International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; 8(4): 111-120

https://www.mathsjournal.com
Received: 20-05-2023
Accepted: 22-06-2023

P Durga Bhavani
Department of Statistics & Computer Applications, ANGRAU, Agricultural College, Bapatla, Andhra Pradesh, India

D Ramesh
Department of Statistics & Computer Applications, ANGRAU, Agricultural College, Bapatla, Andhra Pradesh, India

V Srinivasa Rao
Department of Statistics & Computer Applications, ANGRAU, Agricultural College, Bapatla, Andhra Pradesh, India

M Sunil Kumar
Department of Statistics & Computer Applications, ANGRAU, Agricultural College, Bapatla, Andhra Pradesh, India

Corresponding Author:
P Durga Bhavani
Department of Statistics & Computer Applications, ANGRAU, Agricultural College, Bapatla, Andhra Pradesh, India

Statistical analysis of rainfall in Kadapa District of Andhra Pradesh

P Durga Bhavani, D Ramesh, V Srinivasa Rao and M Sunil Kumar

Abstract
Rainfall analysis is one of the important components in hydrological processes for either using it as a random chance constrained input or for taking a risk at certain level for crop planning. Present study made an attempt to understand the rainfall scenario of Kadapa district, through use of descriptive statistics, trend analysis [Mann-Kendall (or Modified Mann-Kendall) test, Sen's-slope method] and distribution of rainfall for monthly, seasonal and annual rainfall of during the period from 1985 to 2021. It was revealed that the October (116.26 mm) which ranges 11.31 mm to 394.74 mm, while the least average was observed in January (2.36 mm), which indicated that the months of January and October were the extreme precipitation months of the district. From the Mann-Kendall and Sen's slope estimator, it was resulted that the probability value was found to be significant with increasing trend for June (1.24), September (2.14), South-West Monsoon (5.96) and Annual (11.68) rainfall series. Based on Goodness of fit criterion, Log-logistic (3P) and Pearson 5 (4P) was identified as appropriate distribution for South-West monsoon of Kadapa district and it was found that through Log-logistic (3P) distribution, probability of getting rainfall more than 100 mm, 200 mm, 300 mm, 400 mm and 500 mm were identified as 97.6%, 86.5%, 58.2%, 27.2% and 10.6% respectively.

Keywords: Rainfall, trend, slope, distribution

Introduction

Water is a vital natural resource necessary for survival. In many parts of the world, including Andhra Pradesh, rainfall serves as the primary source of water for agricultural production. Rainfall is uneven, erratic and inconsistent, showing great variation both regionally and temporally. It is one of the components in hydrological processes for either using it as a random chance constrained input or for taking a risk at certain level for crop planning. Growth of agriculture and related sectors depend on timely onset of monsoon in adequate amount.

Kadapa district is located in the Rayalaseema region of Andhra Pradesh. The district lies between the coordinates 13° 43' to 15° 14'N latitude and 78° 40' to 79° 55'E longitude. While agriculture remains the most important economic activity of the district, it has also been affected by high levels of instability and uncertainty. Being located in the rain-shadow region of Andhra Pradesh, the district comes under southern zone which ranges between 600-1000 mm. Paddy, Groundnut, Red gram, Cotton, Bengal gram are the major Agricultural crops grown in the district. Rainfed agriculture plays a significant role in Kadapa district, with approximately 90% of agricultural activities relying on natural rainfall for irrigation.

Trend analysis is one of the important statistical techniques used to examine and identify patterns in a rainfall series over a specific period of time. It involves analysing the data to identify whether there is a consistent upward or downward movement, a cyclical pattern, or any other systematic changes over time (Reddy *et al.*, 2022) [8]. Trend analysis is widely used in various fields, including finance, economics, marketing and environmental studies, to understand and forecast future behaviour based on historical data.

The distribution of rainfall, rather than its volume, also plays a crucial role in influencing crop yield in any region. Probability and frequency analysis of rainfall data help to determine the expected rainfall at different probability levels. The variability of rainfall can impact the frequency of floods or instances of drought and climate change studies focus on potential changes in climatic parameters like rainfall and temperature.

By considering the information, present study had been formulated to study the trend and distribution of rainfall in Kadapa district of Andhra Pradesh.

Materials and Methods

In the present study, data pertaining to Monthly, Seasonal and Annual rainfall during the study period (1985-2021) had been utilized for the Kadapa district of Andhra Pradesh. For this, Secondary time series data on daily rainfall during the study period was collected from the Directorate of Economics and Statistics - Government of Andhra Pradesh and the Andhra Pradesh State Development Planning Society (APSDPS, 2022).

Trend analysis

In the present study, non-parametric test namely Mann-Kendall test (Under the assumption i.e., data are independent and randomly ordered) was employed to understand the trends of monthly, seasonal and annual rainfall in the Kadapa district during the study period. So, initially, randomness of each data series was verified by Wallis and Moore phase-frequency test (Wallis and Moore, 1941) [43]. If the randomness of data series was found to be violated, then Modified Mann-Kendall test was tried instead of Mann-Kendall test (Singh *et al.*, 2021) [11]

Mann–Kendall Test

To determine the presence of statistically significant trend in hydrologic climatic variables such as temperature, precipitation and stream flow with reference to climate change, non-parametric Mann-Kendall: M-K test (Mann, 1945; Kendall, 1975) [5] has been employed by a number of researchers as due to certain advantages of it: (i) the data do not need to conform to a particular distribution, thus extreme values are acceptable (ii) missing values are also allowed to be included in the dataset (iii) the test has low sensitivity to abrupt breaks due to heterogeneous time series and (iv) finally, in time series analysis, it is not necessary to specify whether the trend is linear or not.

The M-K test is applicable in cases when the data values x_i of a time series can be assumed to obey the model,

$$x_i = f(t_i) + \varepsilon_i$$

Where, $f(t)$ is a continuous monotonic increasing or decreasing function of time and the residuals ε_i can be assumed to be from the same distribution with zero mean and constant variance. According to this test, the null hypothesis H_0 assumes that there is no trend *i.e.* the observations x_i come from a population where the random variables are independent and identically distributed. The alternative hypothesis H_1 is that the data follow an increasing or decreasing monotonic trend over time.

The M-K statistic *(S)* is computed as follows: $S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i)$

Where

$$\text{Sign}\,(x_j - x_i) = \begin{cases} 1 & \text{if } x_j - x_i > 0 \\ 0 & \text{if } x_j - x_i = 0 \\ -1 & \text{if } x_j - x_i < 0 \end{cases}$$

Where x_i and x_j are the data values at time *j* and *i*, *j* > *i* respectively. If a data value from a later time period is higher (lower) than a data value from an earlier time period, the statistic S is incremented (decremented) by 1. The net result of all such increments and decrements yields the final value of S. The exact distribution of S for n <10 was derived by both Mann (1945) [5] and Kendall (1975). For n ≥10, the statistic S is approximately normally distributed with the mean and variance as follows:

$$E[S] = 0 \quad VAR(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p-1}^{q} t_p(t_p - 1)(2t_p + 5)\right]$$

Where *q* is the number of tied (zero difference between compared values) groups and t_p is the number of ties in the p^th^ group.

The standard test statistic Z is computed as follows and is approximately normally distributed. The presence of a statistically significant trend is evaluated using the Z value. A positive (negative) value of Z indicates an upward (downward) trend. If the computed value of Z> $Z_{\alpha/2}$, the null hypothesis H_O is rejected at α level of significance in a two-sided test.

$$Z = \begin{cases} \frac{S-1}{\sqrt{VAR(S)}} & \text{if } S>0 \\ 0 & \text{if } S=0 \\ \frac{S+1}{\sqrt{VAR(S)}} & \text{if } S<0 \end{cases}$$

Sen's slope estimator

In this study, the magnitude of trend in the time series was determined by using a non-parametric method known as Theil-Sen estimator also known as Sen's slope estimator (Sen, 1968). Sen's method assumes a linear trend *f (t)* in the time series and has been widely used for determining the magnitude of trend in hydro-meteorological time series

$$f(t) = Qt + B$$

Where *Q* is the slope, *B* is a constant and *t* is time. To get the slope estimate Q, the slopes of all the data value pairs is calculated using the equation:

$$Q_i = \frac{x_j - x_k}{j-k} \quad i=1,2,\ldots N$$

Where x_j and x_k are the data values at time j and k (*j*>*k*) respectively. If there are n values x_j in the time series, there will be as many as $N = \frac{n(n-1)}{2}$ slope estimates Q_i are obtained. The median of these N values of Q_i is the Sen's estimator of slope (Q), which is calculated as A positive value of Q indicates an upward (increasing) trend and a negative value indicates a downward (decreasing) trend in the time series.

$$Q = \begin{cases} Q_{\frac{(N+1)}{2}} & \text{if } N \text{ is odd} \\ \frac{1}{2}\left(Q_{\frac{N}{2}} + Q_{\frac{(N+2)}{2}}\right) & \text{if } N \text{ is even} \end{cases}$$

Modified Mann-Kendall (MM-K) test

Even though M-K test is most commonly used test for detecting trend in rainfall data, it assumes that sample data should be serially independent. However, it is well known that from many previous studies, most of rainfall time-series data exhibit serial correlation. The presence of serial correlation in time-series will alter the variance of the M-K test statistic which in turn will affect the ability of the test to assess the significance of the trend correctly (Hamed and Rao, 1998). The presence of positive autocorrelation in the data increases the probability of detecting trend even though actual data have no trend, and vice versa. Yue and Wang (2004) developed Modified Mann-Kendall (MM-K) test, which eliminates the effect of serial correlation present in the time-series data on the M-K test statistic by correcting the variance using Effective Sample Size (ESS). The accuracy of the modified test in terms of its empirical significance level was found to be superior to that of the original Mann-Kendall trend test without any loss of power.

Therefore, the modified variance $V^*(S)$ using ESS is given by:

$$V^*(S) = V(S).\frac{n}{n^*}$$

Where n is the Actual Sample Size (ASS) of data, n/n^* is termed the Correction Factor (C.F) and n^* is the ESS, proposed by Lettenmaier (1976) computed by:

$$n^* = \frac{n}{1 + 2.\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right).\rho_k}$$

Where ρ_k is the lag-k serial correlation coefficient, which can be estimated by the sample lag-k serial correlation coefficient (r_k) given by:

$$r_k = \frac{\frac{1}{n-k}\sum_{t=1}^{n-k}(x_t - \bar{x}_t)(x_{t+k} - \bar{x}_t)}{\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x}_t)^2}$$

$$\bar{x}_t = \frac{1}{n}\sum_{t=1}^{n} x_t$$

Next variance of M-K test is replaced by modified variance and proceeds with the M-K test procedure.

Fitting probability distributions to rainfall data

In the study, different probability distributions *viz.* Exponential, Exponential(2P), Ferchet, Ferchet (3P), Gamma, Gamma (3P), GEV, Gumbel Max, Gumbel Min, Log-Logistic, Log-Logistic (3P), Log Pearson 3, Lognormal, Lognormal (3P), Normal, Pareto, Perason 5, Perason 5(3P), Pearson 6, Pearson 6 (4P), Weibull and Weibull (3P) were used to evaluate the best fit probability distribution for monthly, seasonal and annual rainfall series of the district. Probability density function of the selected distributions were depicted Table-1.

Description of Parameter

- **Shape parameter**

A shape parameter is any parameter of a probability distribution that is neither a location parameter nor a scale parameter (nor a function of either or both of these only, such as a rate parameter). Shape parameters allow a distribution to take on a variety of shapes, depending on the value of the shape parameter. These distributions are particularly useful in modelling applications since they are flexible enough to model a variety of data sets. Examples of shape parameters are skewness and kurtosis.

- **Scale parameter**

In probability theory and statistics, a scale parameter is a special kind of numerical parameter of a parametric family of probability distributions. The larger the scale parameter, the more spread out the distribution. The scale parameter of a distribution determines the scale of the distribution function. The scale is either estimated from the data or specified based on historical process knowledge. In general, a scale parameter stretches or squeezes a graph. The examples of scale parameters include variance and standard deviation.

- **Location parameter**

The location parameter determines the position of central tendency of the distribution along the x-axis. The location is either estimated from the data or specified based on historical process knowledge. A location family is a set of probability distributions where μ is the location parameter. The location parameter defines the shift of the data. A positive location value shifts the distribution to the right, while a negative location value shifts the data distribution to the left. Examples of location parameters include the mean, median and mode.

Table 1: Description of continuous probability distributions

Distribution	Probability density function	Range	Parameters
Exponential	$f(x) = \lambda exp^{(-\lambda x)}$	$0 \leq x < +\infty$ $\lambda > 0$	$\lambda = inverse\ scale\ parameter$
Exponential (2P)	$f(x) = \lambda exp^{[-\lambda(x-\Upsilon)]}$	$0 \leq x < +\infty$ $\lambda > 0$	$\lambda = inverse\ scale\ parameter$ Υ = location parameter
Ferchet	$f(\text{x};\ \alpha, \text{s}, \text{m}) = \frac{\propto}{s}\left(\frac{x-m}{s}\right)^{-1-\propto} e^{-\left(\frac{x-m}{s}\right)^{-\propto}}$	$\propto > 0$ $s > 0$ $-\infty < m < +\infty$	$\propto = shape\ parameter$ $s = scale\ parameter$ $m = location\ parameter$
Gamma (1P)	$\text{f(x)} = \frac{1}{\tau(k)} x^{k-1} \exp(-x)$	$0 < x < \infty$ $k > 0$	$k = shape\ parameter$ $\beta = scale\ parameter$ $\gamma = location\ parameter$ $\tau\ is\ the\ gamma\ function$
Gamma (3P)	$\text{f(x)} = \frac{(x-\gamma)^{k-1}}{\tau(k)\beta^k} x^{k-1} \exp(-\frac{(x-\gamma)}{\beta})$	$k > 0, \beta > 0$ $\gamma > 0,$ $\gamma \leq x \leq \pm\infty$	
Generalized extreme value (GEV)	$\text{f(x)}=\begin{cases}\frac{1}{\beta} exp\left[-(1+kz)^{\frac{-1}{k}}\right](1+kz)^{\frac{-1}{k}} & k \neq 0 \\ \frac{1}{\beta} exp[-z - \exp(-z)] & k = 0\end{cases}$	$1+k\,z > 0$ for $k \neq 0$ $-\infty < x < +\infty$ for $k = 0$ where $z = \frac{(x-\mu)}{\beta}$	$k = shape\ parameter$ $\beta = scale\ parameter$ $\mu = location\ parameter$

Table 2: (Cont.) Description of continuous probability distributions

Distribution	Probability density function	Range	Parameters
Normal	f(x) $= \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$	$-\infty < x < +\infty$ $-\infty < \mu < +\infty$ $\sigma > 0$	$\mu = mean$ $\sigma = standard\ deviation$
Log – normal	f(x) $= \frac{1}{x\,\sigma\sqrt{2\pi}}\exp\left[-\left(\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)\right]$	$\gamma < x < +\infty$ $\sigma > 0, \mu > 0,\ \gamma = 0$	$\mu = shape\ parameter$ $\sigma = scale\ parameter$ $\gamma = location\ parameter$, Yields two parameter lognormal distribution.
Lognormal (3P)	f(x) $= \frac{exp\left[-\frac{1}{2}\left(\frac{\ln x-\gamma}{\sigma}\right)^2\right]}{(x-\gamma)\sigma\sqrt{2\pi}}$		
Pareto	f(x) $= \frac{k\beta^k}{\beta^{k+1}}$	$\beta > 0,\ k > 0$	$\beta = scale\ parameter,$ $k = shape\ parameter$
Gumbel	f(x) $= \frac{1}{\beta}\exp - (z + e^{-z})$ where, $z = \frac{x-\mu}{\beta}$	$\beta > 0$ $-\infty < x < +\infty$	$\beta = scale\ parameter$ $\mu = location\ parameter$
Log-logistic(3P)	f(x) $= \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}\left(1 + \left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{-2}$	$x > 0, \beta > 0,\ \alpha > 0$	$\alpha = scale\ parameter$ $\beta = shape\ parameter,$ $\gamma = location\ parameter$

Table 3: (Cont.) Description of continuous probability distributions

Distribution	Probability density function	Range	Parameters
Pearson 5 (3P)	f(x) $= \frac{exp\left(\frac{-\beta}{(x-\gamma)}\right)}{\beta\tau(\alpha)((x-\gamma)/\beta)^{\alpha+1}}$	$\gamma < x < +\infty$ $\alpha > 0, \beta > 0,\ \gamma = 0$	$\alpha = shape\ parameter$ $\beta = scale\ parameter$ $\gamma = location\ parameter$
Pearson 6 (3P)	f(x) $= \frac{(x/\beta)^{\alpha_1-1}}{\beta B(\alpha_1,\alpha_2)(1+x/\beta)^{\alpha_1+\alpha_2}}$	$\gamma \le x < +\infty$ $\alpha_1 > 0,$ $\alpha_2 > 0,$ $\beta > 0, \gamma = 0$	$\alpha_1 = shape\ parameter$ $\alpha_2 = shape\ parameter$ $\beta = scale\ parameter$ $\gamma = location\ parameter$
Pearson 6 (4P)	f(x) $= \frac{((x-\gamma)/\beta)^{\alpha_1-1}}{\beta B(\alpha_1,\alpha_2)(1+(x-\gamma)/\beta)^{\alpha_1+\alpha_2}}$		
Log-Pearson 3 (3P)	f(x) $= \frac{1}{x\lvert\beta\rvert\tau(\alpha)}\left(\frac{\ln x-\gamma}{\beta}\right)^{\alpha-1} exp\left(-\frac{\ln x-\gamma}{\beta}\right)$	$0 \le x \le e^{\gamma},\ \beta < 0$ $e^{\gamma} \le x \le +\infty,\ \beta > 0$	$\alpha = shape\ parameter$ $\beta = scale\ parameter,$ $\gamma = location\ parameter$
Weibull (1P)	f(x) = k $x^{k-1}\exp(-x^k)$	$x > 0, \beta > 0$	$k = shape\ parameter$ $\beta = scale\ parameter$ $\mu = location\ parameter$
Weibull (3P)	f(x) $= \frac{k}{\beta}\left(\frac{x-\mu}{\beta}\right)^{k-1}\exp - \left(\frac{x-\mu}{\beta}\right)^k$	$0 \le x < +\infty$ $k > 0, \beta > 0, \gamma = 0$	

To identify the best distribution among other distributions to the particular rainfall series, various goodness of fit criterion were utilized, as described below.

Goodness-of-fit assessment

The goodness of fit test measures the discrepancy between observed values and the expected values. In the study, Kolmogorov-Smirnov test, Anderson -Darling test and Chi-Square test were selected as goodness of fit measures.

The null and alternative hypotheses of these tests are

H_0: the data follow the specified distribution;

H_1: the data do not follow the specified distribution.

Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test is used to decide if a sample comes from a population with a specific distribution. The Kolmogorov-Smirnov (K-S) test is based on the empirical distribution function (ECDF). Given N ordered data points $Y_1, Y_2 ... Y_N$, the ECDF is defined as

$$E_N = \frac{n(i)}{N}$$

Where n(i) is the number of points less than Y_i and the Y_i are ordered from smallest to largest value. This is a step function that increases by 1/N at the value of each ordered data point.

Test Statistic: The Kolmogorov-Smirnov test statistic is defined as

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

Where F is the theoretical cumulative distribution of the distribution being tested which must be a continuous distribution (i.e., no discrete distributions such as the binomial or Poisson) and it must be fully specified (i.e., the location, scale, and shape parameters cannot be estimated from the data). The hypothesis regarding the distributional form is rejected if the test statistic, D, is greater than the critical value obtained from a table (Ghosh *et al*, 2016) [1].

Anderson –Darling Test

The Anderson-Darling test (Stephens, 1974) is used to test if a sample of data comes from a population with a specific distribution. It is a modification of the Kolmogorov-Smirnov (K-S) test and gives more weight to the tails than does the K-S test. The K-S test is distribution free in the sense that the critical values do not depend on the specific distribution being tested. The Anderson-Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution.

The Anderson-Darling test statistic is defined as

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

F is the cumulative distribution function (CDF) of the specified distribution. Note that the Yi are the ordered data. The critical values for the Anderson-Darling test are dependent on the specific distribution that is being tested. The test is a one-sided test and the hypothesis that the distribution is of a specific form is rejected if the test statistic is greater than the critical value.

Chi-Square Test: The Chi-square test assumes that the number of observations is large enough so that the chi-square distribution provides a good approximation as the distribution of test statistic.

The Chi-squared statistic is defined as

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where

O_i = observed frequency

E_i = expected frequency

'i'= number observations (1, 2, ……k)

$E_i = F(X_2) - F(X_1)$

F = the CDF of the probability distribution being tested.

The observed number of observation (k) in interval 'i' is computed from $K = 1+\log_2 n$; here n = sample size. This equation is for continuous sample data only and is used to determine if a sample comes from a population with a specific distribution. The hypothesis regarding the distributional form is rejected at the chosen significance level (α) if the test statistic is greater than the critical value defined as; $\chi^2_{1-\alpha,k-1}$ meaning the Chi-Squared inverse CDF with k-1 degrees of freedom and a significance level of α.

Identification of best fitted distribution

As individual rank is associated for each of GOF tests separately, hence it would be difficult to identify the best fitted distribution of data series, based on all three GOF tests. Hence, an approach of scoring has been adopted to find out the best fitted model for each data series. According to this method, among the selected distributions (For e.g., 18 candidate distributions), a highest score of 18 will be given to the one which ranks first and next score (i.e., 17) is awarded to the distribution having rank more than 1 (i.e., 2) is given to the distribution, likewise. A lowest score of 1 is provided to the distribution which ranks 18 and 0 is given when a distribution fails to fit the data. By this way, score will be given to all the distributions for each of the GOF tests ranking separately and the final score is obtained by adding these three scores. A distribution, which have maximum total score from three GOF tests will be considered as the best fitted distribution to the data series.

Results and Discussion

Initially various selected descriptive measures were applied to the monthly, seasonal and annual rainfall of the Kadapa district during the period from 1985 to 2021, as to know the basic behavior of rainfall. The selected measures were namely Mean, Standard Deviation (SD), Coefficient of Variation (CV%), Minimum, Maximum, Skewness and Kurtosis.

Table 4: Descriptive statistics for monthly, seasonal and annual rainfall (mm) of Kadapa district

Monthly	Mean	SD	CV	Skewness	Kurtosis	Maximum	Minimum
January	2.36	5.69	241.55	4.21	20.30	31.68	0.00
February	2.83	7.23	254.96	3.27	10.25	31.23	0.00
March	6.11	11.25	184.07	2.83	7.74	46.76	0.00
April	12.99	15.17	116.73	1.91	4.20	63.91	0.00
May	31.50	28.10	89.20	1.75	4.06	132.81	0.00
June	56.42	49.79	88.25	1.90	3.62	223.99	11.39
July	78.54	51.91	66.09	0.74	-0.51	193.59	9.37
August	95.45	51.33	53.78	0.62	0.68	244.43	9.79
September	105.18	61.70	58.66	0.91	1.31	302.72	15.46
October	116.26	82.76	71.19	1.39	2.49	394.74	11.31
November	82.66	86.12	104.20	1.98	4.13	386.02	3.35
December	21.72	22.33	102.81	1.07	0.24	79.50	0.00
Seasonal							
South-West	335.60	134.04	39.94	0.51	0.38	676.97	75.72
North-East	220.64	121.48	55.06	0.77	0.18	544.09	45.76
Winter	5.19	9.29	179.02	2.63	6.59	39.89	0.00
Annual	612.02	231.45	37.82	0.34	0.01	1145.03	167.29

From Table 2, it was revealed that the highest average rainfall was reported in October (116.26 mm) which ranges 11.31 mm to 394.74 mm, while the least average was observed in January (2.36 mm), which indicated that the months of January and October were the extreme precipitation months of Kadapa District of Andhra Pradesh during the study period (1985-2021). Based on CV%, highest value was observed for February (254.96%), which might be due to heavy irregularities during the period and where the least was observed for August (53.78%). For the same table-4.29, rainfall during the months (Jan-Dec), monsoons and annual period were positively skewed. Lepto kurtic (>3) values were observed during January, February, March, Winter Season, April, November, May and june only, which indicated that those data comprised of extreme outliers. The average annual rainfall of Kadapa district was 612.02 mm.

Trend analysis

In the present study, non-parametric test namely Mann-Kendall test (Under the assumption i.e., data are independent and randomly ordered) was employed to understand the trends of monthly, seasonal and annual rainfall in the Kadapa district of Andhra Pradesh during the study period (1985-2021). So, initially, randomness of each data series was verified by Wallis and Moore phase-frequency test. If the randomness of data series was found to be violated, then Modified Mann-Kendall test was tried instead of Mann-Kendall test.

From Table-3, it was revealed through Wallis and Moore phase-frequency test that the monthly, seasonal and annual rainfall were found to be random in nature, as p-values were greater than 5% level of significance. Hence, Mann-Kendall and Sen's slope estimators were applied only for the random series and it was obtained a significant with increasing trend for June (1.24), September (2.14), South-West Monsoon (5.96) and Annual (11.68) rainfall series. As a consequence, the total rainfall during these significant periods would expect to increase to some extent in the district. Similar kind of report was obtained by Rana *et al.* (2019) [7] that there was significant increasing trend for June and September month.

Table 5: Trend analysis for monthly, seasonal and annual rainfall of Kadapa district

Kadapa district	Wallis and Moore phase frequency test		Mann-Kendall test		Modified Mann-Kendall test		Sen's Slope estimator
Monthly	**Z-statistic**	**P- value**	**Z-statistic**	**P- value**	**Z-statistic**	**P- value**	**Slope (mm/year)**
January	1.73	0.08	1.42	0.16	-	-	0.00
February	0.27	0.79	1.62	0.11	-	-	0.00
March	1.07	0.29	1.25	0.21	-	-	0.03
April	0.93	0.35	1.49	0.14	-	-	0.24
May	0.53	0.59	1.18	0.24	-	-	0.40
June	1.33	0.18	2.97	0.00	-	-	1.24
July	0.93	0.35	1.24	0.21	-	-	0.96
August	0.67	0.51	1.82	0.07	-	-	1.55
September	0.13	0.89	2.42	0.02	-	-	2.14
October	1.87	0.06	0.82	0.41	-	-	0.94
November	0.93	0.35	0.98	0.33	-	-	0.45
December	0.27	0.79	1.20	0.23	-	-	0.30
Seasonal							
South-West	0.67	0.51	2.89	0.00	-	-	5.96
North-East	1.47	0.14	1.74	0.08	-	-	3.51
Winter	0.13	0.89	1.70	0.09	-	-	0.06
ANNUAL	0.27	0.79	3.02	0.00	-	-	11.68

Distribution fitting

In the present study, monthly (June, July, August, September), seasonal (south west and north east monsoon) and annual data of rainfall related to the Kadapa district of Andhra Pradesh were tried to fit for different distributions such as Exponential, Exponential(2P), Ferchet, Ferchet (3P), Gamma, Gamma (3P), Generalized Extreme Value, Gumbel Max, Gumbel Min, Log-Logistic, Log-Logistic (3P), Log Pearson 3, Lognormal, Lognormal (3P), Normal, Pareto, Perason5, Perason5(3P), Pearson6, Pearson6(4P), Weibull, Weibull (3P) as to obtain best fit for different rainfall series. Based on Kolmogorov- Smirnov (KS test), Anderson-Darling test and chi-square goodness of fit test statistic values, three different rankings were been given to each of the distribution of different rainfall series of Kadapa district. No rank was given to the distribution when the concerned test fails to fit the data.

JUNE

The selected distributions were tried to fit for rainfall series of June during the study period of Kadapa district. Based on the highest total score value from the three GOF tests, the best fitted distribution was identified as Log-Pearson 3 (Score: 59), as per table 4. For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.06807, 0.17036 and 1.3555 respectively. The parameters for shape, scale and location of this fitted distribution were estimated as 35.486, 0.12961 and -0.86964 respectively, as represented in table 5. The probability density function of the fitted distribution was also depicted in Figure 1. It was found that probability of getting rainfall more than 100 mm, 200 mm and 300 mm were identified as 13.0%, 2.9% and 1% respectively, as per table 6.

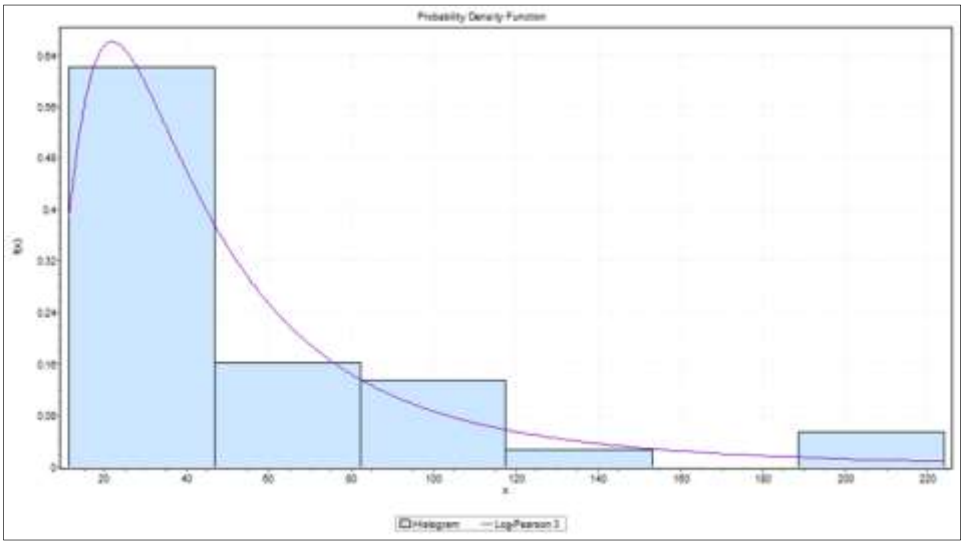

Fig 1: Log-Pearson 3 distribution for June month during the study period of Kadapa district

JULY

Based on the selected criterion, the highest total score value (Score: 53) from the three GOF tests was identified for Gamma (3P), during the study period of July month in Kadapa district, as per table 4. For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.09032, 0.32961 and 2.9331 respectively. The parameters for shape, scale and location of fitted distribution were estimated as 1.6775, 42.639 and 7.0144 respectively, as represented in table 5.The probability density function of the fitted distribution was also depicted in Figure 2. It was found that probability of getting rainfall more than 100 mm, 200 mm and 300 mm were identified as 27.1%, 3.8% and 0.5% respectively, as per table 6.

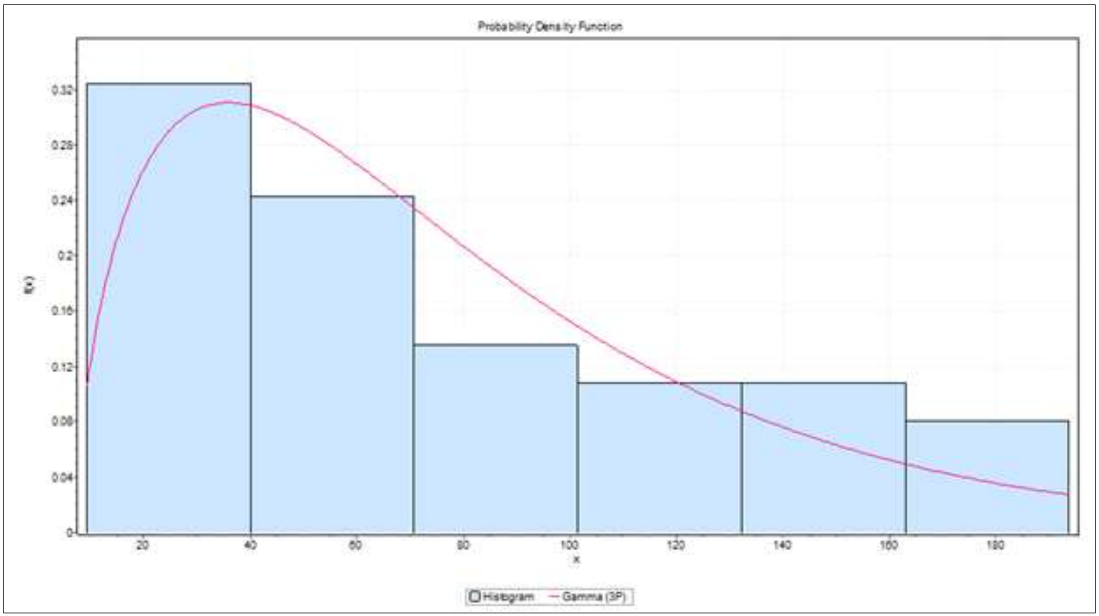

Fig 2: Gamma (3P) distribution for July month during the study period of Kadapa district

August

The selected distributions were tried to fit for rainfall series of August during the study period of Kadapa district. Based on the highest total score value from the three GOF tests, the best fitted distribution was identified as Log-logistic (3P) (Score: 61), as per table 4. For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.06505, 0.1595 and 0.38366 respectively. The parameters for shape, scale and location of this fitted distribution were estimated as 7.6653, 216.17 and -125.85 respectively, as represented in table 5.The probability density function of the fitted distribution was also depicted in Figure 3. It was found that probability of getting rainfall more than 100 mm, 200 mm and 300 mm were identified as 41.7%, 4.1% and 0.6% respectively, as per table 6.

September

Based on the selected criterion, the highest total score value (Score: 61) from the three GOF tests was identified for GEV, during the study period of September in Kadapa district, as per table 4. For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.07385, 0.27819 and 0.43717 respectively. The parameters for scale and location of this fitted distribution were estimated as -0.04539, 51.691 and 77.572 respectively, as represented in table 5. The probability density function of the fitted distribution was also depicted in Figure 4. It was found that probability of getting rainfall more than 100 mm, 200 mm and 300 mm were identified as 47.5%, 7.8% and 0.8% respectively, as per table 6.

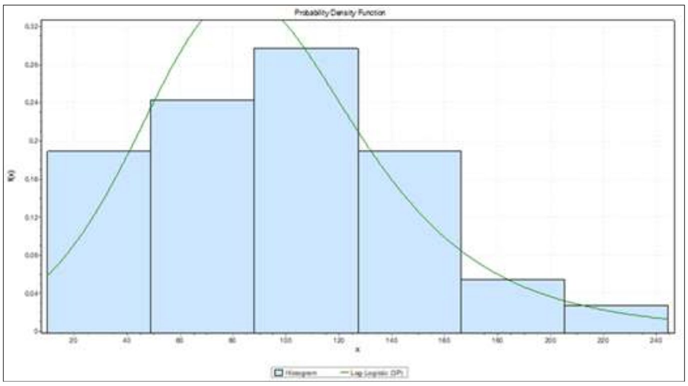

Fig 3: Log-Logistic (3P) distribution for August month during the study period of Kadapa district

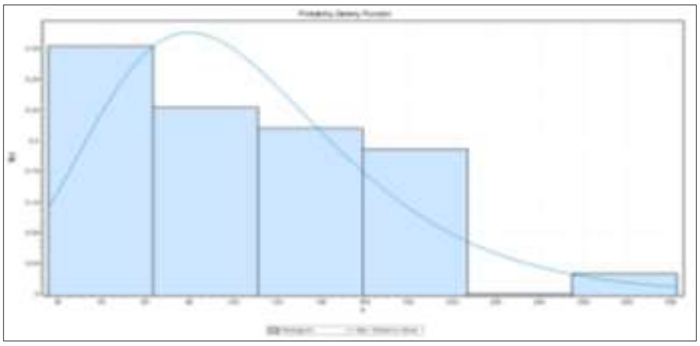

Fig 4: GEV distribution for September month during the study period of Kadapa district

South-West Monsoon

The selected distributions were tried to fit for rainfall series of South-West Monsoon during the study period of Kadapa district. Based on the highest total score value from the three GOF tests, the best fitted distribution was identified as Log-logistic (3P) and Pearson 5 (4P) (Score: 56), as per table 4. For Log-logistic (3P) distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.06713, 0.16115 and 1.9236 respectively. The parameters for shape, scale and location of Log-logistic (3P) distribution were estimated as 9.4105, 690.87 and -366.88 respectively. Similarly, the parameters for Pearson 5 (4P) distribution were estimated as 61.545, 61848 and -685.91, as represented in table 5. The probability density function of the fitted distribution was also depicted in Figure 5. It was found that through Log-logistic (3P) distribution, probability of getting rainfall more than 100 mm, 200 mm, 300 mm, 400 mm and 500 mm were identified as 97.6%, 86.5%, 58.2%, 27.2% and 10.6% respectively, as per table 6.

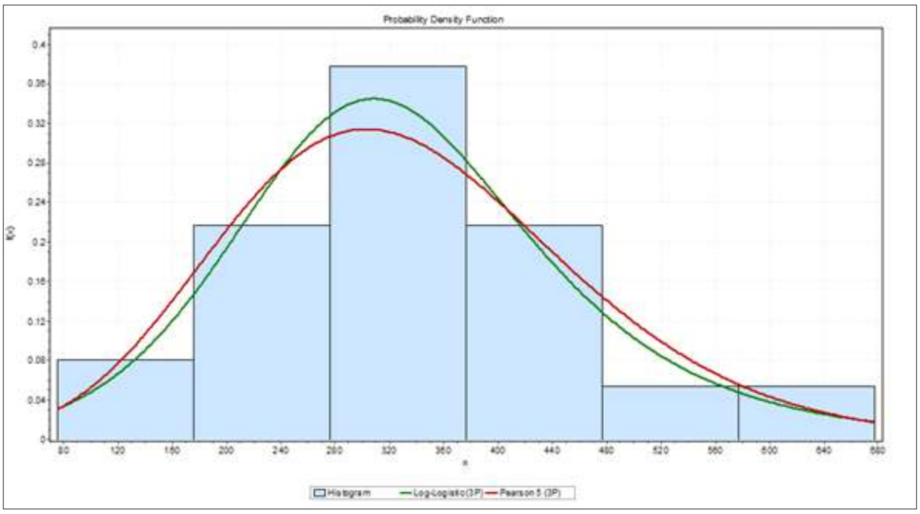

Fig 5: Log-logistic (3P) and Pearson 5 (3P) distribution for South-West Monsoon during the study period of Kadapa district

North-East Monsoon

Based on the selected criterion, the highest total score value (Score: 63) from the three GOF tests was identified for GEV, during the study period of North-East monsoon in Kadapa district, as per table 4. For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.08148, 0.18438 and 0.26079 respectively. The parameters for shape, scale and location of this fitted distribution were estimated as -0.00933, 99.631 and 164.04 respectively, as represented in table 5. The probability density function of the fitted distribution was also depicted in Figure 6. It was found that probability of getting rainfall more than 100 mm, 200 mm, 300 mm, 400 mm and 500 mm were identified as 85%, 50.2%, 22.4%, 8.7% and 3.2% respectively, as per table 6.

Annual: The selected distributions were tried to fit for annual rainfall series during the study period in Kadapa district. Based on the highest total score value from the three GOF tests, the best fitted distribution was identified as Log-normal (3P) and Pearson 5 (4P) (Score: 62), as per table 4. For Log-normal (3P) distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.05722, 0.1484 and 0.24415 respectively. The parameters for shape, scale and location of Log-normal (3P) distribution were estimated as 0.13048, 7.4566 and -1134 respectively. Similarly, the parameters for Pearson 5 (4P) distribution were estimated as 106.11, 2.4556E+5 and -1724, as represented in table 5. The probability density function of the fitted distribution was also depicted in Figure 7. It was found that through Log-normal (3P) probability of getting rainfall more than 100 mm, 200 mm, 300 mm, 400 mm and 500 mm were identified as 99.5%, 97.7%, 92.6%, 82.3% and 67.1% respectively, as per table 6.

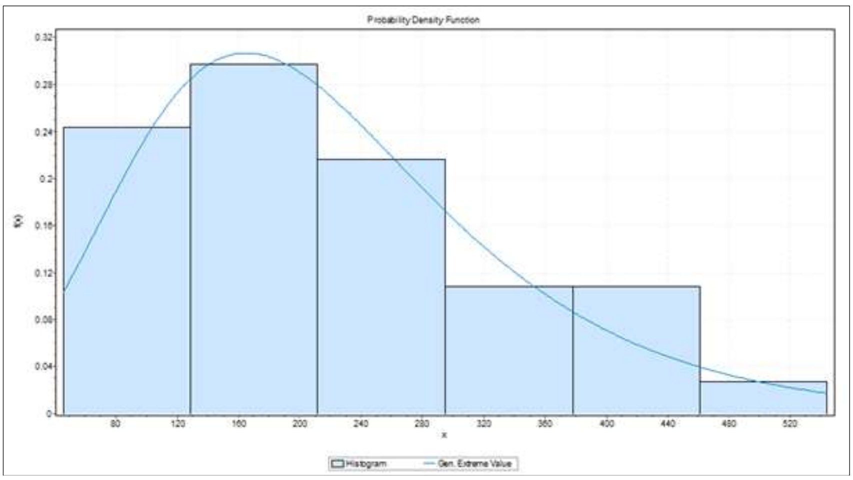

Fig 6: GEV distribution for North-East Monsoon during the study period of Kadapa district

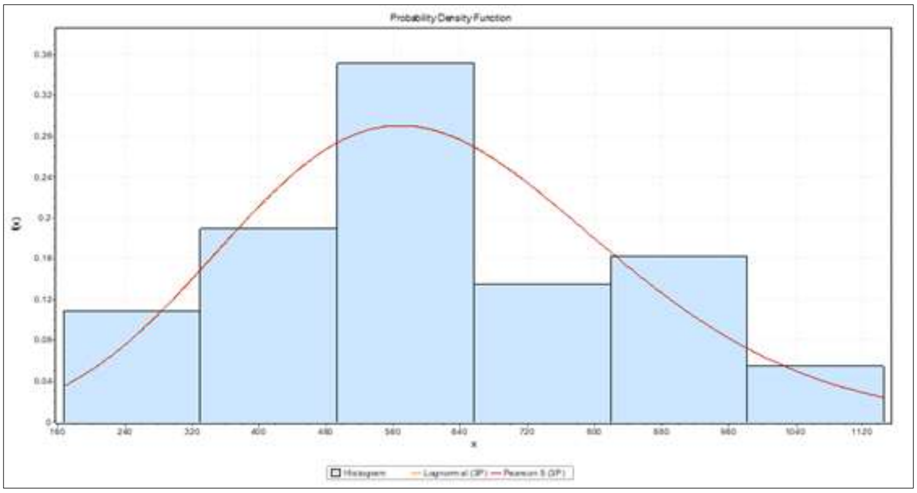

Fig 7: Log-Normal (3P) and Pearson 5 (3P) distribution for Annual period during the study period of Kadapa district

Conclusion

By the study, it was found that the average annual rainfall of Kadapa district was 612.02 mm over the study period. Through Mann-Kendall test, for the period of June (1.24), September (2.14), South-West Monsoon (5.96) and Annual (11.68) rainfall series had significant increasing trend during the period. By fitting different distributions, Log-Pearson 3, Gamma (3P), Log-Logistic (3P) and GEV distribution were identified as the best fit for June, July, August and September respectively. For the South-West Monsoon, it was found that through Log-logistic(3P) distribution probability of getting rainfall more than 100 mm, 200 mm and 300 mm were identified as 97.6%, 86.5% and 58.2% respectively. Similarly for North-East monsoon, probability of getting rainfall more than 100 mm, 200 mm and 300 mm were identified as 85%, 50.2% and 22.4% respectively. These estimates are crucial for proper water resource management and agricultural planning.

Table 6: Statistic values of best fitted probability distribution for Kadapa district

Month/season/Annual	Distribution	Kolmogorov-Smirnov	Anderson-Darling	Chi-Squared test
June	Log-Pearson 3	0.06807	0.17036	1.3555
July	Gamma (3P)	0.09032	0.32961	2.9331
August	Log-Logistic (3P)	0.06505	0.1595	0.38366
September	GEV	0.07385	0.27819	0.43717
South-West Monsoon	Log-Logistic (3P)	0.06713	0.16115	1.9236
	Pearson 5 (4P)	0.06491	0.17868	2.8265
North-East Monsoon	GEV	0.08148	0.18438	0.26079
Annual	Log-Normal (3P)	0.05722	0.1484	0.24415
	Pearson 5 (4P)	0.05747	0.14788	0.2457

Table 7: Score wise best fitted probability distribution with parameter estimates for Kadapa district

Month/season/Annual	Distribution	Score	Parameters estimated
June	Log-Pearson 3	59	$\alpha = 35.486, \beta = 0.12961, \gamma = -0.86964$
July	Gamma (3P)	53	$\alpha = 1.6775, \beta = 42.639, \gamma = 7.0144$
August	Log-Logistic (3P)	61	$\alpha = 7.6653, \beta = 216.17, \gamma = -125.85$
September	GEV	61	$k = -0.04539, \sigma = 51.691, \quad \mu = 77.572$
South-West Monsoon	Log-Logistic (3P)	56	$\alpha = 9.4105, \beta = 690.87, \gamma = -366.88$
	Pearson 5 (4P)	56	$\alpha = 61.545, \beta = 61848, \gamma = -685.91$
North-East Monsoon	GEV	63	$k = -0.00933, \sigma = 99.631, \quad \mu = 164.04$
Annual	Log-Normal (3P)	62	$\sigma = 0.13048, \mu = 7.4566, \gamma = -1134$
	Pearson 5 (4P)	62	$\alpha = 106.11, \beta = 2.4556E+5, \gamma = -1724$

Table 8: Probabilities of rainfall at various point of exceedance for Anantapur district

Month/season/Annual	Distribution	100 mm	200 mm	300 mm	400 mm	500 mm
June	Log-Pearson 3	13.0	2.9	1.0	-	-
July	Gamma (3P)	27.1	3.8	0.5	-	-
August	Log-Logistic (3P)	41.7	4.1	0.6	-	-
September	GEV	47.5	7.8	0.8	-	-
South-West Monsoon	Log-Logistic (3P)	97.6	86.5	58.2	27.2	10.6
	Pearson 5 (4P)	98.0	85.4	57.6	28.9	11.1
North-East Monsoon	GEV	85.0	50.2	22.4	8.7	3.2
Annual	Log-Normal (3P)	99.5	97.7	92.6	82.3	67.1
	Pearson 5 (4P)	99.5	97.7	92.6	82.3	67.2

References

1. Ghosh S, Roy MK, Biswas SC. Determination of the best fit probability distribution for monthly rainfall data in Bangladesh. American Journal of Mathematics and Statistics. 2016;6(4):170-174.
2. Hamed KH, Rao AR. A modified Mann-Kendall trend test for auto correlated data. Journal of hydrology. 1998;204(1-4):82-196.
3. Kendall MG. Rank Correlation Methods. Charles Griffin, London; c1975.
4. Lettenmaier DP. Detection of trends in water quality data from records with dependent observations. Water Resources Research. 1976;12(5):1037-1046.
5. Mann HB. Nonparametric tests against trend. Econometrica. Journal of the econometric society; c1945. p. 245-259.
6. Pal S, Mazumdar D. Stochastic modelling of monthly rainfall volume during monsoon season over gangetic west Bengal, India. Nature Environment and Pollution Technology. 2015;14(4):951.
7. Rana S, Deoli V, Kashyap PS. Temporal analysis of rainfall trend for Udaipur district of Rajasthan. Indian Journal of Ecology. 2019;46(2):306-310.
8. Reddy BNK, Bhanusree D, Kallakuri S, Tuti MD, Rathod S, Meena A, *et al*. Trend Analysis of Rainfall in Telangana State (India) Using Advanced Statistical Approaches. International Journal of Environment and Climate Change. 2022;12(11):3405-3413.
9. Sen PK. Estimates of the regression coefficient based on Kendall's tau. Journal of the American statistical association. 1968;63(324):1379-1389.
10. Sharma, MA, Singh JB. Use of probability distribution in rainfall analysis. New York Science Journal. 1968;3(9):40-49.
11. Singh RN, Sah S, Das B, Potekar S, Chaudhary A, Pathak H, *et al*. Innovative trend analysis of spatio-temporal variations of rainfall in India during 1901–2019. Theoretical and Applied Climatology. 2021;145(1-2):821-838.
12. Stephens MA. EDF statistics for goodness of fit and some comparisons. Journal of the American statistical Association. 1974;69(347):730-737.
13. Wallis WA, Moore GH. A significance test for time series analysis. Journal of the American Statistical Association. 1941;36(215):401-409.
14. Yue S, Wang C. The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. Water resources management. 2004.18(3): 201-218.