MAUSAM, **73**, 1 (January 2022), 139-150

551.509 : 551.577.551.573 (540.29)

Best fit probability distribution analysis of precipitation and potential evapotranspiration of India's highly dense population state - Bihar

VIKRAM KUMAR, SHAKTIBALA* and BHARTESH**

Department of Civil Engineering, Gaya College of Engineering, Bihar – 823 003, India

** Department of Civil Engineering, Poornima Group of Institution, Jaipur – 302 022, India*

*** Department of Civil Engineering, A. P. Goyal Shimla University, Shimla – 171 013, India*

(*Received 21 May 2020, Accepted 29 September 2021*)

e mail : 25.vikram@gmail.com

सार – वर्षा और अन्य मौसम विज्ञान संबंधी भौतिक विशेषताओं के साथ संदिग्धता और गैर-एकरूपता के कारण जल संसाधनों और इसके प्रबंधन की योजना संभवत: बिहार में कृषि उत्पादन पर प्रभाव डाल सकती है जहां किसान ज्यादातर वर्षा पर निर्भर हैं। राज्य में वर्षा और वाष्पीकरण क्षमता अस्थायी वितरण भू-आकृति, जलवायविक और राज्य के अन्य मानवजनित कारकों के कारण अनियमित है। इस शोध पत्र में, राज्य के सभी 37 जिलों के पिछले 102 साल के रिकॉर्ड के आधार पर वार्षिक औसत वर्षा और वाष्पीकरण क्षमता के विभिन्न उपलब्ध संभाव्यता वितरण के बीच सबसे उपयुक्त संभाव्यता वितरण को उजागर करने का प्रयास किया गया है। कोलमोगोरोव स्मिरनोव, एंडरसन डार्लिंग और ची-स्क्वायर जैसे फिट परीक्षणों की अच्छाई की श्रेणियों के आधार पर, सामान्य वितरण को 11 जिलों के लिए सबसे उपयुक्त संभाव्यता वितरण के रूप में देखा गया, इसके बाद 9 जिलों के लिए वेइबुल (3पी), 5 जिलों के वर्षा के लिए बीटा वितरण और बाकी जिलों के लिए अन्य वितरण को लिया गया। जबकि कॉची वितरण सभी जिलों की वाष्पीकरण क्षमता के लिए सबसे उपयुक्त संभाव्यता वितरण रहा और दूसरा सबसे अच्छा गामा (3पी) रहा जिसने कुल जिलों के लगभग 60% को कवर किया,जिसके बाद सामान्य चरम मान वितरण(32%) रहा। इन परिणामों का उपयोग भविष्य के द्रवचालित डिजाइनया नीति विकास के लिए वापसी अवधि के आकलन हेतु जल विज्ञानिक अध्ययन और जल संसाधन योजनाकारों द्वारा किया जा सकता है।

ABSTRACT. Planning of water resources and its management with the ambiguity and non-uniformity accompanying with precipitation and other meteorological physical characteristics may perhaps effect on agricultural production in Bihar where the farmers mostly depend on precipitation. The precipitation and potential evapotranspiration temporal distribution of the state is irregular due to geomorphology, climatic and other anthropogenic factors of the state. In the present study, attempt is taken to expose the best-fit probability distribution among the various available probability distribution of annual average precipitation and potential evapotranspiration based on 102 year of past records of all 37 districts of the state. On the basis of ranks of goodness of fit tests such as Kolmogorov Smirnov, Anderson Darling and Chi-Squared, the normal distribution was observed the best-fit probability distribution for 11 districts followed by Weibull (3P) for 9 districts, the Beta distribution for 5 districts and other distribution for rest districts for precipitation. Whereas Cauchy distribution was come out with the best-fit probability distribution for potential evapotranspiration for all districts and the second best was Gamma (3P) covering almost 60% of the total districts followed by General Extreme Value distribution (32%). The results can be used in future hydraulic design, hydrological study for estimation of return period and water resource planners for policy development.

Key words – Probability distribution, Potential evapotranspiration, Precipitation, Water resource management.

1. Introduction

One of the key natural resources of water on the earth for domestic and agriculture production is precipitation in the form of rainfall in the maximum part of the county which contributes to ponds, aquifer and rivers. The precipitation distribution pattern and potential evapotranspiration of any location are much significant parameters upsetting various agriculture growth, management of water resources, flood and drought study, forestry and tourism and socio-economic systems (Kumar and Sen, 2020). Understanding the variability and estimation of forthcoming precipitation and rate of evapotranspiration at certain location based on the past records has ascertained to be difficult in terms of future occurrence. The well-ordered understanding the precipitation form of water and wise use of precipitation can be improved by analyzing precipitation and potential evapotranspiration using probability distributions (Subudhi, 2007). The understanding of the best fit distribution is not to explore the properties of precipitation, its occurrence or circulation, but to use the past recorded sequences of precipitation and other meteorological parameters as feedbacks for another hydraulic/hydrological modeling to see the various mechanism subsequently (Nanda *et al*., 2018). Distribution fitting is the way of selecting a statistical distribution that best fits to past records. The use of statistics and probability distribution has a wide range of important applications in climate research (Storch, 1999).

Bruhn *et al*. (1980) first applied the probability distribution over the minimum and maximum temperature, precipitation and humidity using the daily recorded data (1946-1975) at Geneva, New York for developing the model for agriculture. Study results showed that precipitation series is better following the Gamma distribution whereas minimum and maximum temperature is described by Normal (3P) distribution and humidity is considered as Normal distributed. The simulation results further argue that Beta distribution could further reduce the simulation results of meteorological variables. For the future prediction of the precipitation in the region (Bhakar *et al*., 2008) found that best fitting of the precipitation data is using the Gumbel distribution for monthly and weekly data. In the Himalayan and mountainous catchment (Kumar *et al*., 2017) interpreted the annual mean precipitation of complete Uttarakhand state covering wide spectrum of elevation difference and found that Weibull distribution performance was the best in more than 7 districts followed by Chisquared (2P) and log-Pearson. The Normal distribution was observed best in describing the 24 hr precipitation at Himalayan, Karakoram, Hindu Kush and Pamir mountain (Amin *et al*., 2016). In Bangladesh (Alam *et al*., 2018 and Sanjib, 2016) studied the probability distribution over the monthly precipitation of more than 25-year data for better understanding of agriculture and engineering design and found that Generalized Extreme Value (GEV) distribution offers a good fit for Bangladesh.

Francesco and Maurizio (2013) developed new coupled spatial (CS) distribution for precipitation and temperature in USA after analyzing records of about 8000 locations for temperature and 12000 locations for precipitation during the 1895-1997. Not only the meteorological but hydrological aspects have also been studied on rivers in China (Zhanling, 2014) and showed that GEV, Burr and Weibull distributions were best distribution and Exponential and Pareto 2 distributions performs worst. Probability distribution function is also being applied to retrieve the precipitation data from the satellite (Wilheit *et al*., 1991). Log-Pearson (3P) distribution followed by Pearson (3P) performed well in Nigeria cities for peak daily precipitation data of 54 years (Olofintoye *et al*., 2009).A new joint probability distribution function was developed for study the precipitation and evapotranspiration effect together by considering the rainfall and ET_0 well distributed by Pearson (3P) in Henan Province of China (Zhang *et al*., 2017). In Japan, annual precipitation of 32 stations showed that, the log-Pearson (3P) distribution is best suitable and GEV could be considered as a second alternative (Yue and Michio, 2007). Four-parameter kappa probability distribution was found best fir study the flow duration curve using the daily streamflow data in most of the gauging locations in US (Blum *et al*., 2017). For quickness, only key worthy studies have been reported above just to elucidate the best-fit distribution assortment.

Above past several studies irrespective of the purposes shows that individual probability distribution has their own characteristics with wide applicability. Evaluating the suitable probability distribution in Bihar state where agriculture is the foremost profession of the rural or poor inhabitants is indispensable. In Bihar, to the authors' information, very limited or even there has lack of meticulous study concerning about the probability distribution on the any meteorological parameters. Furthermore, the choice of a suitable probability distribution is the foremost problems in any other engineering practice. Present study provides an opportunity to understand and decide the best probability distribution types among all possible twenty (20) types for annual precipitation and potential evapotranspiration (PE) which are two key parameters to understand for agriculture perspective. Present study aims to contribute to not only suggesting the best fit distribution for all the 37 districts of the Bihar using the more than 100 years of records but also constructive suggestions and understanding for the better productivity under the decline in the water resources. Author's perspective is to emphasis the need to judiciously and efficaciously manage our natural resources for the present and keeping the future needs. The following structure is followed in the paper: section 2 describes the general area description and physiology, section 3 summaries the data used and methodology with approach followed for identifying the best distribution, section 4 will be emphasized on the results obtained with the discussions and at last section will be focused towards the recommendations and conclusion.

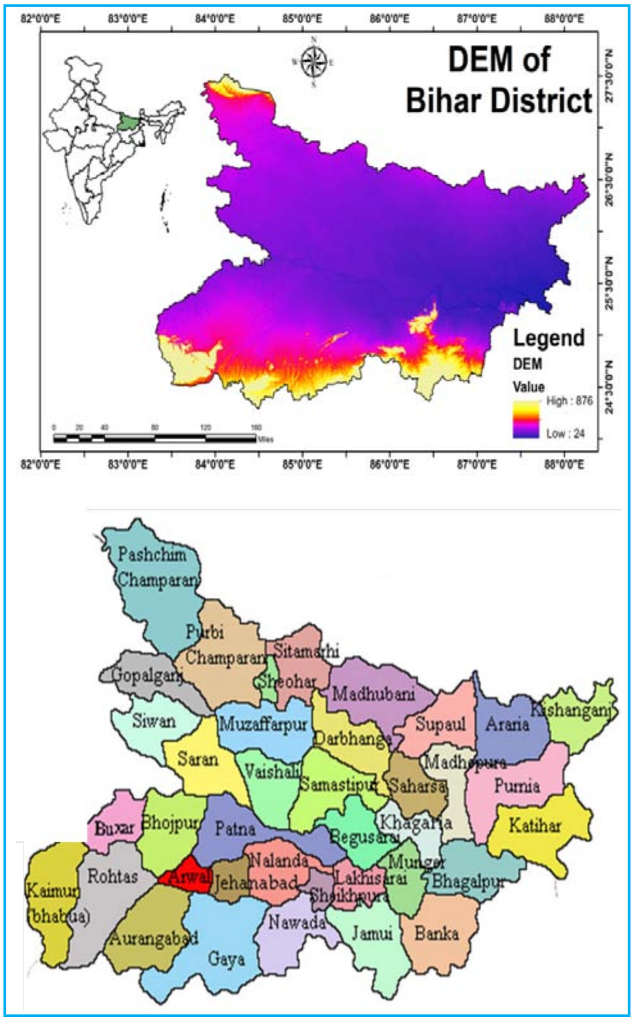


Fig. 1. DEM of the study location and district distribution

2. Study area

Bihar lies in north-east India in the Ganga plains located between N24° 15'-27° 23' Latitudes and E83° 20'-88º 00' Longitudes (Fig. 1) and one of the noticeable states with 38 districts of India that contributes to GDP through the agriculture. As per the agro-climatic based classification, whole Bihar is distributed in three broad regions, *i.e.*, North-West (NW), North-East (NE) and South (S). The NW region comprises of 13 districts with annual precipitation ranges between 1040-1450 mm. NE region has 8 districts with annual precipitation ranges between 1200-1700 mm and the last S region has remaining 17 districts with annual precipitations ranges to 990-1300. The basic physiography of 37 districts is summarised in the Table 1. Arwal district is formed few years back and its data were not available, thus, Arwal has excluded in the present study. The Government also acknowledges and reported that uneven spatial and temporal precipitation along with the rate of evaporation across the state is one of the major challenges for the irrigation planning and scheduling.

TABLE 1

Summary of probability distribution with the function expression

S. No.	Distribution Type	Distribution Function Expression
1.	Beta	$f(x) = [(\frac{1}{B(\alpha_1,\alpha_2)})(\frac{(x-a)^{\alpha_1-1}(b-x)^{\alpha_2-1}}{(b-a)^{\alpha_1+\alpha_2-1}})]$
2.	Cauchy	$f(x) = [\pi\sigma\,(1+(\frac{x-\mu}{\sigma})^2)]^{-1}$
3.	Chi-Squared	$f(x) = \frac{(x-\Upsilon)^{\frac{v}{2}-1} exp^{(\frac{-(x-\Upsilon)}{2})}}{(2^{\frac{v}{2}})(r(\frac{v}{2}))}$
4.	Chi-Squared (2P)	$f(x) = \frac{x^{\frac{v}{2}-1} exp^{(\frac{-x}{2})}}{(2^{\frac{v}{2}})(r(\frac{v}{2}))}$
5.	Exponential	$f(x) = \lambda exp^{(-\lambda x)}$
6.	Exponential (2P)	$f(x) = \lambda exp^{[-\lambda(x-\Upsilon)]}$
7.	Gamma	$f(x) = \frac{(x)^{\alpha-1} exp^{-\frac{(x)}{\beta}}}{(\beta^{\alpha})(\Gamma\alpha)}$
8.	Gamma (3P)	$f(x) = \frac{(x-\Upsilon)^{\alpha-1} exp^{-\frac{(x-\Upsilon)}{\beta}}}{(\beta^{\alpha})(\Gamma\alpha)}$
9.	Gen. Extreme Value	$f(x) = \frac{1}{\sigma}[1-k(\frac{x-\mu}{\sigma})]^{\frac{1}{k-1}} exp^{[-(1-k(\frac{x-\mu}{\sigma})]^{\frac{1}{k}}}$
10.	Log-Gamma	$f(x) = \frac{[\ln(x)]^{\alpha-1}}{x\beta^{\alpha}(\Gamma\alpha)} exp^{-\frac{ln(x)}{\beta}}$
11.	Log-Pearson 3	$f(x) = \frac{1}{x\,\lvert\beta\rvert\,(\Gamma\alpha)} \frac{(\ln(x)-\Upsilon)^{\alpha-1}}{\beta} exp^{-\frac{ln(x-\Upsilon)}{\beta}}$
12.	Lognormal	$f(x) = \frac{\exp(\frac{-1}{2}(\frac{\ln x}{\sigma})^2)}{x\alpha\sqrt{2\pi}}$
13.	Lognormal (3P)	$f(x) = \frac{\exp(\frac{-1}{2}(\frac{\ln x-\Upsilon}{\sigma})^2)}{x\alpha\sqrt{2\pi}}$
14.	Normal	$f(x) = \frac{\exp(\frac{-1}{2}(\frac{x-\mu}{\sigma})^2)}{\alpha\sqrt{2\pi}}$
15.	Pearson 5	$f(x) = \frac{\exp(\frac{-\beta}{(x)})}{\beta\Gamma(\alpha)\frac{(x)}{\beta}^{\alpha+1}}$
16.	Pearson 5 (3P)	$f(x) = \frac{\exp(\frac{-\beta}{(x-\Upsilon)})}{\beta\Gamma(\alpha)\frac{(x-\Upsilon)}{\beta}^{\alpha+1}}$
17.	Pearson 6	$f(x) = \frac{(x)^{\alpha_1-1}}{B(\alpha_1\alpha_2)(1+x)^{\alpha_1+\alpha_2}}$
18.	Pearson 6 (4P)	$f(x) = \frac{(\frac{x-\Upsilon}{\beta})^{\alpha_1-1}}{\beta B(\alpha_1\alpha_2)(1+\frac{x-\gamma}{\beta})^{\alpha_1+\alpha_2}}$
19.	Weibull	$f(x) = \frac{\alpha}{\beta}(\frac{x}{\beta})^{\alpha-1} exp^{(-\frac{x^{\alpha}}{\beta})}$
20.	Weibull (3P)	$f(x) = \frac{\alpha}{\beta}(\frac{x-\gamma}{\beta})^{\alpha-1} exp^{(-\frac{x-\gamma^{\alpha}}{\beta})}$

In the study area depth of water table lies between 5M in the NE region to 20M in the S region districts. Repeated floods and drought are a thoughtful issue which upsets the farmers of the state. Hydro-geologically the state is categorized into consolidated/fissured formations and unconsolidated/Alluvial formation. Despite the fact that the land is appropriate for a wide scope of yields, rice, wheat and maize are the major cereal produce of Bihar. Sugar cane and jute are two significant cash crops of Bihar in limited districts.

3. Data set and approach used

The precipitation and potential evapotranspiration data from 1901 to 2002 is used for exploring the best fit probability distribution analysis which is being available at http://www.indiawaterportal.org/met_data/ (India Water Portal). The various probability distributions deliberated for fitting the precipitation and potential evapotranspiration data are as follows: Beta, Cauchy, Chi-Squared, Chi Squared (2P), Exponential, Exponential (2P), Gamma, Gamma (3P), Generalized Extreme Value (GEV), Log-Gamma (LG), Log-Pearson 3 (LP), Log-normal, Log-normal (3P), Normal, Pearson 5, Pearson 5 (3P), Pearson 6, Pearson 6 (4P), Weibull, Weibull (3P). The expressions of the above used 20 different distributions is summarized in the Table 1.

Second step is to study the goodness of fit test which are generally applied for inspection of the presumed best fir probability distribution model among the different distribution type. In the present case, nonparametric distribution free tests such as Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Chi-square (χ^2) were used at 5% significance level for best fit distribution on precipitation and potential evapotranspiration data.

3.1. *Kolmogorov-Smirnov (KS) Test*

The KS test is a goodness-of-fit statistic that looks at an observational appropriation function (F_x), with a identified distribution function (F_y). Alternatively, the determined change between (F_x) and (F_y) is the KS test. Commonly KS test is employed as an option in contrast to the chi-square decency of fit test. The KS statistic (D) can be figured out using the relation:

$$D = \max\left|F_x(x) - F_y\right| \tag{1}$$

Obviously, a large difference indicates an inconsistency between the observed data and the statistical model. Some restrictions towards the using of the KS test are that it can be applied merely to continuous distributions and be likely to be extra sensitive towards the centre of distribution than its ends.

3.2. *Anderson-Darling (AD) test*

The AD test was first make known by Theodore Wilbur Anderson and Donald Darling to put extra weight or judicious control towards the ends of the distribution. This test could be of significant when selection of the best-fitted distribution to precipitation and other maxima data. The test statistic of AD is expressed as :

$$AD = -n - \frac{1}{N}\sum_{i=1}^{n} (2i-1)\{\ln(x_i) + \ln[1-(x_{n+1-i})]\} \tag{2}$$

where, $\left[x_{(1)} < .... < x_{(n)}\right]$ is the well-ordered sample of size n and $F(x)$ is the basic hypothetical combined distribution to which the precipitation and evapotranspiration is matched. The null-hypothesis is that $\left[x_{(1)} < ...... < x_{(n)}\right]$ comes from the underlying distribution $F(x)$ is rejected if AD is larger than the critical value AD_α at a given significance level (α). The drawback with AD test is that the critical values has to be evaluated independently for individual distribution type.

3.3. *Chi-square (χ^2) test*

χ^2test is a procedure which involves checking of a particular distribution type for a certain observed precipitation/potential evapotranspiration occurrence times in a sample, regardless of whether they are reasonable for that example or not. Using O to define "recorded data" and E to define "predicted data" then the χ^2test statistic will be expressed using the relation:

$$\chi^2 = \frac{\sum(O-E)^2}{E} \tag{3}$$

Above relation holds good for continuous precipitation and potential evapotranspiration data only and only be used to decide if the precipitation and potential evapotranspiration data comes from a population with a definite distribution type (Sharma and Singh, 2010). This test could be a substitute to the AD and KS goodness-of-fit tests. The major drawback with the test is that, the $\chi^2$is highly sensitive to size of the data that is being used for the analysis. As the sample increases, the difference in the absolute value becomes less and lesser percentage of the predicted value.

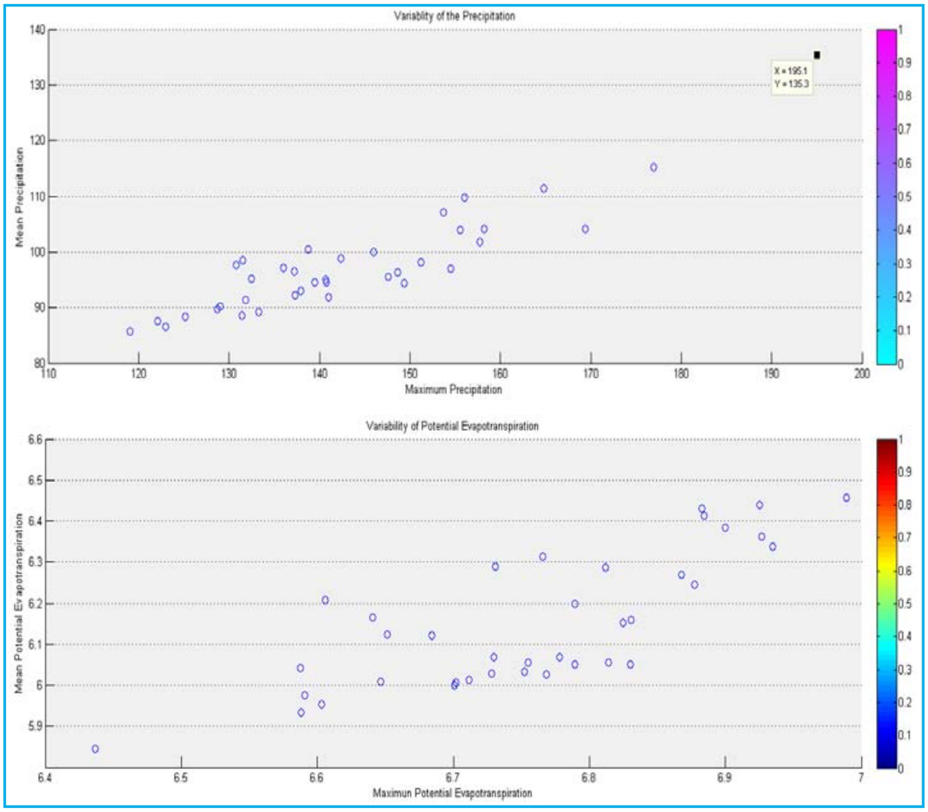


Fig. 2. Variation in the precipitation and evapotranspiration from the annual average value

The next step is to do the ranking of all the applied goodness of fit test (KS, AD, χ^2) on theoretical different probability distributions based on minimum rank. All the used probability distributions are ranked with each goodness of fit test (rank 1 is the best-fit). The combined rank of all 3 goodness fit test to yield a ranking value and the smallest ranking score among all the 20 probability distribution will be the best.

4. Results and discussion

4.1. *Precipitation and potential evapotranspiration data analysis*

The precipitation and potential evapotranspiration data of 37 districts of the Bihar for the period of 1901-2002 (102 years) were analyzed to understand the basic statical characteristics. Monthly data of precipitation and potential evapotranspiration were analyzed for maximum and mean precipitation and potential evapotranspiration, with standard deviation and skewness which is summarized in the Table 2. Table 2 with the Fig. 2 evidently shows the less variation in the precipitation among the different districts but more variability in the potential evapotranspiration.

The lowest amount of average monthly precipitation and potential evapotranspiration occurred in Buxar (85.7cm) and Kishanganj (5.85mm) respectively similarly the maximum average monthly precipitation occurred in Kishanganj (135.3cm) and maximum potential evapotranspiration in Aurangabad (6.45mm). Out of the 37 districts, 11 districts, *viz*., Araria, Banka, Bhagalpur, Jamui, Katihar, Kishanganj, Paschim Champaran, Purbi Champaran Purnia, Sitamadhi and Supaul have mean monthly precipitation more than the 100mm whereas Araria, Kishanganj, Madhubani and Supaul having potential evapotranspiration less than 6mm which favors the condition of the agriculture. Very slight amount of the precipitation recorded during the winter season (January to March) and early summer (April) where water level situation goes worst.

Kumar (2010) and Zakwan *et al*. (2019) reported the decline in the rainfall trend in the Bihar which is challenging to cope up the future agriculture issue. There is more positive increasing trend evaluated at5 districts namely Chapra, Darbhanga, Motihari, Muzaffarour and Supaul in the month of February, March, April, May and October, November, December and January months. Whereas the February, March, April and May months

TABLE 2

Summary statistics of precipitate and evapotranspiration data of Bihar

District	Precipitation (cm)				Potential Evapotranspiration (mm)			
	Maximum	Mean	Std Dev	Skewness	Maximum	Mean	Std Dev	Skewness
Araria	177.0	115.2	18.2	0.4	6.6	5.9	0.1	1.5
Aurangabad	141.0	91.9	19.2	-0.1	7.0	6.5	0.1	1.8
Banka	158.2	104.1	18.3	0.4	6.8	6.1	0.2	1.8
Begusarai	140.7	95.0	16.4	0.1	6.8	6.1	0.1	1.9
Bhagalpur	138.7	100.4	16.4	0.1	6.8	6.1	0.2	1.8
Bhojpur	122.1	87.5	15.5	-0.3	6.9	6.4	0.1	1.7
Buxar	119.1	85.7	15.0	-0.3	6.9	6.4	0.1	1.6
Darbhanga	154.5	97.0	17.4	0.3	6.6	6.0	0.1	1.7
Gaya	147.6	95.6	19.7	0.0	6.9	6.3	0.1	1.9
Gopalganj	128.7	89.7	15.6	-0.2	6.7	6.3	0.1	1.3
Jahanbad	129.0	90.2	16.8	-0.2	6.9	6.4	0.1	1.8
Jamui	157.7	101.9	18.1	0.3	6.8	6.0	0.1	1.9
Kaimur	131.4	88.6	16.8	-0.1	6.9	6.4	0.1	1.6
Katihar	156.0	109.7	16.1	0.0	6.7	6.0	0.1	1.6
Khagariya	137.2	96.4	16.2	0.1	6.8	6.0	0.1	1.8
Kishanganj	195.1	135.3	19.3	0.2	6.4	5.8	0.1	1.4
Lakhisarai	132.5	95.2	16.4	0.0	6.8	6.1	0.1	1.9
Madhepura	130.8	97.6	14.8	-0.2	6.7	6.0	0.1	1.7
Madhubani	131.5	98.5	14.8	-0.1	6.6	6.0	0.1	1.5
Munger	136.0	97.1	16.5	0.1	6.8	6.1	0.1	1.9
Muzaffarpur	139.5	94.5	17.1	0.1	6.7	6.1	0.1	1.6
Nalanda	137.9	93.0	16.8	0.0	6.9	6.2	0.1	1.9
Nawada	151.3	98.2	18.4	0.1	6.8	6.2	0.1	1.9
Paschim Champaran	164.8	111.4	19.9	0.0	6.6	6.2	0.1	1.1
Patna	131.9	91.4	16.2	-0.1	6.9	6.3	0.1	1.8
PurbiChamparan	146.0	100.0	17.8	0.0	6.6	6.2	0.1	1.3
Purnia	153.7	107.1	15.8	0.0	6.7	6.0	0.1	1.6
Rohtas	133.2	89.2	17.1	-0.2	6.9	6.4	0.1	1.7
Samastipur	149.4	94.3	16.7	0.2	6.7	6.1	0.1	1.8
Saran	125.2	88.3	15.3	-0.2	6.8	6.3	0.1	1.6
Saharsha	148.6	96.4	16.2	0.3	6.7	6.0	0.1	1.7
Shaikhpura	140.8	94.6	16.8	0.0	6.8	6.2	0.1	1.9
Sheohar	142.3	98.8	18.0	0.1	6.7	6.1	0.1	1.5
Sitamadhi	155.6	103.9	19.0	0.2	6.6	6.0	0.1	1.4
Siwan	123.0	86.5	15.1	-0.2	6.8	6.3	0.1	1.3
Supaul	169.4	104.2	17.4	0.5	6.6	6.0	0.1	1.5
Vaishali	137.3	92.2	16.4	0.0	6.8	6.2	0.1	1.7

may experience increases in evapotranspiration, particularly under RCP4.5 climate change scenario in 2050 than the January, July, August and September month. The district which may have a higher evapotranspiration under climate change (RCP4.5) in 2050 will be Gaya during the month of February, March, April and May and Bhagalpur, Purnea and Supaul in the month of October, November, December and January (Tesfaye *et al*., 2017). The monthly average precipitation and potential evapotranspiration of all the 37 districts specifies that the evapotranspiration is strongly positively skewed for all the districts whereas for the precipitation it shows negative skewness for 13 districts (Aurangabad, Bhojpur, Buxar, Gopalganj, Jahanabad, Kaimur, Madhepura, Madhubani, Patna, Rohtas, Saran and Siwan) and positive skewness for rest other districts. The skewness is greater than 1 for the evapotranspiration, which is observed as highly skewed whereas the precipitation has skewness less than 1thusconsidered as moderately skewed and this indicated the deviation from its mean precipitation value.

The above analyzed precipitation data indicated an intensification in the precipitation during the monsoon months during the future climate change, that could enhance the possibilities of high floods at most of the districts. Other impact could be lesser ground water recharge which upsets the farmers, as they completely rely on this source during the non-rainy season and in absence of enough precipitation (Randhawa *et al*., 2014). The increase in the possibility of non-rainy days and degree of evapotranspiration are under the projected climate change, negotiating the profit from maximum precipitation and enhance growing periods. Besides, expanding temperature would shorten the development time of crops and consequently an all-inclusive developing period alone couldn't be converted into a superior production. Analyzed data indicates that, almost all the districts have appropriate precipitation during the monsoon month (June - September) to compensate the evapotranspiration demand and *vice versa* from the rest other months (October - May). This, consequently, stresses the requirement for designing adaptation measures that capitalize on the opportunities which accompany with climate change while limiting its negative impacts.

4.2. *Best fit probability distribution*

This section would be focusing on the results documented for the best-fit probability distribution for all the 37 districts of the Bihar which is well recognized for the agriculture and main depended variables are precipitation and evapotranspiration. The precipitation and evapotranspiration data were assessed with twenty different probability distribution as described in the Table 1 to determine the suitable describing the data variation, the summary of the obtained parameters is presented in Annexure (A, B). Estimation of the probability distribution parameters of mean annual precipitation and potential evapotranspiration was established on the hypothesis that data were identically distributed and independent arbitrary events. These estimated parameters are earlier not available in the literature but presently obtained parameters along with the distribution could be used for any water related modelling and policy making with the agriculture crop water requirement point of view.

The best-fit distribution, such as Beta, Cauchy, Chi-Squared, Chi Squared (2P), Exponential, Exponential (2P), Gamma, Gamma (3P), Generalized Extreme Value (GEV), Log-Gamma (LG), Log-Pearson 3 (LP), Log-normal, Log-normal (3P), Normal, Pearson 5, Pearson 5 (3P), Pearson 6, Pearson 6 (4P), Weibull, Weibull (3P) were applied on the used data and the best-fit outcome was considered as the lowest goodness-of-fit result. The probability distribution for each district with the lowest ranking was considered as the best-fit probability distribution. In the present study the method "maximum likelihood" have been applied for ascertaining the best fit probability distribution functions as suggested by (Willems *et al*., 2007). Moreover, the best fit probability distribution for all 37 districtsin case of precipitation directed that no single distribution scored consistently best whereas in case of potential evapotranspiration it is (Cauchy distribution). Thus in case of potential evapotranspiration, second best were analyzed in the subsequent section.

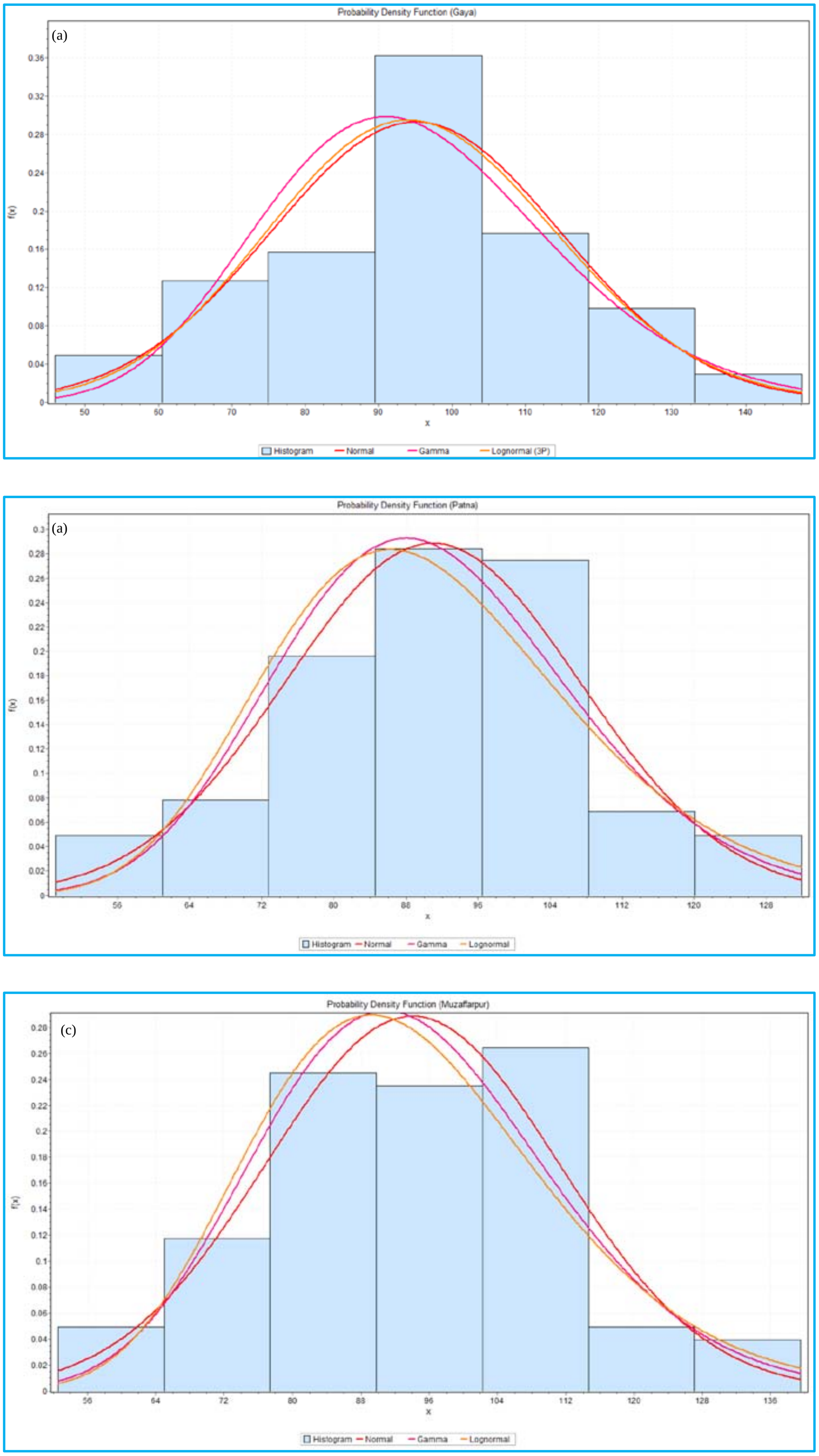
Using Kolmogorov-Smirnov (KS) goodness of test, it is observed that in precipitation Weibull distribution best fit for 10 districts (Aurangabad, Gaya, Gopalganj, Jahanbad, Madhepura, Nalanda, Purbi Champaran, Purnia, Shaikhpura, Siwan), Normal distribution best fit for 8 districts (Araria, Banka, Bhagalpur, Kaimur, Kishanganj, Patna, Rohtas, Supaul). Beta distribution, GEV distribution and Weibull (3P) distribution shows best fit in 4 districts each. The other distribution type such as Cauchy, Exponential, Exponential (2P), Gamma, Gamma (3P), Log-Gamma (LG), Pearson 5, Pearson 5 (3P), Pearson 6 and Pearson 6 (4P) were not suitable in this goodness of test. But the results were not same for the potential evapotranspiration, the Cauchy distribution is the best for all the districts the second best is Gamma distribution which covering the 23 districts and third best is GEV distribution covering the 14 districts (Table 3). Table 3 enlist the names of the districts characterizing the best fit distribution individually.

TABLE 3

Summary of best fit statistics results for Bihar districts

District	Precipitation (cm)				Potential Evapotranspiration*			
	By KS	By AD	By χ2	Best Fit (Rank)	By KS	By AD	By χ2	Best Fit (Rank)
Araria	Normal	Normal	Weibull	Normal (4)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Aurangabad	Weibull	Weibull (3P)	Cauchy	Weibull (7)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Banka	Normal	Normal	Weibull	Normal (4)	GEV	Pearson 5 (3P)	Pearson 5 (3P)	Pearson 5 (3P) (7)
Begusarai	Weibull (3P)	Weibull (3P)	Gamma	Weibull (3P) (7)	GEV	GEV	GEV	GEV (6)
Bhagalpur	Normal	GEV	Cauchy	GEV (7)	GEV	Pearson 6 (4P)	Pearson 6 (4P)	Pearson 6 (4P) (7)
Bhojpur	Beta	Beta	Gamma (3P)	Beta (4)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Buxar	Beta	Beta	Beta	Beta (3)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Darbhanga	GEV	Pearson 6 (4P)	Weibull	Gamma (3P) (10)	GEV	GEV	GEV	GEV (6)
Gaya	Weibull	Normal	Normal	Normal (4)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Gopalganj	Weibull	Weibull (3P)	Log Normal	Log Normal (13)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Jahanbad	Weibull	Weibull (3P)	Log Normal	Weibull (3P) (5)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Jamui	GEV	GEV	Pearson 6	GEV (5)	GEV	GEV	GEV	GEV (6)
Kaimur	Normal	Weibull (3P)	Chi Square	Normal (11)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Katihar	Weibull (3P)	Normal	Gamma (3P)	GEV (8)	Gamma (3P)	Pearson 5	Log Normal	Pearson 5 (9)
Khagariya	Weibull (3P)	Beta	Weibull (3P)	Weibull (3P) (5)	GEV	GEV	GEV	GEV (6)
Kishanganj	Normal	Normal	Pearson 5	Weibull (3P) (11)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Lakhisarai	Weibull (3P)	GEV	Chi Square (2P)	GEV (5)	GEV	GEV	GEV	GEV (6)
Madhepura	Weibull	Normal	Weibull	Normal (6)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Madhubani	Chi Square (2P)	Log Normal	Normal	Normal (10)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Munger	Beta	Beta	Beta	Beta (3)	GEV	GEV	GEV	GEV (6)
Muzaffarpur	Log Normal	Log Normal	Normal	Normal (5)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Nalanda	Weibull	Normal	Weibull (3P)	Weibull (3P) (7)	GEV	GEV	GEV	GEV (6)
Nawada	Gamma (3P)	Normal	Weibull	Normal (9)	GEV	GEV	GEV	GEV (6)
Paschim Champaran	GEV	Normal	Weibull	Weibull (3P) (7)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Patna	Normal	Normal	GEV	Normal (5)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
PurbiChamparan	Weibull	Normal	Log Normal	Normal (8)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Purnia	Weibull	Normal	Weibull	Weibull (7)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Rohtas	Normal	Weibull (3P)	Cauchy	Weibull (6)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Samastipur	Log Pearson	Normal	Weibull	Log Pearson (8)	GEV	GEV	GEV	GEV (6)
Saran	Beta	Beta	Normal	Beta (4)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Saharsha	GEV	Normal	Weibull	Weibull (3P) (13)	GEV	GEV	GEV	GEV (6)
Shaikhpura	Weibull	Normal	Gamma (3P)	Weibull (3P) (10)	GEV	GEV	GEV	GEV (6)
Sheohar	Beta	Beta	Cauchy	Beta (14)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Sitamadhi	Chi Square	Pearson 6 (4P)	Chi Square	Pearson 5 (3P) (10)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Siwan	Weibull	Weibull (3P)	Pearson6 (4P)	Weibull (3P) (8)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Supaul	Normal	Normal	Weibull	Normal (4)	Gamma (3P)	Gamma (3P)	Gamma (3P)	Gamma (3P) (6)
Vaishali	Log Pearson (3P)	Normal	Log Pearson (3P)	Log Pearson (3P) (7)	GEV	GEV	GEV	GEV (6)

* Due to Cauchy best distribution for all districts, second best is analyzed in case of potential evapotranspiration



Figs. 3(a-c). Probability Density Function of (a) Gaya district, (b) Pata district and (c) Muzaffarpur district

Similarly using the another goodness of fit Anderson-Darling (AD) Test, it is also observed that the Normal distribution fits well for 17 districts followed by Weibull distribution in 7 districts, Beta distribution in 6 districts and GEV in 3 districts in case of precipitation. Whereas for potential evapotranspiration, Gamma distribution fits well in 22 districts coving maximum districts (60%) followed by GEV (32%) and other distribution types. In case of Chi-Squared goodness of fit test, Weibull distribution well fits for 10 districts (27%) followed by Cauchy and Normal distribution together covering 8 districts (21%). There is as any specific distribution pattern and altogether different districts having distinct distribution characteristics for the precipitation. But in case of the potential evapotranspiration, it is showing the same behavior as shown in case of AD goodness test "Gamma distribution fits well in 22 districts coving maximum districts (60%) followed by GEV (32%) covering 12 districts". The other distribution type such as Exponential, Exponential (2P), Gamma, Log-Gamma (LG), Pearson 5 and Pearson 6 (4P) were not suitable in this goodness of test. It is suggested not to use the Exponential, Exponential (2P), Gamma, Log-Gamma (LG), Pearson 5 and Pearson 6 (4P) probability distribution for any of the hydraulic design and flood estimation studies.

It is practically not possible to judge which probability distribution to select as a best fit to the precipitation and potential evapotranspiration data based merely by evaluating the individual goodness of fit test. Hence in order to finalize the best distribution, the smaller the combined rank of all three goodness of fit, the better the choice of the probability distribution would be considered. Sharma and Singh (2010) studied the best suited statically distribution fit over the precipitation data by the same procedure as described above and compare with different model techniques. The overall ranks of the 37 districts combined shows that Normal distribution fit well the precipitation in the 11 districts covering 30% of the Bihar followed by the Weibull (3P) in 9 districts which covered rest 24% and the third best is Beta distribution which covers the 5 districts accounting 13.5% of the total districts. Other best fit distribution for the precipitation are Gen. Extreme Value, Weibull, Log-Pearson 3, Gamma (3P), Lognormal (3P) and Pearson 5 (3P). Table 3 shows the summary of the score with the district name. The 3P Pearson probability distributions, which usually considered for understanding the procedure of the skewness in the assessment of future precipitation, were found to be the best-fit for Sitamadi district only. From the above results, it can be assumed that the precipitation distribution in Bihar are positively and negatively skewed and the Normal distribution and Weibull (3P) distributions can be suitably apply for the forecast of future precipitation in the climate change and region having the shortage of water resources.

Similarly, by integrating the best three probability distributions together and expect Cauchy distribution which is best for all districts, Gamma (3P) ranks the first covering 22 districts which covers 60% of the total districts followed by the GEV distribution in 12 districts for the potential evapotranspiration. Interesting here to notice that most of the districts is well fitted by relatively simple probability distributions with single shape parameter which rules their asymptotic behavior. In case of the evapotranspiration where maximum districts get fitted by more than one parameter would fit the ends of the evapotranspiration series better. Suhaila and Jemain (2007) also reported that in Malaysia, Gamma and Exponential distribution with other parameters performed better rather than using the single parameters distributions. Fig. 3 shows the probability density function plot of few selected districts with the normal, Gamma and Log Norml (3P) distributions. Because of the space constraint, different figures of P-P plots, probability difference plots and probability density function plots for different stations are excluded herein.

In summary, the normal distribution was seen as the best-fit probability distribution for 11 districts followed by Weibull (3P) for 9 districts, the Beta distribution for 5 districts and other distribution for rest districts for precipitation. Whereas Cauchy distribution was found to be the best-fit probability distribution for potential evapotranspiration for all districts and the second best was Gamma (3P) covering almost 60% of the total districts followed by General Extreme Value distribution (32%). In general, the Pearson distribution is extensively used in the modeling of maximum hydrology and water resources distributions of precipitation and potential evapotranspiration. Table 3 displays the various likelihood distribution ranks towards the different probability distribution and as per the project requirement with the risk and uncertainty acceptance we may choose the best fit distribution. This study is of its first kind in the agriculture dominating and most populous state therefore it allows us to improve future studies related to water resources and agriculture that might guard economy loss and food security. One another argument which is that Exponential, Exponential (2P), Gamma, Log-Gamma, Log-Pearson 3P and Pearson 5 could be avoided for the best fit distribution which indicates constraint of one parameter distribution type in extreme precipitation analysis. The results may be recommended for future forecast and developing the suitable modelling approach for precipitation studies by adding more distribution and joint distribution studies.

5. Conclusions

The two key variables of meteorological are precipitation and evapotranspiration and their temporal and spatial distribution varies across the gauging locations. The variation could be because of the climatic factor, geographical settings and other environmental challenges which is difficult to understand. Climate change issues is most significant concerns for the scientific community involved in the water resource and agriculture planning. In order to minimize the uncertainty in the using the past records and analyzing an effort has been made with well-organized procedure to understand the best fit distribution of the precipitation and potential evapotranspiration of 37 districts of the Bihar where agriculture is lifeline for many poor people.

The 20 different possible probability distributions such as Beta, Cauchy, Chi-Squared, Chi Squared (2P), Exponential, Exponential (2P), Gamma, Gamma (3P), Generalized Extreme Value (GEV), Log-Gamma (LG), Log-Pearson 3 (LP), Log-normal, Log-normal (3P), Normal, Pearson 5, Pearson 5 (3P), Pearson 6, Pearson 6 (4P), Weibull, Weibull (3P) were applied on precipitation and potential evapotranspiration data of 102 years. The normal distribution was seen as the best-fit probability distribution for the 11 districts followed by Weibull (3P) for 9 districts, the Beta distribution for 5 districts and other distribution for rest other districts for precipitation. Whereas Cauchy distribution was found to be the best-fit probability distribution for potential evapotranspiration for all districts and the second best was Gamma (3P) covering almost 60% of the total districts followed by General Extreme Value distribution (32%). Furthermore, the choice of a suitable probability distribution is the foremost problems in any other engineering practice. There is still requirement for additional studies using other probability distribution type or joint distribution studies to confirm its wide applicability in Bihar.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgment

The author would like to thank anonymous reviewer(s) for valuable comments and positive suggestions to improve the quality of this paper.

Disclaimer : The contents and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Alam, M. A., Emura, K., Farnham, C. and Yuan, J., 2018, "Best-fit probability distributions and return periods for maximum monthly rainfall in Bangladesh", *Climate*, **6**, 1, 9.

Amin, M. T., Rizwan, M. and 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.

Bhakar, S. R., Iqbal, M., Devanda, M., Chhajed, N. and Bansal, A. K., 2008, "Probability analysis of rainfall at Kota", *Indian Journal of Agricultural Research*, **42**, 3, 201-206.

Blum, Annalise G., Stacey A. Archfield and Richard, M. Vogel., 2017,"On the probability distribution of daily streamflow in the United States", *Hydrology and Earth System Sciences*, **21**, 6, 3093-3103.

Bruhn, J. A., Fry, W. E. and Fick, G. W., 1980, "Simulation of daily weather data using theoretical probability distributions", *Journal of Applied Meteorology*, **19**, 9, 1029-1036.

Francesco, De Paola and Giugni, Maurizio, 2013, "Coupled spatial distribution of rainfall and temperature in USA", *Procedia Environmental Sciences*, **19**, 178-187.

Kumar, V. and Sen, S., 2020, "Assessment of spring potential for sustainable agriculture: A case study in lesser Himalayas", *Applied Engineering in Agriculture*, **36**, 1, 11-24.

Kumar, V., Jain, S. K. and Singh, Y., 2010, "Analysis of long-term rainfall trends in India", *Hydrol. Sci. J.*, **55**, 4, 484-496.

Kumar, V., Shanuand Jahangeer, 2017, "Statistical distribution of rainfall in Uttarakhand, India", *Appl. Water Sci.*, **7**, 4765-4776.

Nanda, A., Sen, S., Jirwan, V., Sharma, A. and Kumar, V., 2018, "Understanding plot-scale hydrology of Lesser Himalayan watershed - A field study and HYDRUS-2D modelling approach", *Hydrological Processes*, **32**, 9, 1254-1266.

Olofintoye, O. O., Sule, B. F. and Salami, A. W., 2009, "Best-fit Probability distribution model for peak daily rainfall of selected Cities in Nigeria", *New York Science Journal*, **2**, 3, 1-12.

Randhawa, R. K., Singh, H. and Kang, M. S., 2014, "Global warming and its possible impact on agriculture in India", *Adv. Agron*., **123**, 65-121.

Sanjib, G., Roy, S. and Biswas, S. C., 2016, "Determination of the best fit probability distribution for monthly rainfall data in Bangladesh", *Am. J. Math. Stat*, **6**, 170-174.

Sharma, M. A. and Singh, J. B., 2010, "Use of Probability Distribution in rainfall Analysis", *New York science Journal*, 40-49.

Storch, H. and Zweirs, F. W., 1999, "Statistical Analysis in Climate Change Research", *Cambridge University Press*, Cambridge

Subudhi, R., 2007, "Probability analysis for prediction of annual maximum daily rainfall of hakapada block of Kandhamal district in Orissa", *Indian J. Soil Conser*., **35**, 84-85.

Suhaila, J. and Jemain, A. A., 2007, "Fitting daily rainfall amount in Peninsular Malaysia using several types of exponential distributions", *J. Applied. Sci. Res*., **3**, 1027-1036.

Tesfaye, K., Aggarwal, P. K., Mequanint, F., Shirsath, P. B., Stirling, C. M., Khatri-Chhetri, A. andRahut, D. B., 2017, "Climate variability and change in Bihar, India: challenges and opportunities for sustainable crop production", *Sustainability*, **9**, 11.

Wilheit, Thomas T., Alfred, T. C., Chang and Long, S. Chiu, 1991, "Retrieval of monthly rainfall indices from microwave radiometric measurements using probability distribution functions", *Journal of Atmospheric and Oceanic Technology*, **8**, 1, 118-136.

Willems, P., Guillou, A. and Beirlant, J., 2007, "Bias correction in hydrologic GPD based extreme value analysis by means of a slowly varying function", *J. Hydrol*., **338**, 3-4, 221-236.

Yue, Sheng and Michio, Hashino, 2007, "Probability distribution of annual, seasonal and monthly precipitation in Japan", *Hydrological sciences journal*", **52**, 5, 863-877.

Zakwan, M. and Ara, Z., 2019, "Statistical analysis of rainfall in Bihar",*Sustainable Water Resources Management*, **5**, 4, 1781-1789.

Zhang, Jinping, Yong, Zhao and Zhihong, Ding, 2017, "Research on the joint probability distribution of rainfall and reference crop evapotranspiration", *Paddy and water environment*, **15**, 1, 193-200.

Zhanling Li, Zhanjie, Li, Wei, Zhao and Yuehua, Wang, 2014, "Probability Modeling of Precipitation Extremes over Two River Basins in Northwest of China", *Advances in Meteorology*, 2015, Article ID 374127, 13pages.