particular distribution for fitting the climatic water balance for overall India and for each meteorologically homogeneous zone. The rejection frequency is defined as the ratio of number of grid points which did not fit the time series of water balance, D , (P–PET) for the selected distribution, to the total number of grid points in the region at a given significance level. Lower the rejection frequency, better the distribution performance in a particular region according to K–S test at a significance level of 0.05. Further to understand the appropriateness of each of the probability distributions for various homogeneous meteorological zones, the present study analysed the agreement and disagreement of a given distribution for spatial averages (figure 2) and for various time scales for each grid point (figures 3–9).

The Normal distribution was not found to be a promising distribution for fitting SPEI for all zones and all over India for various time scales, therefore this distribution was excluded for zonal wise analysis comparison. Here onwards, the suitability of candidate distribution analysis at zonal level was made among GEV, log-logistic and PT-III distributions only. For shorter accumulation time periods (3 and 6 months), PT-III was identified as the best distribution compared to the other candidate distributions. Whereas, at longer accumulated time durations of 12 and 24 months all distributions except Normal were found to be

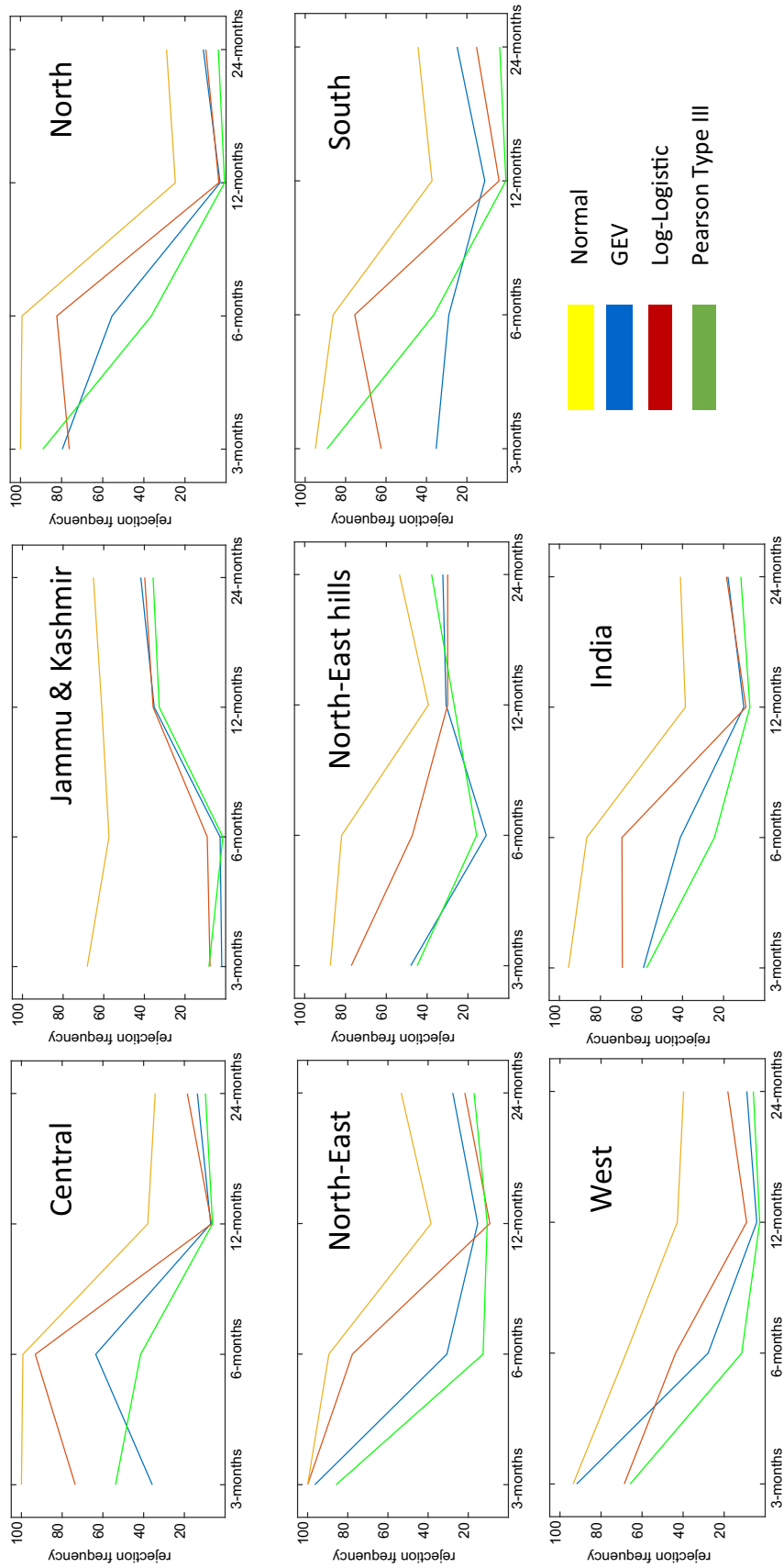


Figure 2. Spatial averaged K-S test rejection frequency (%) for SPEI probability distributions for various meteorologically homogeneous zones and all over India.

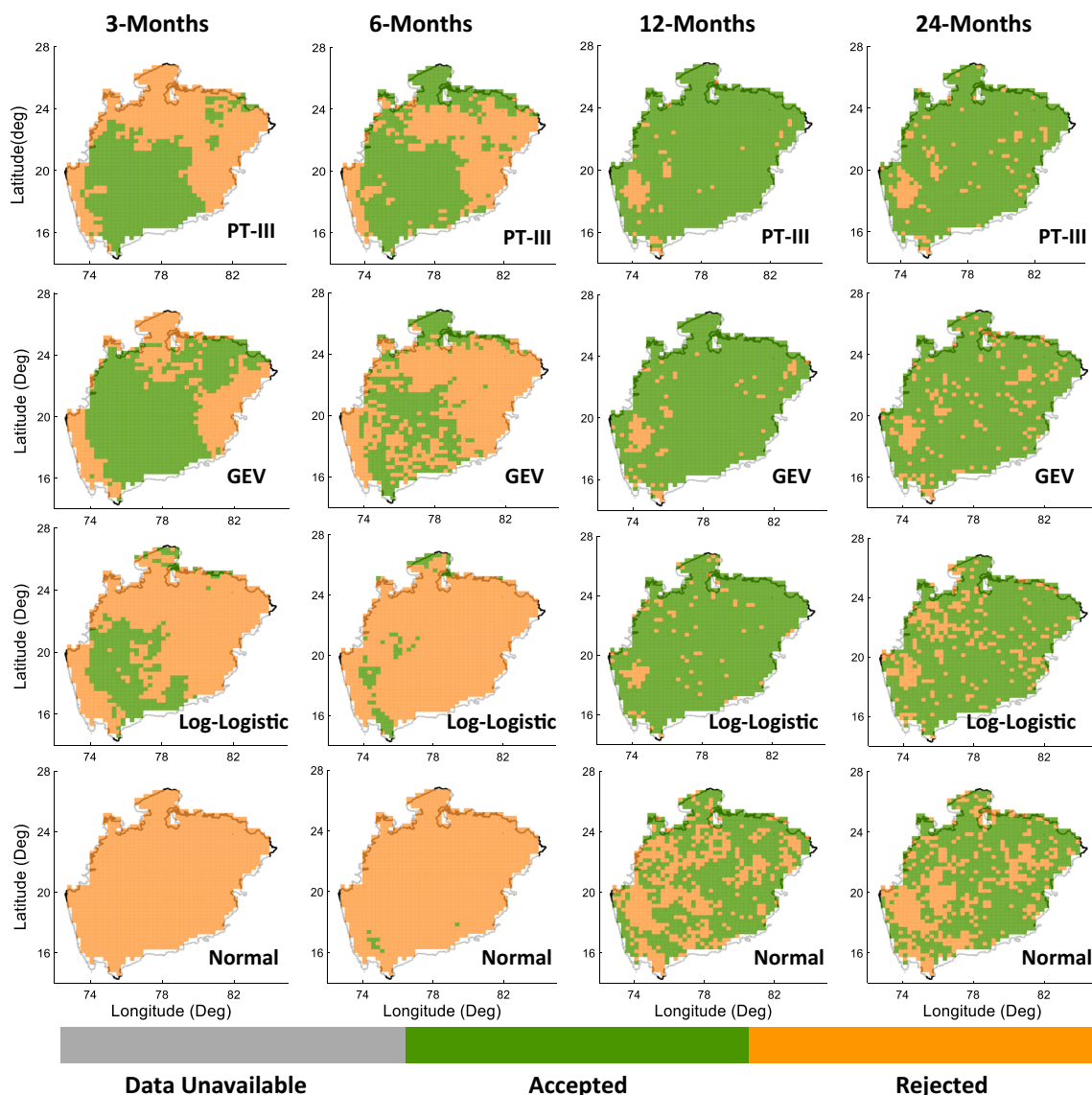


Figure 3. Results from K–S test for Central zone with accepted and rejected grid points at a significance level of 0.05 for each probability distribution for the accumulated periods of 3, 6, 12 and 24 months.

working with almost with same rejection frequencies with K–S test at 0.05 significance level.

For central zone, the GEV distribution was found to be more suitable at 3-month time accumulation period with spatial averaged K–S rejection frequencies estimated as 35.76% compared to other distributions. Whereas, for 6-, 12- and 24-month accumulation periods the PT-III was found to be the best fitting distribution. It can be also noted that, at 12-month time accumulation period, all distributions have performed with less rejection frequencies (less than 1% difference). Figure 3 shows the accepted and rejected grids with K–S test for various accumulated periods and with each distribution for Central Zone. Moreover, the difference between the disagreements with a

given distribution is more for the short accumulation periods of 3-months (35.76 for GEV, 73.59 for LL and 53.8 for PT-III) and 6-months (63.54 for GEV, 93.22 for LL and 41.5 for PT-III) compared to long accumulation periods of 12-months (7.07 for GEV, 6.62 for LL and 6.2 for PT-III) and 24 months (13.62 for GEV, 18.56 for LL and 9.7 PT-III). Therefore, for Central zone the GEV, LL and PT-III distributions have performed reasonably well at 12- and 24-month scales.

Figure 4 shows the accepted and rejected grids with K–S test for various accumulated periods and with each distribution for North Zone. The North zone’s K–S rejection frequencies were obtained as 79.68%, 76.27%, and 89.17% for GEV, LL, PT-III distributions respectively for 3-month scale

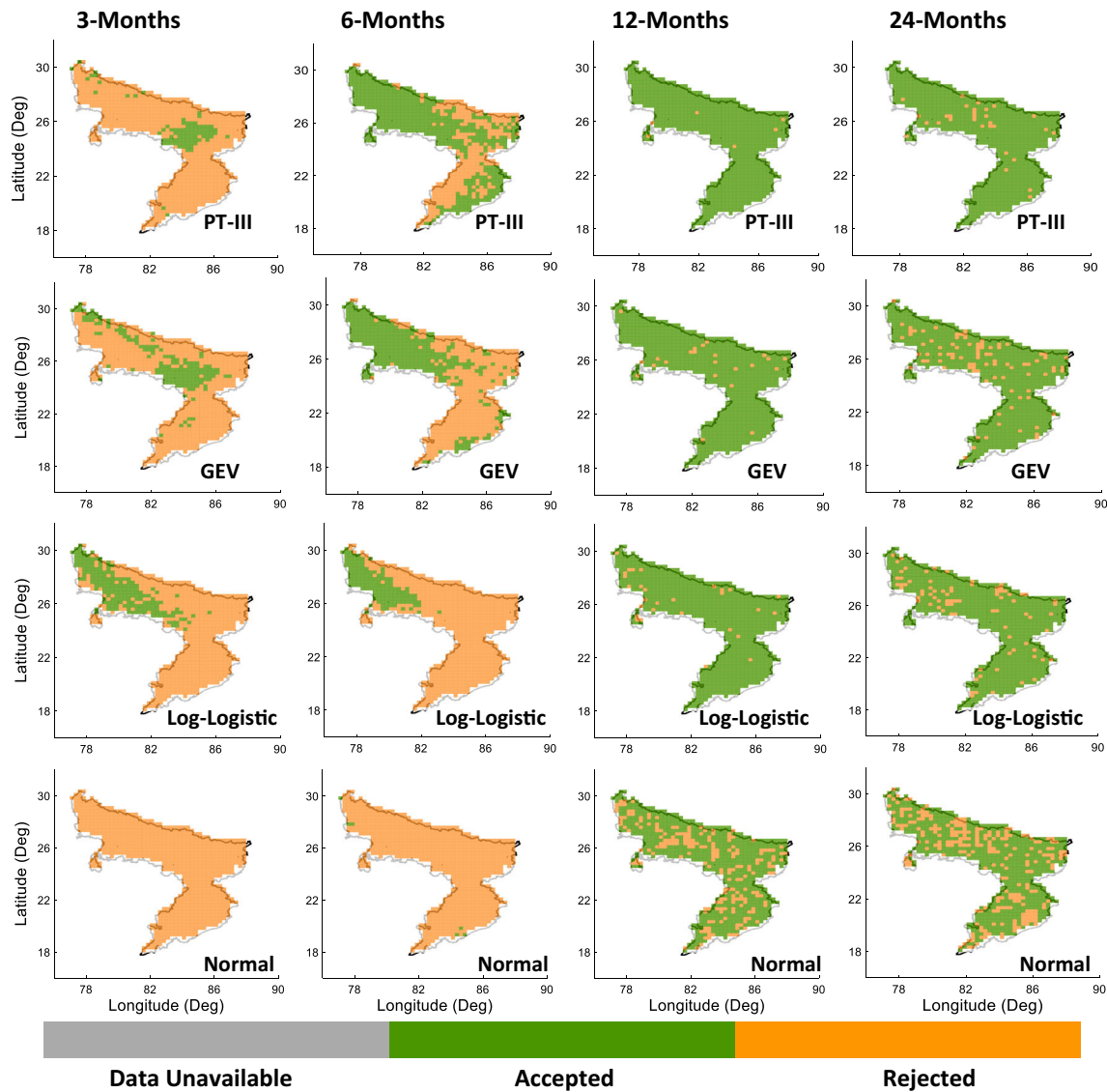


Figure 4. Results from K–S test for North zone with accepted and rejected grid points at a significance level of 0.05 for each probability distribution for the accumulated periods of 3, 6, 12 and 24 months.

(figures 2 and 4). Whereas, at 6-month scale the estimated K–S rejection frequencies were obtained as 55.47%, 82.36% and 36.37% for GEV, LL and PT-III distributions respectively (figures 2 and 4). Therefore, with the high K–S rejection values, none of the distributions under consideration were found to be promising at 3-month and 6-month accumulated periods for the fitting of SPEI values for North zone. However, for higher accumulated time periods of 12-month (2.91 for GEV, 3.52 for LL and 0.72 for PT-III) and 24-months (11.07 for GEV, 9.85 for LL and 3.77 for PT-III) the PT-III was noticed to perform well for fitting the SPEI. Particularly, for annual scale drought studies of SPEI, the PT-III distribution was found to be the best for North zone.

Figure 5 shows the accepted and rejected grids with K–S test for various accumulated periods and with each distribution for South Zone. For South zone, the PT-III distribution has performed as promising distribution to formulate the SPEI drought index for all accumulation periods. It can also be noted that LL distribution was found to be suitable with K–S rejection frequencies as 4.08 % and 15.24 % compared to PT-III rejection frequencies as 5.03% and 12.89% for 12- and 24-month accumulation periods respectively (figures 2 and 5). Figure 6 shows the accepted and rejected grids with K–S test for various accumulated periods and with each distribution for West Zone. Similar to North zone, for the West zone also none of the distributions under consideration fit the

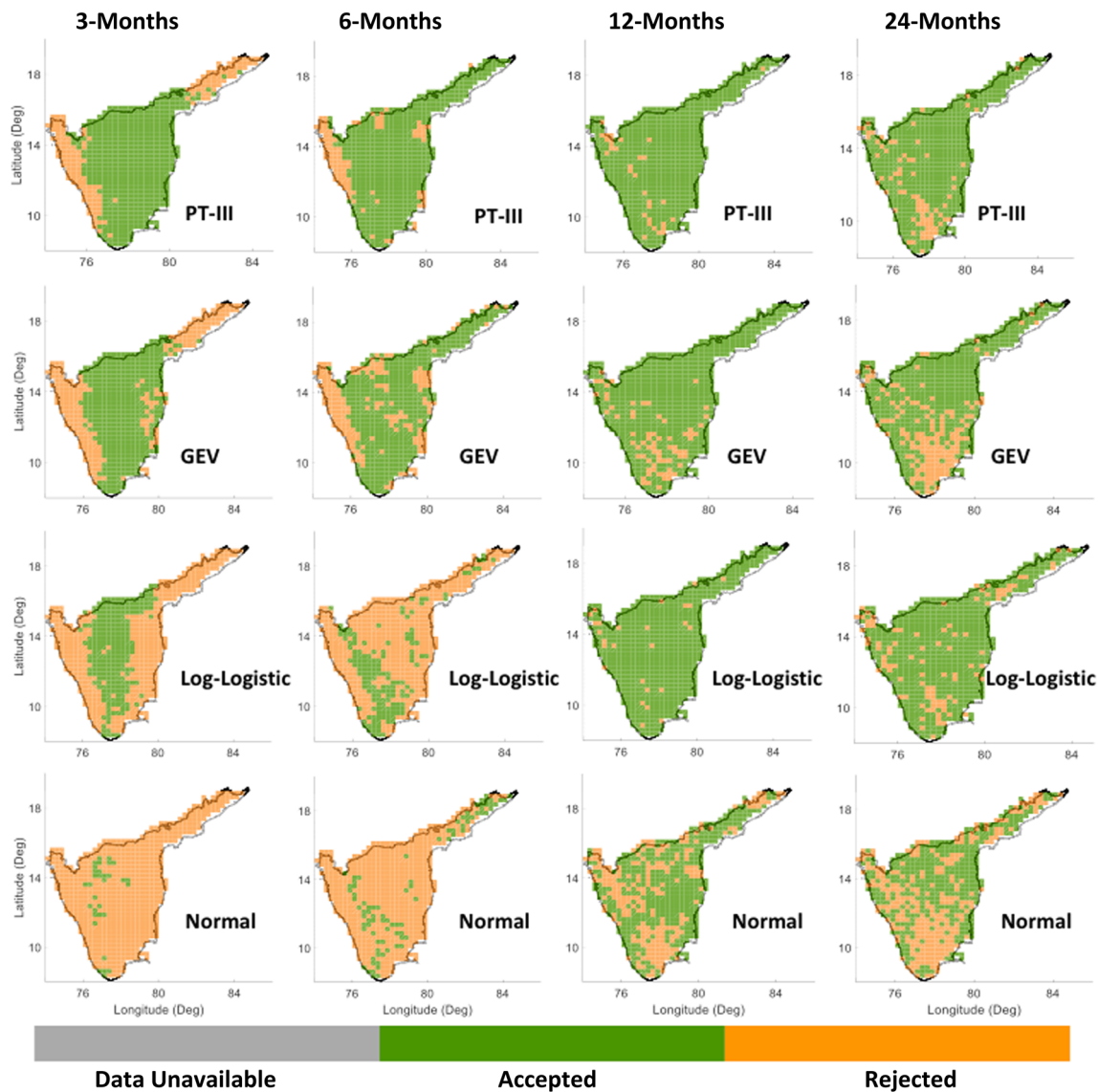


Figure 5. Results from K–S test for South zone with accepted and rejected grid points at a significance level of 0.05 for each probability distribution for the accumulated periods of 3, 6, 12 and 24 months.

3-month accumulation period of D with spatial averaged K–S rejection frequencies estimated as 91.86% for GEV, 68.69% for LL and 65.8 for PT-III. For the remaining accumulation time periods PT-III distribution was found to be suitable distribution for the formulation of SPEI for Western zone (figures 2 and 6). Figure 7 shows the accepted and rejected grids with K–S test for various accumulated periods and with each distribution for North-East Zone. For North-East zone also none of the distribution has shown as best fit for 3-month accumulation time period for the formulation of SPEI with rejection frequencies estimated as 96.39% for GEV, 100% and for 85.81 for PT-III. For longer time accumulation periods such as 12 and 24 months, the PT-III was identified as the

best distribution in the formulation of SPEI for North-East zone (figures 2 and 7).

Figure 8 shows the accepted and rejected grids with K–S test for various accumulated periods and with each distribution for North-East hills Zone. For the North-East hills (figures 2 and 8) zone, GEV distribution has been identified as the best performing distribution for shorter time scales of 3- and 6-months with rejection frequencies as 48.03% and 11.02%. However, for the 12-month time scale all three distributions performed similarly (30.01% for GEV, 29.27% for LL and 26.77% for PT-III). For 24-month time scale the performance of LL distribution was found to be the best. Figure 9 shows the accepted and rejected grids with K–S test for various accumulated periods and with each

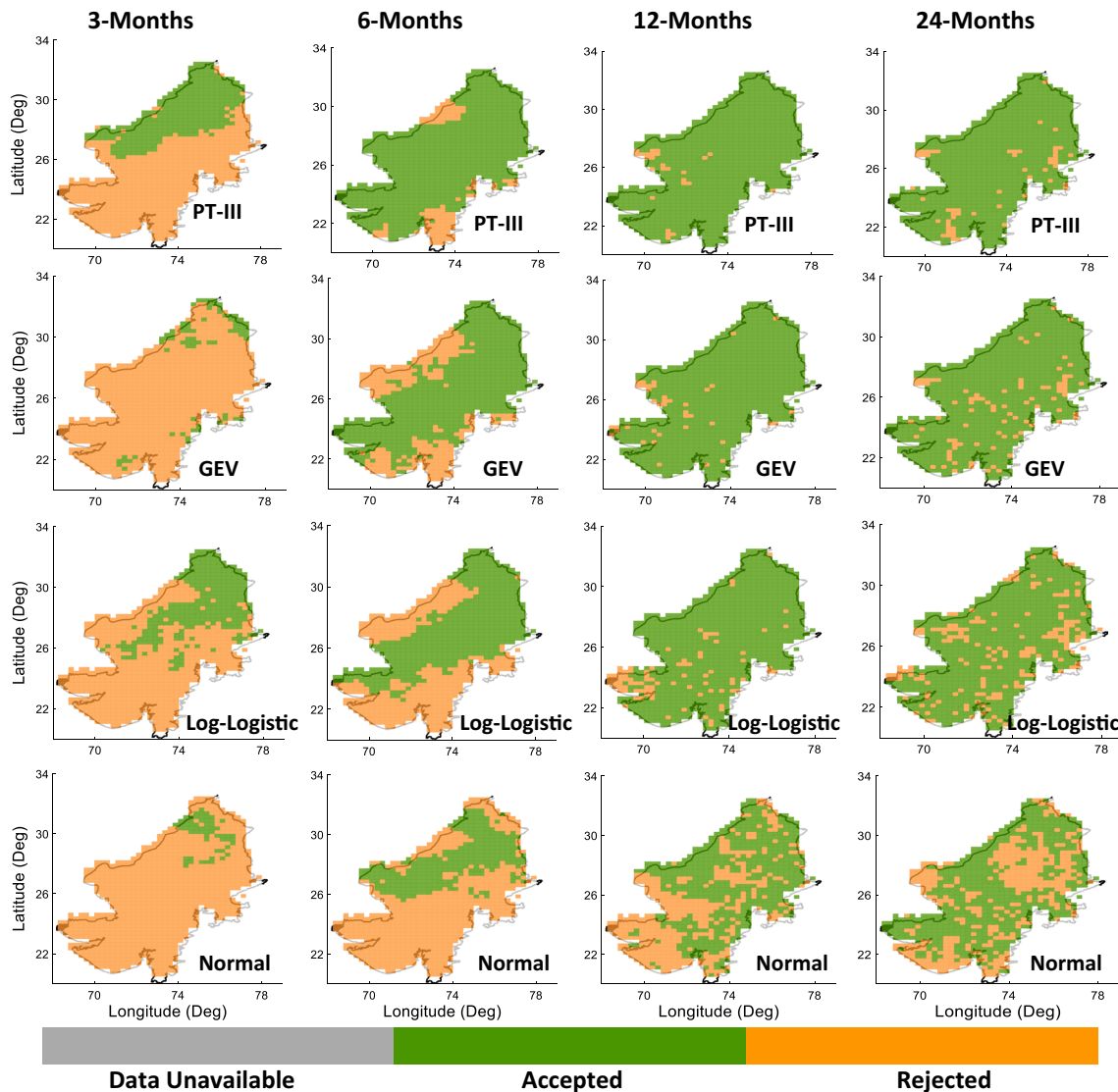


Figure 6. Results from K–S test for West zone with accepted and rejected grid points at a significance level of 0.05 for each probability distribution for the accumulated periods of 3, 6, 12 and 24 months.

distribution for Jammu and Kashmir Zone. For Jammu and Kashmir zone, GEV was the best fitting distribution at shorter time scales (1.92% and 2.74% for 3- and 6-month time scales). For longer time scales the performance of all three distributions was comparable (figures 2 and 9), with none of the distributions able to fit the data in the North of Jammu and Kashmir zone correctly.

Table 1 gives the performance of each probability distribution to fit the SPEI values for various time scales of 3, 6, 12 and 24 months for all over India. The Normal distribution was observed to fit the SPEI with highest rejection frequency for various accumulated time periods of D values all over India. For all accumulation periods PT-III distribution was noted to be having the least rejection frequency all over India according to the K–S test.

However, it is to be noted that at longer time scales (12 and 24 months) the difference in rejection frequencies for all the distributions is very small. Therefore, for long-accumulation periods, we make use of AIC to determine the best performing distribution all over India. The present study used AIC relative ranking to further investigate the best fitted distribution all over India for various time scales. As AIC produces only relative ranking, we cannot comment on the absolute accuracy or fit of the distribution depending on it (Sienz *et al.* 2012). The AIC values (equation 23) are estimated for each grid point for a given probability distribution at various time scales all over India. The AIC values of each distribution are ranked and the best distribution is assigned to each grid point based on the minimum AIC value for a given accumulated

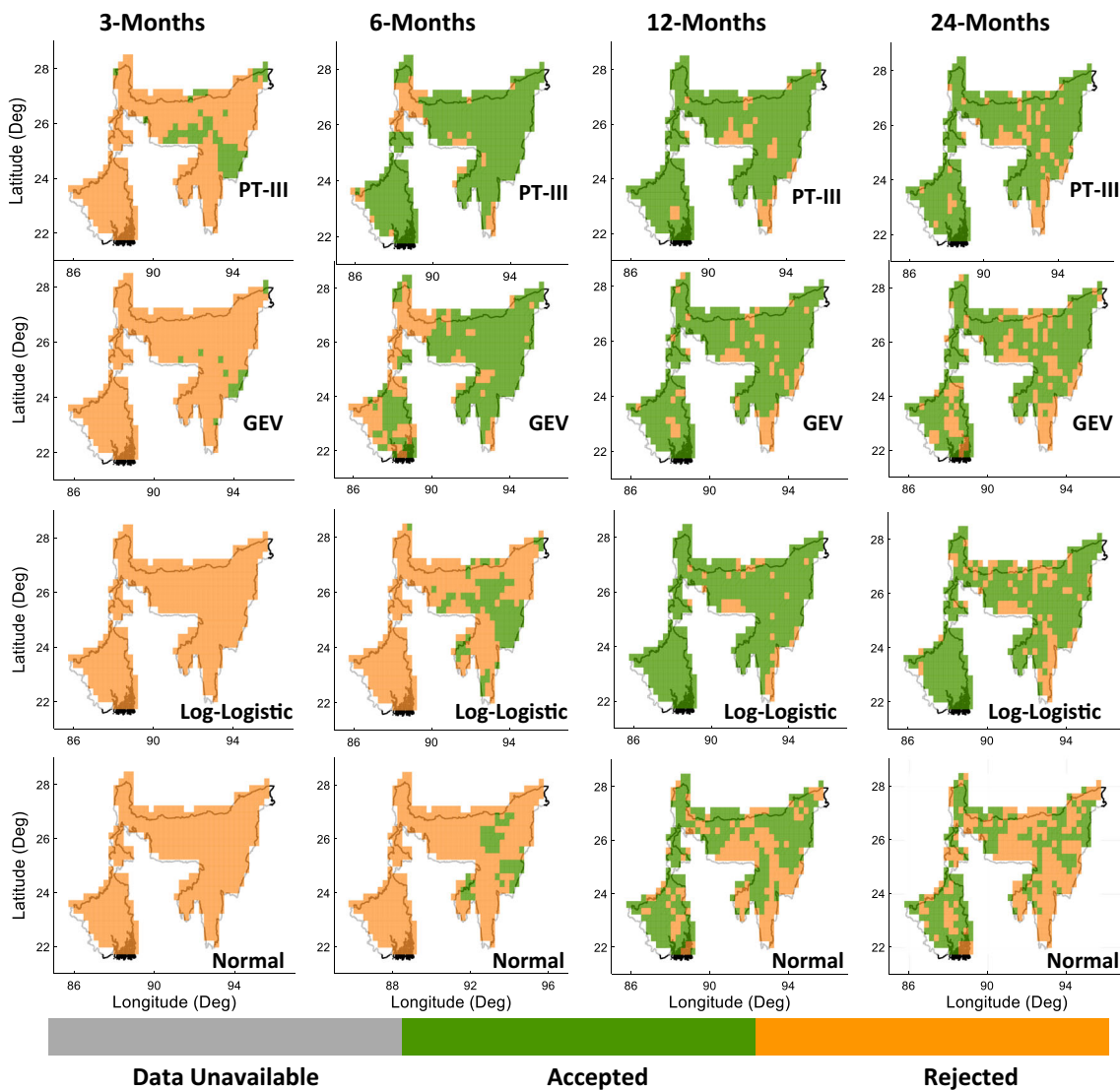


Figure 7. Results from K–S test for North-East zone with accepted and rejected grid points at a significance level of 0.05 for each probability distribution for the accumulated periods of 3, 6, 12 and 24 months.

time period. Figure 10 shows the spatial comparison of SPEI distributions with best relative fit, based on AIC for various time scales all over India. For 3 months accumulated period, the PT-III distribution has produced good relative fit for all over India except for the Western Ghats and parts of West zone with about 82.89% of total grids as shown in figure 10. Whereas, the Normal distribution and LL distributions have been performed reasonably good for the arid climate of West zone with 41% of total grids combinedly at 3-month timescale. Based on AIC, the PT-III distribution was found to be best fitted distribution for 6-month accumulated period for all over India with about 71.91% of total grid points. GEV distribution was identified as the best distribution according to AIC in parts of Central, North and South zones covering a total of 25.42% grid points.

For the 12-month time scale, the GEV distribution has performed best with 50.26% of the total grids and LL distribution with 20.33% of the total grids, mostly distributed over the South and North-East regions. Similarly, for the 24-month time scale also, the GEV distribution has performed best with 58.81% of the total grids followed by LL and PT-III distributions at 25.28% and 17.33%, respectively. Furthermore, the LL distribution has been noted to perform reasonably well for the North, North-East, North-East hills and South zones at 12- and 24-month accumulated periods (figure 10) while no such pattern has been observed in the case of PT-III distribution.

With the K–S rejection frequencies and AIC relative distribution rankings, the present study concludes that PT-III distribution is capable of producing good relative fit for SPEI calculation at

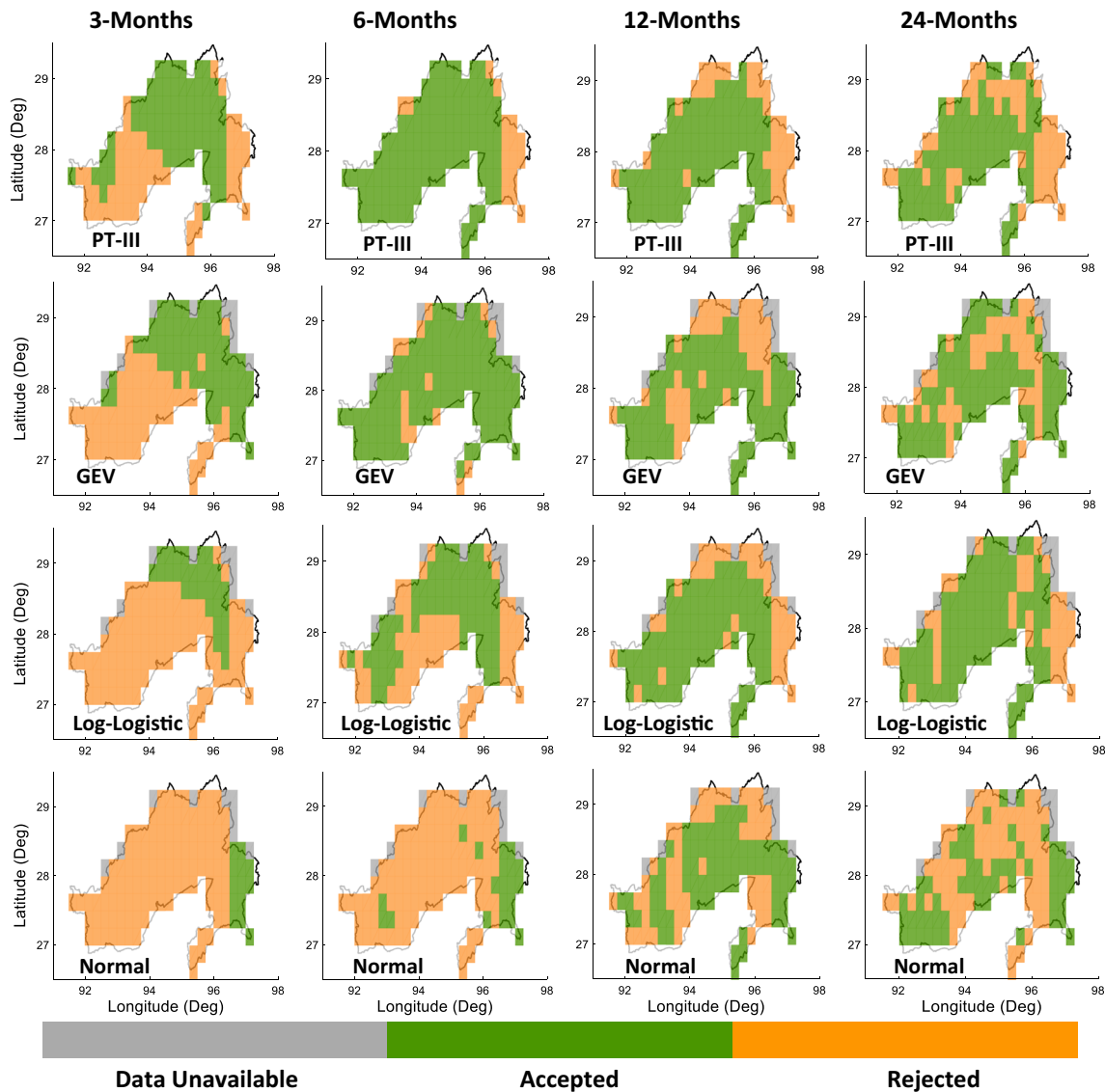


Figure 8. Results from K–S test for North-East hills zone with accepted and rejected grid points at a significance level of 0.05 for each probability distribution for the accumulated periods of 3, 6, 12 and 24 months.

short accumulated periods of 3- and 6-month time scales all over India and for all meteorologically homogeneous zones. Whereas, for long-accumulated periods, although GEV, LL and PT-III distributions can all be adopted for all zones taking into consideration the K–S test results, the use of GEV distribution is recommended over other distributions due to its better fit for most of the grid points which is determined by AIC. There are few evidences in the literature for the adaptation of GEV distribution in the extreme rainfall analysis over India (Alam *et al.* 2015; Chavan and Srinivas 2017). Therefore, the present study recommends PT-III distribution and GEV distribution at shorter and longer time scales respectively, in the drought characterisation all over India with SPEI as drought index.

To study and validate the performance of SPEI resulting from a selected candidate probability distribution after testing for goodness of fit, the present study used Drought Severity Index (DSI), which is a remote sensing-based drought index accounting for the terrestrial evapotranspiration and vegetation greenness products such as Normalized Difference Vegetation Indices (NDVI) derived from satellite observations (Shah and Mishra). A comparison is made between the SPEI-12 calculated using GEV distribution and DSI to evaluate how SPEI performs when compared to a remote sensing data-based drought index. For this purpose, we chose 2002 which has been considered as the worst drought in the modern Indian history. In figure 11, we can observe that the drought intensity patterns as estimated by both SPEI and

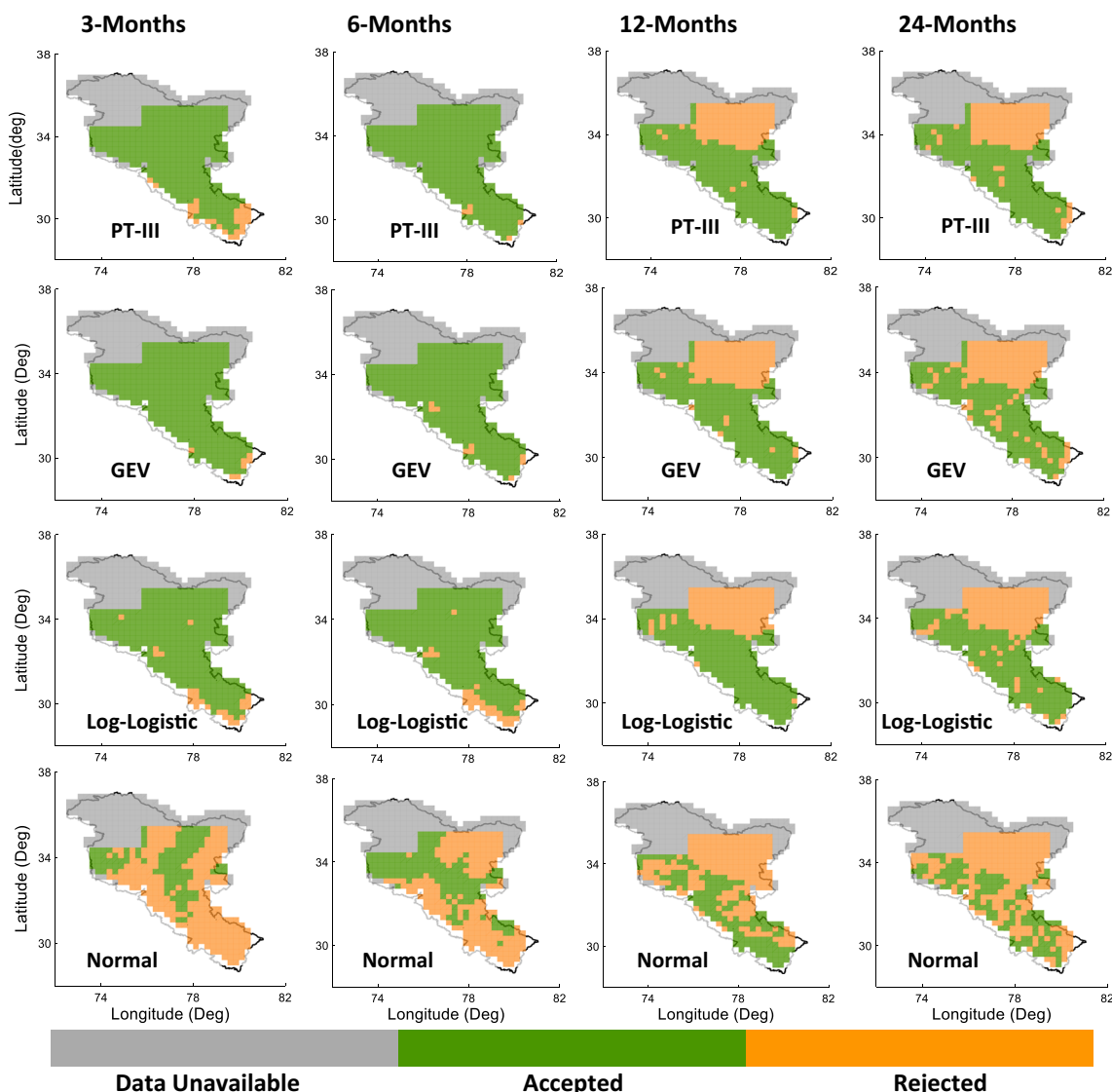


Figure 9. Results from K–S test for Jammu and Kashmir zone with accepted and rejected grid points at a significance level of 0.05 for each probability distribution for the accumulated periods of 3, 6, 12 and 24 months.

Table 1. *K–S rejection frequencies (%) for the probability distributions for various SPEI accumulation periods all over India.*

Distribution	3 months	6 months	12 months	24 months
GEV	59.08	41.08	10.15	17.87
Pearson Type III	57.5	24.46	7.17	11.59
Log Logistic	69.32	69.53	9.09	18.59
Normal	95.57	86.65	38.6	41.04

DSI are in agreement for most regions of India with exceptions being some parts of Central and North-East India. Even though both the index values can theoretically have a range of unlimited negative and positive values, the same value of SPEI and DSI does not indicate the same intensity of drought. This limits us to comparing intensity

patterns rather than intensity metrics. Given that, DSI has been extensively used for drought assessment over India (Shah and Mishra, 2015), the comparison provides value in the drought assessment using SPEI. The spatial extent of the drought as estimated by both the indices with 75% area with SPEI and 71% with DSI. This minor

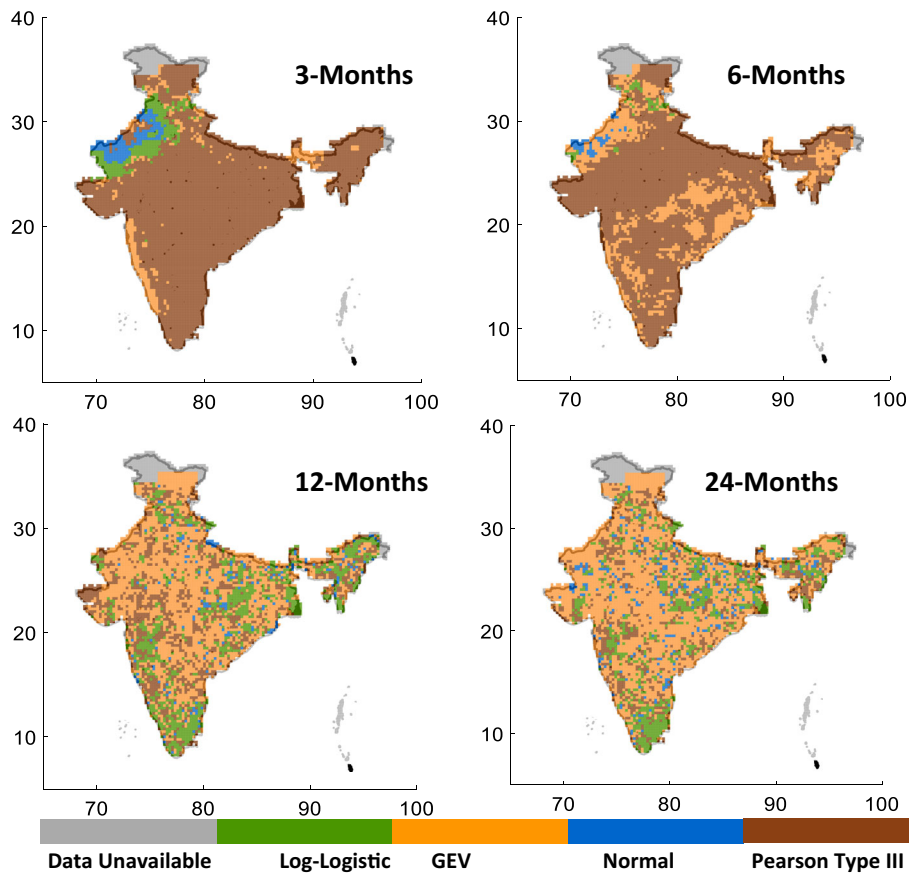


Figure 10. Spatial comparison of SPEI distribution with best relative fit based on AIC for various time scales all over India.

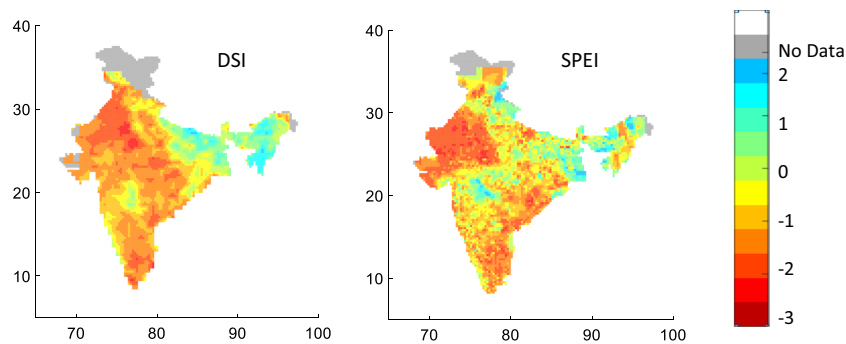


Figure 11. Spatial comparison of GEV based SPEI drought intensities with DSI drought intensities for the year 2002.

difference in spatial extent is due to the estimation of some areas in North-East India as drought by SPEI, but not by DSI.

As India is a country associated with strong seasonality due to monsoon rainfall, it is important that a drought index adopted should capture the seasonality present in the rainfall. The Standardized Precipitation Anomaly Index (SPAI) proposed by Chanda and Maity (2015) accounts for anomalies of the rainfall for capturing the strong seasonal and periodic monsoon-dominated climatology of India. The present study compared

the SPEI based on the selected candidate distribution with SPAI for Central zone to study the ability of SPEI to capture the seasonality. SPAI works by fitting the distribution to the rainfall deficit observed in a particular accumulation period when compared to long term average rather than the actual rainfall in the period. Figure 12 shows the 3-month SPAI-3 and SPEI-3 values for the non-monsoon months of January and May, and for monsoon month of August (accumulated over June, July and August) for all over India from 1951 to 2014. For the year 2002,

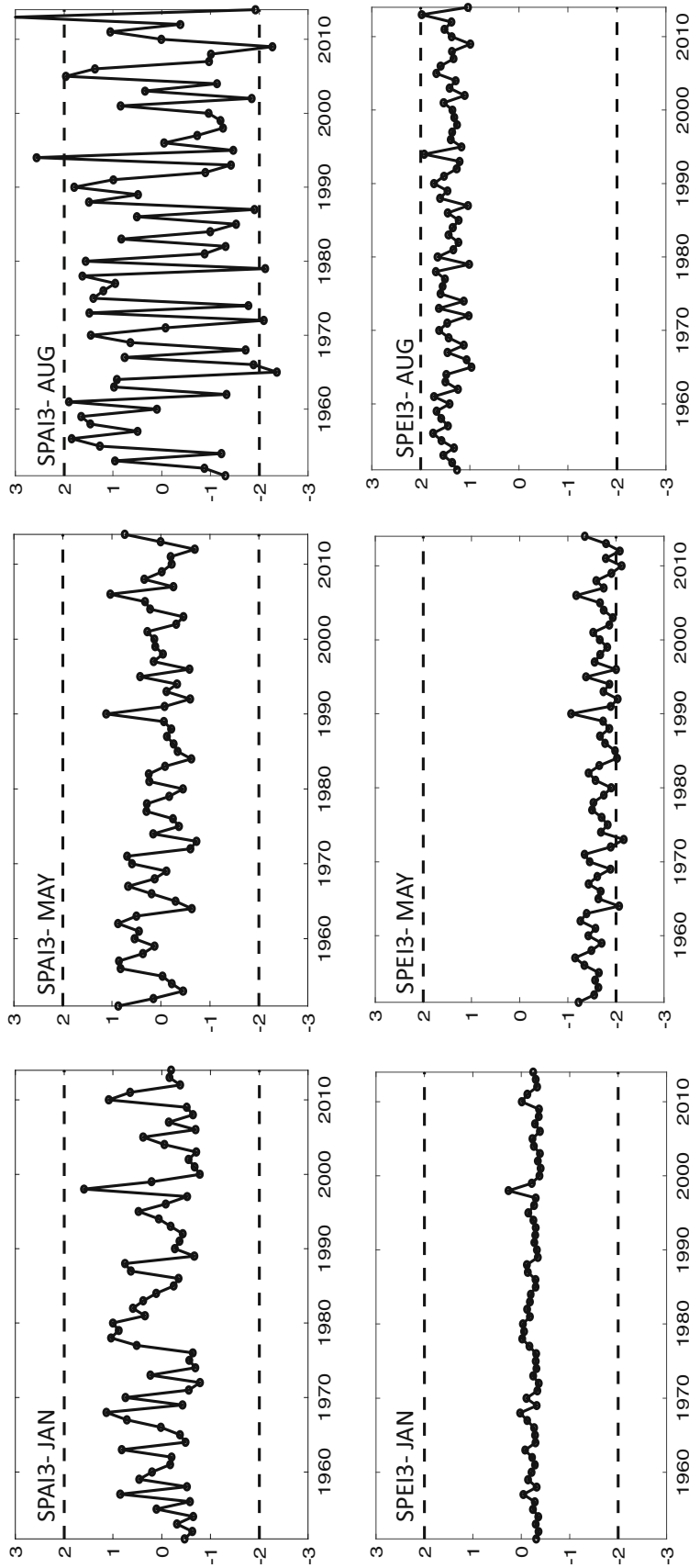


Figure 12. (a) SPI-3 values for Central zone for the months of January, May and August and (b) SPEI-3 values for Central zone for the months of January, May and August.

which is considered to be drought year all over India (Mallya *et al.* 2016), the SPAI-3 index values estimated for the months January, May and August are -0.55 , -0.31 and -1.84 respectively indicating dry conditions corresponding to the effect of rainfall deficit. That is, a deficiency of rainfall in August which is a monsoon month has much more impact than the deficiency of rainfall in January/May which are non-monsoon months (Chanda and Maity 2015). In case of SPEI, the index values for the year 2002 for January, May and August are -0.33 , -1.86 and 1.10 , respectively. It can be observed that even though 2002 was one of the worst drought years, SPEI3 value in August indicating wet conditions. This happens as we fit the distribution to the entire time series without factoring in the seasonal aspects of the rainfall. From figure 12, the values of SPEI3 for a particular month always tend to lie in a confined range (>1 for August and <-1 for May) which does not represent the drought scenario well in case of seasonal rainfall. Therefore, it can be concluded that the SPEI-3 with selected candidate distribution was not able to capture the seasonal variability in comparison to SPAI.

4. Conclusions

The present study compared the performance of a set of probability distributions to fit the climatic water balance time series of SPEI formulation. Four probability distributions of Generalized Extreme Value, Log-Logistic, Pearson Type III and Normal distributions were compared in terms of K-S rejection frequencies and AIC relative ranking for seven meteorologically homogeneous zones and all over India. The Normal distribution was found to be the distribution with highest rejection frequencies, also with less number of AIC ranked grid points.

For North, West, and North-East zones, none of the distributions under consideration fit the 3-month accumulation period of $D(P-PET)$ based on the K-S rejection frequencies ($>70\%$). GEV distribution for Central, and Jammu and Kashmir zones and Pearson Type III distribution for rest of the zones were observed as the best fitting distributions at 3-month time scale. For 6-month time-scale Pearson Type III distribution was identified as the best fitting distribution for all zones except North-East hills where GEV was observed as the best distribution according to

K-S test. For 12- and 24-month time scales, all the distributions were observed to have similar (with difference $\approx 5\%$) K-S rejection frequencies.

The relative ranking by AIC revealed Pearson Type III distribution as the best fit for SPEI values all over India at shorter time scales with total grid points as 82.89% for 3-month time scale and 71.91% for 6-month time scale. Whereas, for longer time scales, GEV distribution has performed best with 50.26% and 58.81% of the total grids at 12-month and 24-month time scale respectively. The study also concludes that LL distribution performs reasonably well for the North, North-East, North-East hills and South zones at 12 and 24-month accumulated periods. Considering this, it would be advantageous to adopt Pearson Type III over LL all over India and for each meteorological homogeneous zone for short accumulated periods of 3- and 6-month time scales. Although, both GEV, LL and PT-III distributions have similar rejection frequencies at long accumulation periods of 12 and 24-month scales, GEV provided better fit for most of the grid points according to AIC.

The GEV based SPEI12 and remote sensing-based Drought Severity index (DSI) drought spatial extents were found to be comparable for most of regions in India. Also, the comparison of SPEI-3 with SPAI-3 revealed that SPEI is unable to capture the seasonality associated with the Indian rainfall patterns. Considering the variability of rainfall characteristics over India, the recommendation of this study is to adopt a suitable probability distribution for each zone for a given time scale in the SPEI drought characterisation. The results of the present study will be useful in the drought characterisation and river water management adaption with respect to each meteorological homogeneous zone of India.

Acknowledgements

The research work presented in the manuscript is funded by Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India through Start-up Grant for Young Scientists (YSS) Project no. YSS/2015/002111. The authors sincerely thank the Editor and the anonymous reviewers for reviewing the manuscript and providing insightful comments.

References

- Aadhar S and Mishra V 2017 High-resolution near real-time drought monitoring in South Asia; *Sci. Data* **4** 170145.
- Akaike H 1974 A new look at the statistical model identification; *IEEE T. Automat. Contr.* **19(6)** 716–723.
- Alam N M, Raizada A, Jana C, Meshram R K and Sharma N K 2015 Statistical modeling of extreme drought occurrence in Bellary District of Eastern Karnataka; *Proc. Nat. Acad. Sci. India B.* **85(2)** 423–430.
- Alam N M, Sharma G C, Moreira E, Jana C, Mishra P K, Sharma N K and Mandal D 2017 Evaluation of drought using SPEI drought class transitions and log-linear models for different agro-ecological regions of India; *Phys. Chem. Earth Parts A/B/C* **100** 31–43.
- Allen R G, Pereira L S, Howell T A and Jensen M E 2011 Evapotranspiration information reporting: I. Factors governing measurement accuracy; *Agr. Water Manag.* **98(6)** 899–920.
- Bhalme H N and Mooley D A 1980 Large-scale droughts/floods and monsoon circulation; *Mon. Weather Rev.* **108(8)** 1197–1211.
- Bordi I, Fraedrich K, Jiang J M and Sutera A 2004 Spatio-temporal variability of dry and wet periods in eastern China; *Theor. Appl. Climatol.* **79(1–2)** 81–91.
- Chakravarty I M, Roy J D and Laha R G 1967 *Handbook of methods of applied statistics*.
- Dai A 2011 Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008; *J. Geophys. Res. Atmos.* **116(D12)**.
- Chanda K and Rajib Maity 2015 Meteorological Drought Quantification with Standardized Precipitation Anomaly Index (SPAI) for the Regions with Strongly Seasonal and Periodic Precipitation; *J. Hydrol. Eng., American Society for Civil Engineering (ASCE)* **20(12)** 06015007-1–06015007-8, [https://doi.org/10.1061/\(asce\)he.1943-5584.0001236](https://doi.org/10.1061/(asce)he.1943-5584.0001236).
- Chavan S R and Srinivas V V 2017 Regionalization based envelope curves for PMP estimation by Hershfield method; *Int. J. Climatol.* **37(10)** 3767–3779.
- Doorenbos J and Pruitt W O 1977 Background and development of methods to predict reference crop evapotranspiration (ET₀); *Crop Water Requirements. FAO Irrigation and Drainage Paper* **24** 108–119.
- Ghosh S and Mujumdar P P 2007 Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment; *Water Resour. Res.* **43(7)**.
- Goswami B N, Venugopal V, Sengupta D, Madhusoodanan M S and Xavier P K 2006 Increasing trend of extreme rain events over India in a warming environment; *Science* **314(5804)** 1442–1445.
- Guhathakurta P 2003 Droughts in districts of India during the recent all India normal monsoon years and its probability of occurrence; *Mausam* **54(2)** 542–544.
- Hargreaves G H 1975 Moisture availability and crop production; *T. ASAE* **18(5)** 980–984.
- Hargreaves G H and Samani Z A 1982 Estimating potential evapotranspiration; *J. Irr. Drain Div-asce* **108(3)** 225–230.
- Hargreaves G H and Samani Z A 1985 Reference crop evapotranspiration from temperature; *Appl. Eng. Agric.* **1(2)** 96–99.
- Homdee T, Pongput K and Kanae S 2016 A comparative performance analysis of three standardized climatic drought indices in the Chi River basin, Thailand; *Agr. Nat. Resour.* **50(3)** 211–219.
- Иванов (Ivanov) НН (1954) Об определении величин испаряемости. (About determination of evapotranspiration); *Известия Всес Геогр Общества Т* **86(2)** 189–195.
- Jensen M E, Burman R D and Allen R G (eds) 1990 Evapotranspiration and irrigation water requirements. ASCE Manuals and Reports on Engineering Practices, No. 70.
- Kumar K N, Rajeevan M, Pai D S, Srivastava A K and Preethi B 2013 On the observed variability of monsoon droughts over India; *Weather Clim. Extremes* **1** 42–50.
- Liu Y, Zhu Y, Ren L, Singh V P, Yang X and Yuan F 2017 A multiscalar Palmer drought severity index; *Geophys. Res. Lett.* **44(13)** 6850–6858.
- Mallya G, Mishra V, Niyogi D, Tripathi S and Govindaraju R S 2016 Trends and variability of droughts over the Indian monsoon region; *Weather Clim. Extremes* **12** 43–68.
- McKee T B, Doesken N J and Kleist J 1993 The relationship of drought frequency and duration to time scales; In: *Proceedings of the 8th Conference on Applied Climatology*, American Meteorological Society, Boston, MA, **17(22)** 179–183.
- Nath R, Nath D, Li Q, Chen W and Cui X 2017 Impact of drought on agriculture in the Indo-Gangetic Plain, India; *Adv. Atmos. Sci.* **34(3)** 335–346.
- Pai D S, Sridhar L, Guhathakurta P and Hatwar H R 2011 District-wide drought climatology of the southwest monsoon season over India based on standardized precipitation index (SPI); *Nat. Hazards* **59(3)** 1797–1813.
- Pai D S, Sridhar L, Rajeevan M, Sreejith O P, Satbhai N S and Mukhopadhyay B 2014 Development of a new high spatial resolution (0.25×0.25) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region; *Mausam* **65(1)** 1–18.
- Parthasarathy B, Rupa Kumar K and Munot A A 1996 *Homogeneous regional summer monsoon rainfall over India: interannual variability and teleconnections*; Indian Institute of Tropical Meteorology.
- Penman H L 1948 Natural evaporation from open water, bare soil and grass; *Proc. R. Soc. Lond. A* **193(1032)** 120–145.
- Rajeevan M and Bhate J 2009 A high resolution daily gridded rainfall dataset (1971–2005) for mesoscale meteorological studies; *Curr. Sci.* **96(4)** 558–562.
- Rajeevan M, Gadgil S and Bhate J 2010 Active and break spells of the Indian summer monsoon; *J. Earth Syst. Sci.* **119(3)** 229–247.
- Rajeevan M, Bhate J and Jaswal A K 2008 Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data; *Geophys. Res. Lett.* **35(18)**.
- Saikranthi K, Rao T N, Rajeevan M and Bhaskara Rao S V 2013 Identification and validation of homogeneous rainfall zones in India using correlation analysis; *J. Hydrometeorol.* **14(1)** 304–317.
- Shah R D and Mishra V 2015 Development of an experimental near-real-time drought monitor for India; *J. Hydrometeorol.* **16(1)** 327–345.
- Van der Schrier G, Jones P D and Briffa K R 2011 The sensitivity of the PDSI to the Thornthwaite and Penman-

- Monteith parameterizations for potential evapotranspiration; *J. Geophys. Res.-Atmos.* **116(D3)**.
- Sheng Y and Hashino M 2007 Probability distribution of annual, seasonal and monthly precipitation in Japan; *Hydrol. Sci. J.* **52(5)** 863–877.
- Shepard D 1968 A two-dimensional interpolation function for irregularly spaced data; Proc. 1968 ACM Nat. Conf., pp 517–524.
- Sienz F, Bothe O and Fraedrich K 2012 Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias; *Hydrol. Earth Syst. Sci.* **16(7)** p. 2143.
- Srivastava A K, Rajeevan M and Kshirsagar S R 2009 Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region. *Atmos. Sci. Lett.* **10(4)** 249–254.
- Stagge J H, Tallaksen L M, Gudmundsson L, Van Loon A F and Stahl K 2015 Candidate distributions for climatological drought indices (SPI and SPEI). *Int. J. Climatol.* **35(13)** 4027–4040.
- Stagge J H, Tallaksen L M, Gudmundsson L, Van Loon A F and Stahl K 2016 Response to comment on ‘candidate distributions for climatological drought indices (SPI and SPEI)’. *Int. J. Climatol.* **36(4)** 2132–2138.
- Stagge J H, Tallaksen L M, Xu C Y and Van Lanen H A J 2014 Standardized precipitation-evapotranspiration index (SPEI): Sensitivity to potential evapotranspiration model and parameters; *IAHS Publ.*, pp. 367–373.
- Thornthwaite C W 1948 An approach toward a rational classification of climate; *Geogr. Rev.* **38(1)** 55–94.
- Vicente-Serrano S M, Beguería S and López-Moreno J I 2010 A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index; *J. Clim.* **23(7)** 1696–1718.
- Vicente-Serrano S M, Van der Schrier G, Beguería S, Azorin-Molina C and Lopez-Moreno J I 2015 Contribution of precipitation and reference evapotranspiration to drought indices under different climates; *J. Hydrol.* **526** 42–54.
- Vicente-Serrano S M and Beguería S 2016 Comment on ‘Candidate distributions for climatological drought indices (SPI and SPEI)’ by James H Stagge et al., *Int. J. Climatol.* **36(4)** 2120–2131.
- Wang H, Chen Y, Y Pan, Chen Z and Ren Z 2019 Assessment of candidate distributions for SPI/SPEI and sensitivity of drought to climatic variables in China; *Int. J. Climatol.*, <https://doi.org/10.1002/joc.6081>
- World Meteorological Organization 2012 Standardized Precipitation Index User Guide, WMO-No. 1090, ISBN 978-92-63-11090-9.
- Wu H, Svoboda M D, Hayes M J, Wilhite D A and Wen F 2007 Appropriate application of the (Anon., n.d.)standardized precipitation index in arid locations and dry seasons; *Int. J. Climatol.* **27(1)** 65–79.

Corresponding editor: RAJIB MAITY