International Journal of Statistics and Applied Mathematics 2023; SP-8(5): 39-46

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(5): 39-46

https://www.mathsjournal.com
Received: 20-07-2023
Accepted: 25-08-2023

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 Anantapur 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 Anantapur 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 the Anantapur district during the period from 1985 to 2021. It was revealed that the highest average rainfall was reported in September (112.01) which ranges 15.93 to 244.07, while the least average was observed in January (1.99), which indicated that the months of January and September 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 (1.12), South-West Monsoon (3.52) and Annual (6.84) rainfall series. Based on Goodness of fit criterion, Log-Normal (3P) was identified as appropriate distribution for South-West monsoon of Anantapur district and it was found that probability of getting rainfall more than 100, 200 and 300 were identified as 96.4%, 80.4% and 49.6% 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.

Anantapur district is the southern-most part of the Rayalaseema region of Andhra Pradesh. The district lies between the coordinates 13° 40' to 15° 15'N latitude and 76° 50' to 78° 30'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 scarce rainfall zone which ranges between 500-670. The major crops in terms of area are 86% ground nut, 3.3% paddy and 10.7% other crops. This district is now producing important crops like Sweet Orange, Sapota, Pomegranate, Mango, Banana, Papaya, Guava, Melon and Vegetables. 90% of agriculture is under rainfed conditions in this district.

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) [11]. 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 Anantapur 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 Anantapur 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 selected district of Andhra Pradesh during the study period. So, initially, randomness of each data series was verified by Wallis and Moore phase-frequency test (Wallis and Moore, 1941) [15]. If the randomness of data series was found to be violated, then Modified Mann-Kendall test was tried instead of Mann-Kendall test (Naveena *et al*, 2023) [9].

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) [6] 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) [6] 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 pth 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 (-K) test

Even though M-K test is most co only 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) [3]. 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) [16] developed Modified Mann-Kendall (-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:

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

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) [5] 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), Ga a, Ga a (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	$f(x) = \lambda exp^{(-\lambda x)}$	$0 \le x < +\infty$ $\lambda > 0$	$\lambda = inverse\ scale\ parameter$
Exponential (2P)	$f(x) = \lambda exp^{[-\lambda(x-\Upsilon)]}$	$0 \le x < +\infty$ $\lambda > 0$	$\lambda = inverse\ scale\ parameter$ Υ = location parameter
Ferchet	$f(\mathrm{x};\ \alpha, \mathrm{s}, \mathrm{m}) = \frac{\propto}{s}\left(\frac{x-m}{s}\right)^{-1-\propto} e^{-\left(\frac{x-m}{s}\right)^{-\propto}}$	$\propto > 0$ $s > 0$ $-\infty < m < +\infty$	$\propto = shape\ parameter$ $s = scale\ parameter$ $m = location\ parameter$
Ga a (1P)	$\mathrm{f(x)} = \frac{1}{\tau(k)} x^{k-1} \exp(-x)$	$0 < x < \infty$ $k > 0$	$k = shape\ parameter$ $\beta = scale\ parameter$ $\gamma = location\ parameter$ $\tau\ is\ the\ gamma\ function$
Ga a (3P)	$\mathrm{f(x)} = \frac{(x-\gamma)^{k-1}}{\tau(k)\beta^k} x^{k-1} \exp(-\frac{(x-\gamma)}{\beta})$	$k > 0, \beta > 0\ \gamma > 0,$ $\gamma \le x \le \pm\infty$	
Generalized extreme value (GEV)	$\mathrm{f(x)}=\begin{cases}\frac{1}{\beta} exp\left[-(1+kz)^{\frac{-1}{k}}\right](1+kz)^{\frac{-1}{k}} & k \ne 0\\ \frac{1}{\beta} exp[-z - \exp(-z)] & k = 0\end{cases}$	$1+k\,z > 0$ for $k \ne 0$ $-\infty < x < +\infty$ 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)	$f(x) = \frac{1}{x\|\beta\|\tau(\alpha)}\left(\frac{\ln x-\gamma}{\beta}\right)^{\alpha-1} exp\left(-\frac{\ln x-\gamma}{\beta}\right)$	$0 \le x \le e^{\gamma},\ \beta < 0$ $e^{\gamma} \le x \le +\infty,\ \beta > 0$	$\alpha = shape\ parameter$ $\beta = scale\ parameter,$ $\gamma = location\ parameter$
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) [2].

Anderson –Darling Test

The Anderson-Darling test (Stephens, 1974) [14] 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 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 (Sharma and Singh, 2010) [13]. 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 (Sharma and Singh, 2010) [13]. According to this method, among the selected distributions (For eg: 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 Anantapur district during the period from 1985 to 2021, as to know the basic behaviour of rainfall. The selected measures were namely Mean, Standard Deviation (SD), Coefficient of Variation (CV%), Minimum, Maximum, Skewness and Kurtosis.
From Table 2, it was revealed that the highest average rainfall was reported in September (112.01) which ranges 15.93 to 244.07, while the least average was observed in January (1.99), which indicated that the months of January and September were the extreme precipitation months of Anantapur District of Andhra Pradesh during the study period (1985-2021). Based on CV%, highest values was observed for February (246.7%), which might be due to heavy irregularities of rainfall during the period and where the least was observed for October (56.29%). Skewness measures the asy etry of a distribution around the mean. For the same table-4.3, rainfall during the months (Jan-Dec) and monsoons were positively skewed. The maximum skewness (2.29) was obtained for Winter season, where as the Annual period was negatively skewed (-0.19). Kurtosis provides an idea about the flatness or peakedness of the frequency distribution curve. Lepto kurtic (>3) values were observed for the periods of March, February, December, November, winter season and January only, which indicated that those data comprised of extreme outliers.

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

Month	Mean	SD	CV	Skewness	Kurtosis	Maximum	Minimum
January	1.99	3.67	184.96	2.16	3.84	14.42	0.00
February	2.50	6.17	246.70	3.34	11.69	29.51	0.00
March	6.99	15.28	218.59	4.14	19.38	84.55	0.00
April	16.38	17.89	109.21	1.53	1.82	65.75	0.00
May	42.60	29.72	69.77	0.54	-0.07	119.97	0.00
June	57.36	36.75	64.07	1.28	1.84	169.33	9.16
July	60.97	51.51	84.48	1.43	2.29	235.19	4.19
August	74.02	43.69	59.02	0.39	-0.65	171.28	7.00
September	112.01	67.47	60.24	0.44	-1.09	244.07	15.93
October	101.19	56.95	56.29	0.29	-0.62	225.61	7.14
November	36.84	43.29	117.51	2.16	5.65	202.23	1.41
December	6.48	9.72	150.07	2.81	9.97	49.37	0.00
Seasonal							
Southwest monsoon	304.36	122.02	40.09	0.19	-0.44	566.95	56.33
Northeast monsoon	144.51	73.17	50.63	0.20	0.02	335.47	16.72
winter	4.49	7.30	162.70	2.29	5.01	29.51	0.00
Annual	519.33	168.95	32.53	-0.19	0.02	819.79	86.26

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 Anantapur district of Andhra Pradesh during the study period 1985-2021. So, initially, randomness of each data series was been verified by Wallis and Moore test and If the randomness of data series had violated, modified Mann-Kendall test was tried instead of Mann-Kendall test.

Table 5: Trend analysis for monthly, seasonal and annual rainfall (year) of Anantapur 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**
January	2.13	**0.03**	-	-	1.64	0.10	0.00
February	1.33	0.18	1.02	0.24	-	-	0.00
March	0.27	0.79	1.08	0.28	-	-	0.03
April	1.07	0.29	1.57	0.12	-	-	0.28
May	0.93	0.35	1.33	0.18	-	-	0.69
June	0.13	0.89	2.11	0.04	-	-	**1.12**
July	0.53	0.59	1.01	0.31	-	-	0.61
August	1.07	0.29	0.77	0.44	-	-	0.56
September	1.47	0.14	1.53	0.13	-	-	1.29
October	1.07	0.29	0.51	0.61	-	-	0.76
November	1.33	0.18	0.72	0.47	-	-	0.18
December	0.13	0.89	0.98	0.33	-	-	0.04
Seasonal							
Southwest monsoon	1.47	0.14	2.01	0.03	-	-	3.52
Northeast monsoon	1.87	0.06	0.85	0.40	-	-	1.02
winter	0.27	0.79	1.04	0.32	-	-	0.03
Annual	0.27	0.79	2.13	0.03	-	-	6.84

From Table-3, it was revealed through Wallis and Moore phase-frequency test that the monthly [except (January)], seasonal and annual rainfall were found to be random in nature, as p-values were greater than 5% level of significance. Hence, Modified Mann-Kendall test was applied only for the non-random series i.e., January and it was obtained a non-significant trend. 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 (1.12), South-West Monsoon (3.52) and Annual (6.84) 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 Agrawal *et al*. (2021) that there was significant increasing trend for the south-west monsoon and annual period.

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 Anantapur district of Andhra Pradesh were tried to fit for different distributions such as Exponential, Exponential(2P), Ferchet, Ferchet (3P), Ga a, Ga a(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 Anantapur district. No rank was given to the distribution when the concerned test fails to fit the data.

June Month

Initially, the selected distributions were tried to fit for rainfall series of June during the study period of Anantapur district. Based on the highest total score (Score: 63) value from the these GOF tests i.e., KS test, AD test and Chi-Square test, the best fitted distribution was identified as Generalized Extreme Value (GEV), as per Table 4. Test statistic values of KS test, AD test and Chi-Square test were obtained as 0.09, 0.25 and 0.33 respectively. The parameters for shape, scale and location of this fitted distribution were estimated as 0.10076, 25.762 and 39.656 respectively, as represented in Table 5. It was found that probability of getting rainfall more than 100 and 200 were identified as 11.5% and 0.8% respectively, as per Table 6. Similar kind of distribution (GEV) was identified as appropriate to fit weekly rainfall series in Pantnagar, as reported by Sharma and Singh (2010) [13].

July Month

Based on selected criterion, the highest total score value (Score: 55) from the three GOF tests was identified for Ga a (3P), during study period of July month in Anantapur district, as as per Table 4. For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.10959, 0.40967 and 4.4859 respectively. The parameters for shape, scale and location of this fitted distribution were estimated as 1.1225, 50.785 and 3.9659 respectively, as represented in Table 5. It was found that probability of getting rainfall more than 100, 200 and 300 were identified as 18.1%, 2.7% and 0.4% respectively, as per Table 6.

August Month

The selected distributions were tried to fit for rainfall series of August during the study period of Anantapur district, as per the Table 4. Based on the highest total score value from the three GOF tests, the best fitted distribution was identified as GEV (Score: 63). For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.07883, 0.27405 and 1.8403 respectively. The parameters for shape, scale and location of this fitted distribution were estimated as -0.12395, 40.291 and 55.21 respectively, as represented in Table 5. It was found that probability of getting rainfall more than 100 and 200 were identified as 26.1% and 0.9% respectively, as per Table 6. Similar kind of distribution (GEV) was identified as appropriate to fit maximum hourly during the study period of Kurnool region, as reported by Mallikarjuna *et al.* (2011) [7].

September Month

The selected distributions were tried to fit to for rainfall series of September during the study period of Anantapur district. Based on the highest total score value from the three GOF tests, the best fitted distribution was identified as Pearson 6(4P) (Score: 54), as per Table 4. For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.13839, 0.52942 and 4.4318 respectively. The parameters for this fitted distribution were estimated as shape $\alpha_1 = 2.2298$, $\alpha_2 = 13.532$, scale (598.49) and location (7.9921), as represented in Table 5. It was found that probability of getting rainfall more than 100, 200 and 300 were identified as 46.7%, 12.6% and 3.3% respectively, as per Table 6.

South-West Monsoon

Based on selected criterion, the highest total score value (Score: 56) from the three GOF tests was identified for Log-Normal (3P), during the study period of South-West monsoon in Anantapur district, as per Table 4. For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.07553, 0.1909 and 1.0416 respectively. The parameters for shape, scale and location of this fitted distribution were estimated as 0.09211, 7.1711 and -1002.4 respectively, as represented in Table 5. It was found that probability of getting rainfall more than 100, 200, 300, 400 and 500 were identified as 96.4%, 80.4%, 49.6%, 20.8% and 5.9% respectively, as per Table 6.

North-East Monsoon

Based on the highest total score value (Score: 59) obtained from the three GOF tests, both GEV and Normal distributions were identified as best-fit for North-East monsoon in Anantapur district, as per Table 4. For the GEV distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.10304, 0.29781 and 0.35618 respectively. Conversely, the Normal distribution showed test statistic values of 0.09281, 0.2668, and 0.44734 for the same tests. The parameters for shape, scale and location of GEV distribution were estimated as -0.26041, 72.658 and 117.81 respectively. On the other hand, the parameters for Normal distribution were 73.168 (mean) and 144.51(SD), as represented in Table 5. It was found that through GEV distribution, probability of getting rainfall more than 100, 200 and 300 were identified as 71.9%, 23.03% and 1.7% respectively, as per Table 6. Similar kind of distribution (Normal) was identified as appropriate to fit annual during the study period of Sudan, as reported by Mohamed and Ibrahim (2015) [8].

Annual

The selected distributions were tried to fit for Annual rainfall series during the study period of Anantapur district, as per Table 4. Based on the highest total score value from the three GOF tests, the best fitted distribution was identified as Pearson 6(4P) (Score: 58). For this fitted distribution, test statistic values of KS test, AD test and Chi-Square test were obtained as 0.07377, 0.22362 and 2.0219 respectively. The parameters for this fitted distribution were estimated as shape $\alpha_1 = 5489.4$, $\alpha_2 = 1820.7$, scale (2059.4) and location (-5691.3), as represented in Table 5. It was found that probability of getting rainfall more than 100, 200, 300, 400 and 500 were identified as 99.5%, 97.5%, 90.8%, 76.3% and 54.4% respectively, as per Table 6.

Conclusion

By the study, it was found that the average annual rainfall of Anantapur district was 519.33 over the study period. Through Mann-Kendall test, for the period of June (1.12), Southwest monsoon (3.52) and annual (6.84) had significant increasing trend during the period. By fitting different distributions, Generalized Extreme Value, Ga a (3P), Perason6(4P), Log-Normal(3P) and Normal distribution were identified as the best fit for June, July, September, South-West Monsoon and North-East Monsoon respectively. For the South-West Monsoon, probability of getting rainfall more than 100, 200 and 300 were identified as 96.4%, 80.4% and 49.6% respectively. Similarly for North-East monsoon, probability of getting rainfall more than 100, 200 and 300 were identified as 71.9%, 23.03% and 1.7% respectively. These estimates are crucial for proper water resource management and agricultural planning.

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

Month/season/Annual	Distribution	Kolmogorov-Smirnov	Anderson-Darling	Chi-Squared test
June	Gen. Extreme Value	0.09	0.25	0.33
July	Ga a (3P)	0.10959	0.40967	4.4859
August	Gen. Extreme Value	0.07883	0.27405	1.8403
September	Pearson 6 (4P)	0.13839	0.52942	4.4318
South–West Monsoon	Lognormal (3P)	0.07553	0.1909	1.0416
North-East Monsoon	Gen. Extreme Value	0.10304	0.29781	0.35618
North-East Monsoon	Normal	0.09281	0.2668	0.44734
Annual	Pearson 6 (4P)	0.07377	0.22362	2.0219

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

Month/season/Annual	Distribution	Score	Parameters estimated
June	Generalized extreme value	63	$k = 0.10076, \sigma = 25.762, \quad \mu = 39.656$
July	Ga a (3P)	55	$\alpha = 1.1225, \beta = 50.785, \gamma = 3.9659$
August	Generalized extreme value	63	$k = -0.12395, \sigma = 40.291, \quad \mu = 55.21$
September	Pearson 6 (4P)	54	$\alpha_1 = 2.2298, \quad \alpha_2 = 13.532, \beta = 598.49, \gamma = 7.9921$
South-west monsoon	Log-normal (3P)	56	$\sigma = 0.09211, \mu = 7.1711, \gamma = -1002.4$
North-east monsoon	Normal	59	$\sigma = 73.168, \mu = 144.51$
	Generalized extreme value	59	$k = -0.26041, \sigma = 72.658, \quad \mu = 117.81$
Annual	Pearson 6 (4P)	58	$\alpha_1 = 5489.4, \quad \alpha_2 = 1820.7, \beta = 2059.4, \gamma = -5691.3$

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

Month/season/Annual	Distribution	100	200	300	400	500
June	Generalized extreme value	11.5	0.8	-	-	-
July	Ga a (3P)	18.1	2.7	0.4	-	-
August	Generalized extreme value	26.1	0.9	-	-	-
September	Pearson 6 (4P)	46.7	12.6	3.3	-	-
South-west monsoon	Log-normal (3P)	96.4	80.4	49.6	20.8	5.9
North-east monsoon	Generalized extreme value	71.9	23.0	1.7	-	-
North-east monsoon	Normal	72.9	22.4	1.7	-	-
Annual	Pearson 6 (4P)	99.5	97.5	90.8	76.3	54.4

References

1. Agarwal S, Suchithra AS, Singh SP. Analysis and interpretation of rainfall trend using Mann-Kendall's and Sen's slope method. Indian Journal of Ecology. 2021;48(2):453-457.
2. 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.
3. Hamed KH, Rao AR. A modified Mann-Kendall trend test for auto correlated data. Journal of hydrology. 1998;204(1-4):82-196.
4. Kendall MG. Rank Correlation Methods. Charles Griffin, London; c1975.
5. Lettenmaier DP. Detection of trends in water quality data from records with dependent observations. Water Resources Research. 1976;12(5):1037-1046.
6. Mann HB. Nonparametric tests against trend. Econometrica. Journal of the econometric society; c1945. p. 245-259.
7. Mallikarjuna P, Jyothy SA, Hemanath K. Probability Distribution Analysis of Maximum Hourly Ranifall Intensity–A Case Study. Hydrology. 2011;34(1):42-49.
8. Mohamed TM, Ibrahim AAA. Fitting probability distributions of annual rainfall in Sudan. Journal of Engineering and Computer Sciences. 2015;17(2):34-39.
9. Naveena K, Sajan DP, Surendran U, Shada B. An advanced approach for Assessment of Spatiotemporal Variation in Rainfall and its Impact on Stream Flow Discharge in Chaliyar River Basin of Kerala; c2023.
10. 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.
11. 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.
12. Sen PK. Estimates of the regression coefficient based on Kendall's tau. Journal of the American statistical association. 1968;63(324):1379-1389.
13. Sharma, MA, Singh JB. Use of probability distribution in rainfall analysis. New York Science Journal. 2010;3(9):40-49.
14. Stephens MA. EDF statistics for goodness of fit and some comparisons. Journal of the American statistical Association. 1974;69(347):730-737.
15. Wallis WA, Moore GH. A significance test for time series analysis. Journal of the American Statistical Association. 1941;36(215):401-409.
16. 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.