Scientific Publishing House

Trends in Ecological and Indoor Environmental Engineering

RESEARCH ARTICLE
https://dx.doi.org/10.62622/TEIEE.023.1.1.24-34

OPEN ACCESS
ISSN 2956-9761

DETERMINATION OF SUITABLE PROBABILITY DISTRIBUTION OF RAINFALL IN PAKISTAN CONSIDERING MULTIPLICITY

Tasir Khan[1*], Fazeela Mureed[2], Mohammad Anwar[3], Owais Khan[4], Muhammad Ali[5], Waheed Ullah[6]
[1]School of Mathematics and Statistics, Gansu Key Laboratory of Applied Mathematics and Complex Systems, Lanzhou University, Lanzhou 730000, China
[2]School of Environmental Science and Engineering, North China Electric Power University, Beijing, China
[3]School of Public Administration, Department of Land Resources Management, Zhejiang Gongshang University, Zhejiang Hangzhou 310018, China
[4]Department of Environmental Science, Abdul Wali Khan University, Mardan, Pakistan
[5]School of geography and environment, Shandong normal university, Jinan, China
[6]Department of Environmental Sciences, COMSATS University Islamabad, Abbottabad, Pakistan
*Corresponding email: tasirkhan.state@gmail.com

Received: 19.12.2023; ***Accepted:*** 28.12.2023; ***Published:*** 30.12.2023

Cite this article: Khan, T., Mureed, F., Anwar, M., Khan, O., Ali, M., & Ullah, W. (2023). Determination of Suitable Probability Distribution of Rainfall in Pakistan Considering Multiplicity. *Trends in Ecological and Indoor Environment Engineering*, 1(1), 24–34.

Extreme rainfall events are occasional, and understanding their intensity and frequency is important for long-term planning and for public safety. The current study aims to investigate the stability of extreme precipitation events in different regions of Pakistan, their independence and the uniformity of their occurrence. A rainfall frequency analysis (RFA) was conducted based on information from 8 weather stations namely, the distribution and moments L of the annual maximum precipitation in Pakistan were investigated; Using goodness-of-fit criteria, the study determined the possible distribution of precipitation. Relatively Absolute Error (RAE) results are based on the most appropriate GUM distribution, revealed that the Sialkot, Multan, Faisalabad and DI Khan stations produce very low errors (0.266, 0.847, 0.075, 0.856, 1.671, 2.522, 3.659, 4.524). It was found that the most suitable distribution is LN distribution for Peshawar, Sibbi and Karachi Stations. The GEV distribution also performs well with small errors for various return periods (0.266, 0.847, 0.075, 0.856, 1.671, 2.522, 3.659, 4.524), corresponding to the return periods 2, 5, 25, 50, 100, 200, 500 and 1000 years. In contrast, P3 has the advantage of a return period of 10 years for all stations, as they all produce similar results for this particular return period. The current research provides valuable insights into estimating extreme rainfall at stations where rainfall data are available.

Keywords: extreme rainfall; L-moments; multiple frequency analysis; Pakistan.

INTRODUCTION

Hydraulic and hydrological design is a main step in planning any water project. Any problems during the design phase can result in design failure, regardless of how the other steps are taken correctly. Hydrologist is concerned with water-related issues, including those associated with the quality, quantity, and availability of water, collectively referred to as hydrological events in the field of humanities. Random methods are frequently used to understand the bases of uncertainty that occur in the physical process to produce observed hydrological events, for example, river flow and precipitation evaluations depend on future or past events. Some statistical methods are provided to summarize and minimize the frequency analysis and uncertainty of observation data. Frequency analysis in hydrology aims to determine the likelihood of specific events occurring by estimating the quintile QT for a T return period. Here, Q represents the measure of events that occur at a particular location and time (Khan et al., 2017). Any rise in temperature, mainly in the light of reduced precipitation over Saudi Arabia, could have a significant impact on water supply and agriculture. Precipitation varies widely, wet areas tend to be wet, and dry areas become drier. These may be prerequisites for flash floods produced by droughts caused by heavy rains and insufficient rainfall. Any change in climate change can lead to changes in extreme weather events, such as heavy rains, high temperatures and cold snaps, as well as prolonged droughts (Almazroui et al., 2012).

The generalize normal is an appropriate distribution for the maximum return period to estimate regional quantile estimation and for generalize extreme value distribution based on the overall region of the low return cycle of relative absolute bias and relative mean square error. For high weather modelling, the common statistical distributions are as follows: log-Pearson III, generalized extreme value (GEV), logistic distribution, generalized Pareto and log normal distribution. These distributions are evaluated using frequency techniques based on both Partial L-moments (PL-moments) and L-moments in various studies (Khan et al., 2017; Shahzadi et al., 2013; Zakaria et al., 2012). The study found that PL-moments are more suitable than L-moments for these distributions. In recent times, extreme weather events such as heat waves, droughts, floods, wildfire and sandstorm, have increased in intensity and frequency in several parts of the world. In fact, in South America there has been a decrease in the proportion of cold nights and an increase in warm nights (Fawad et al., 2018).

Extreme rainfall events are occasional, and understanding their intensity and frequency is important for long-term planning and for public safety (Alam et al., 2021). However, due to limited records and the need infer the distribution at a location without observations, it is difficult to evaluate the possibility of extreme events (Cooley et al., 2007). Change in regional and global precipitation characteristics is one of the most relevant features of climate variation in warming regions; but there is little consensus on the observation and the predictable changes in spatial precipitation patterns. The increasing rates of extreme precipitation values are influenced by various factors, including vertical velocity outline and its change (Darnthamrongkul & Mozingo, 2021; Yue & Hashino, 2007; Salam et al,. 2022). It has been identified that the probability distribution types of monthly, seasonal and annual precipitation are basically the same. Overall, the log-Pearson type-III distribution (LP3) and Pearson type-III (P3) distribution are acceptable distribution types used to represent precipitation, such as in Japan. For monthly observation, P3 is the most suitable distribution, and for seasonal observations, LP3 is the most suitable distribution, while yearly the log normal (LN) distribution provides the best fitted observations, with log normal (LN), generalized extreme value (GEV) and Pearson type-III (P3) distribution as possible alternatives (Koh et al., 2008). The study concludes that the precipitation data were optimized for regionalization, the region's climate and geographical homogeneity were classified to estimate the probability distributions to be GEV distributions between the application distribution, and the results found that the regional analysis process could greatly reduce the relative bias (RBias) and relative root mean square error (RRMSE) in predicting design rainfall.

Rwanda is renowned for its "thousand hills," and its rugged topography, in conjunction with human activities, which contributes to local fluctuations in precipitation patterns throughout the country. This in turn has led to devastating floods over the past few years. The country has suffered severe drought events, floods and land sides related to ENSO (El-Niño-Southern Oscillation) events. Heavy rainfall in 1997/1998 engulfed the floor plan and caused other related environmental damage. Similarly, drought events in 1999/2000 had a significant impact on the Umutara, Bugesera, and Mayaga areas (Wagesho & Claire, 2016). Heavy precipitation events caused several devastating floods in North America. The most destructive was the 1993 flooding in the upper Mississippi river, which caused an estimated $18 billion in damage (Salam et al., 2021). These events show that societies remain vulnerable to extreme weather, and it also raises questions about whether the frequency of precipitation events that produce floods has changed (Kunkel et al., 1999).

There are many probability modes useful for continuous random variables, like each year maximum rainfall. In Brazil, a more simplified theoretical probability model is fitting. Common observations, such as 2 and 3 parameters Gumbel distributions, generalized extreme value and log-normal (Back et al., 2011; Cassalho et al., 2018; Senapeng & Busababodhin, 2017).

However, continuous random variables can be represented by multiple probability models. The best suited to model the data series is selected by non-parameter testing to estimate the relationship between the theoretical frequency and the observational frequency. In the hydrological area, the fit tests, Anderson-Darling, chi-square and Kolmogorov-Smirnov test can be highlighted (Back et al., 2011; Cooley et al., 2007).

The current study aims to investigate the stability of extreme precipitation events in different regions of Pakistan, their independence and the uniformity of their occurrence. Thus, the study focuses on rainfall frequency analysis (RFA) based on information from 8 meteorological stations: namely: (i) to establish the most accurate distribution and L moments of annual maximum rainfall in Pakistan; (ii) determine their possible distribution using means of goodness-of-fit; (iii) calculate quantiles using the selected distributions.

MATERIALS AND METHODS

Study area and data

Pakistan's climate is changeable and has many characteristics. Among them, the most important are wind speed, temperature, humidity, precipitation, and altitude. Some regions of Pakistan are hot, dry and desert. In the study, meteorological data for 36 years was used to measure the probable rainfall trend, and eight stations from the major cities were selected from the Pakistan Meteorological Department (PMD), Islamabad. The eight stations are located in Sialkot, Peshawar, Multan, Lahore, Faisalabad, DI Khan, Sibbi, Karachi, in the states of Sindh, Punjab, and Khyber Pakhtunkhwa. The annual maximum rainfall (AMR) data from PMD were obtained and is used in this study. The data of AMR, spanning a long-term period from 1981 to 2016 inclusive, were analysed across several decadal periods.

This paper investigates the characteristics of maximum rainfall at all eight sites mentioned. Figure 1 shows the geographic location of the annual maximum precipitation.

Figure 1. The studied stations on the map of Pakistan

Table 1 lists the visual statistic for each site, among which mean, kurtosis, skewness, minimum, maximum rainfall, as well as the latitude, altitude, and longitude. The AMR series has a record length of 36 years. The mean rainfall exceeds 69.332 but lower 403.262. The skewness of the Faisalabad and DI Khan stations 2, which can be interpreted as a significant degree of skewness. The skewness of the remaining stations is in the range of 0.794 to 1.466 and can be seen as around symmetric to asymmetric. As can be seen from Table 1, the maximum rainfall is between 217.30 to 839.60. The station latitude is determined by coordinates 24°54' ... 34°0' north latitude and 67°4' ... 74°31' east longitude. Extreme precipitation is explained by the altitude above sea level, which is a major influencing factor. In this study, heights ranged from 4 to 359 m. Peshawar station has the highest altitude of 359 m, relative to all other stations, while the Sialkot and Lahore stations (256 m, 215 m high) are close to the Peshawar station in terms of height. Nevertheless, the Karachi station has the lowest altitude 4 m.

Preliminary analysis of AMR series

Before frequency analysis, annual sequences should correspond certain statistical conditions, such as homogeneity, independence and stationarity. These are common assumptions for frequency analysis in extreme events, such as precipitation extremes, extreme temperature, and floods. To assess stability, trends, homogeneity, and independence, the Augmented Dickey-Fuller test (ADF-test), Mann-Kendall test (MK-test), Wald-Wolfowitz test (WW-test) and Mann-Whitney U test (MWU-test), were applied to the annual maximum rainfall series.

Mann-Kendall test (MK-test) for assessing trends

The non-parametric MK-test is typically used to study monotonic patterns of decreasing or increasing trends in the data being studied (Mann, 1945; Gilbert, 1987). It allows you to understand whether the trend is positive or negative. The null hypothesis (H_o) of the MK-test states that there is no monotonic trend in the AMR sequence. The t_p is random variable is identically distributed. The MK-test statistics are as:

$$Z_{MK} = \begin{cases} \frac{(S-1)}{\sqrt{var[R]}} \text{ if } S > 0, \\ 0 \text{ if } S = 0, \\ \frac{(S+1)}{\sqrt{var[R]}} \text{ if } S < 0, \end{cases} \quad (1)$$

where

$$S = \sum_{f=1}^{n-1} \sum_{e=f+1}^{n} \text{sign}(X_e - X_f), \quad (2)$$

$$\text{Sign}(X_e - X_f) = \begin{cases} 1 \text{ if } (X_e - X_f) > 0, \\ 0 \text{ if } (X_e - X_f) = 0, \\ \text{ if } (X_e - X_f) < 0, \end{cases} \quad (3)$$

$$\text{Var}(s) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{g} t_p (t_p - 1)(2t_p + 5)\right], \quad (4)$$

where x_e and x_f are observation of time e and f individually. In the data set, g denotes the number of binding groups, and t_p represents the number of data points in the P^{th} binding groups.

If $n > 8$ the S statistic obey the normal distribution. The statistics S obey the normal distribution given that $Z_{MK} \geq Z_{1-\alpha/2}$, H_o is rejected.

Estimation of stationarity by ADF-test

The ADF-test, established by (Dickey & Fuller, 1979), is an augmented version of the originally created Dickey Fuller (DF) test and is usually applicable when it is necessary to assess the stationarity of a sequence. In the ADF-test, the variables under study are subject to an autoregressive process. The DF-test parametric high order correlation related to the original DF-test. The central difference between the two tests is that the DF-test contains past effects and only one autoregressive term. But the ADF-test applies to a larger and more intricate time series model. Autoregressive process of the Dickey Fuller test as follows:

$$y_t = \rho y_{t-1} + \varepsilon_{t,}\ t = 1, 2, 3, \ldots\ldots, T \quad (5)$$

where ρ is the autoregressive and $\varepsilon_{t,}$ is a random element of the model, which matches the characteristics of the white noise process. In the null hypothesis H_0, where ρ =1, the sequence is a unit root test, indicating that the variable studied is non-stationary, for example $I(1)$. In the alternative hypotheses H_1 where $|\rho| < 1$, the series does not contain the unit root indicating that the sequence is stationary, denoted as $I(0)$. To calculate the DF-test, alternative equation involves subtracting y_t-1 from both side of equation (5):

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t \quad (6)$$

where $\beta = \rho - 1$.

The DF-test statistics are equated as:

$$t_{DF} = \frac{\rho - 1}{S_{\hat{P}}} \quad (7)$$

In equation (7), ρ is the estimated value of ρ, and $S_{\hat{P}}$ is the estimated standard error of ρ. Under H_0, t_{DF} follows the DF distribution. The critical values of the DF distribution found through simulation have been established by (Dickey & Fuller, 1976). The DF equation (5) is extended to include a linear trend in the equation as follows:

$$y_t = \alpha_0 + \rho y_{t-1} + \varepsilon_t, \quad (8)$$

$$y_t = \alpha_0 + \alpha_1 t + \rho y_{t-1} + \varepsilon_t. \quad (9)$$

If the components of the DF models are auto correlated, equation (5) can be transformed as:

$$y_t = \rho y_{t-1} + \sum_{i=1}^{p-1} r_i \Delta y_{t-i} + \varepsilon_t. \quad (10)$$

Referred to as the ADF-test, it is expressed as:

$$y_t = (\rho - 1) y_{t-1} + \sum_{i=1}^{p-1} r_i \Delta y_{t-i} + \varepsilon_t. \quad (11)$$

In ADF-test, selecting the appropriate choice of hypothesis (p) can be a challenging task. According to (Schwert & Statistics, 2002), it recommends to set the maximum hypothesis order as $p_{max} = 12\frac{T}{100}$. The effectiveness of the test is affected by the presence of autocorrelation. When p is small, the test tends to be less effective, and when p is large, the test may suffer from inefficiency. Furthermore, equation 11 can be extended to include a linear trend as follows:

$$y_t = d_t + y_{t-1} + \sum_{i=1}^{p-1} r_i \Delta y_{t-i} + \varepsilon_t, \quad (12)$$

where,

$$d_t = \sum_{i=0}^{p} \alpha_i t^i \text{ for } p = 0.1. \quad (13)$$

Here, d_t is the portion of equation 12, and it is worth noting that the asymptotic distribution of the AD-test statistic is the same as the ADF-test statistic.

WW-test for assessing independence

Independence in this context means that the annual maximum rainfall occurrence at one site does not affect the presence or absence of maximum rainfall at any other site being observed; each site is independent of the other. Hydrological variable checks are usually performed on assumptions of independence, such as annual average, totals, maximums or minimums, seasonal, monthly and other intervals of time, as well as extreme number of non-annual samples (like partial duration sequences). The nonparametric The WW-test, originally created by (Wald & Wolfowitz, 1943), is often used to identify the instances of independence within a given series of values (Huang et al., 2018; Santos et al., 2009). It is also used to examine the existence of the data trends. Let, $x_1, x_2, x_3, \ldots, x_n$ be the observed value of the variable. Then, recommend the Q statistic, and subsequently recommend the R statistic, as shown by (Huang et al., 2018).

$$Q = \sum_{i=1}^{n-1} x_i x_{i+1} + x_1 x_n \quad (14)$$

The normal distribution of the R statistic is as follows:

$$\bar{Q} = \frac{(S_1^2 - S_2)}{n-1} \quad (15)$$

$$\text{Var}(Q) = \frac{(S_2^2 - S_4)}{n-1} - \bar{Q}^2 + \frac{(S_1^4 - 4S_1^2 S_2 + 4S_1 S_3 + S_2^2 - 2S_4)}{(n-1)(n-2)} \quad (16)$$

Where $S_q = nm'_q$ and m'_q are the q^{th} moments relative to the sample origin, the test moments relative to the sample origin. The test statistic can be expressed as:

$$R = \frac{(Q - \bar{Q})}{\sqrt{\text{var}(Q)}}. \quad (17)$$

The test statistic R is used to examine the independence of a data set at the significance level α. In the case when the standard normal variable $\frac{u\alpha}{2}$, which corresponds to excess probability $\frac{\alpha}{2}$, is greater than the scale of statistics u, then the null hypothesis H_0 must be accepted; this indicates the independence of the studied variables.

MWU-test for homogeneity

The MWU-test is a nonparametric test established by (Mann & Whitney, 1947) to determine whether two samples are drawn from the same population. When the assumption of normality is violated, the MWU-test can replace the t-test. This test is besides it is known as the U test and is commonly used to assess the assumption of consistency. To conduct this test, consider two independent sample sizes, denoted as p and q, with $p \leq q$. Let N be the total number of observations, where $N = p + q$. Arrange all the samples in ascending order. The MWU-test depends on the minimum valve of "U", which is determined by the minimum variables V and W, as explained by (Mann & Whitney, 1947; Mello et al., 2013):

$$V = Y - \left\{\frac{P(P+1)}{2}\right\}, \quad (18)$$

$$W = pq - V, \quad (19)$$

$$U = \min(V, W), \quad (20)$$

where the primary sample size is p and the second sample size are q. Y is rank of the primary sample p in the united N series. $Y = \sum_{i=1}^{n} R_i$, where R_i is the rank series N of the primary sample. The number of times of the components in first sample in ranking follows the components in the second sample, q. W describes the case where the primary sample, p, follows the second under the null hypothesis. The test statistics is determined as:

$$u = \left[\frac{U - \bar{U}}{\sqrt{Var(U)}}\right], \quad (21)$$

$$\bar{U} = \frac{pq}{2}, \quad (22)$$

$$Var(U) = \left[\frac{pq(p+q+1)}{12}\right]. \quad (23)$$

In the presence a parallel rank, the variance formula is as follows:

$$Var(U) = \left(\frac{pq}{12}\right)\left[(N+1) - \sum_{i=1}^{i} \frac{J_i^3 - J_i}{N(N-1)}\right], \quad (24)$$

where J_i represents the number of observations of the sharing level K and are the number of parallel levels.

Probability distributions of applicants (PD)

In this study, maximum rainfall frequency analysis (MRFA) used several distributions. These distributions include two-parameter distribution and seven multi-parameter distributions that are commonly used in five extreme event frequency analysis. These distributions include generalized extreme value (GEV), Log-Normal, Pearson type-III (P3), and log-Pearson type-III (LP3), Gumbel (GUM) distributions.

Most of these distributions have been proposed for on-site rainfall, wind and temperature analysis (Fawad et al., 2018).

Estimation parameters of PDs by the method of L-moments

In this study, the L-moments method is used for frequency analysis for extreme precipitation to estimate the parameters of the partial discharge. This technique, proposed by (Hosking & Wallis, 1997), is based on a linear combination of sequential statistics that have been ranked in descending or ascending direction (Huang et al., 2018). The L-moments are more consistent because they are less sensitive to outliers, making them suitable when the sample size is small. They have also been related to the maximum likelihood and method of moments approaches in statistical estimation (Alam et al., 2016). The L-moment is also expressed by probability weighed moments (PWM):

$$\beta_r = E[X\{F(X)\}^R], \quad (25)$$

where $r = 1, 2, 3, \ldots$.

The L-moments λ_{r+1} can be found as a linear combination probability weighted moment:

$$\lambda_{r+1} = \sum_{k=0}^{r} P_{r,k}^{*} \beta_r, \quad (26)$$

$$P_{r,k}^{*} = (-1)^{r-k}\binom{r}{k}\binom{r+k}{k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}. \quad (27)$$

The following equations can determine the first four overall L-moments $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ involving position, ratio, L kurtosis and L skewness, separately:

$$\lambda_{1,} = \beta_0, \quad (28)$$

$$\lambda_2 = 2\beta_1 - \beta_0, \quad (29)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0, \quad (30)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0, \quad (31)$$

$$\text{L-CV: } \tau = \frac{\lambda_3}{\lambda_2}, \quad (32)$$

$$\text{L-skewness: } \tau_3 = \frac{\lambda_3}{\lambda_2}, \quad (33)$$

$$\text{L-kurtosis: } \tau_4 = \frac{\lambda_4}{\lambda_2} \quad (34)$$

Repeatedly use b_r can be unbiased estimator of β_r which as:

$$b_r = n^{-1} \sum_{j=r+1}^{n} \frac{(j-1)(j-2)\ldots.(j-r)}{(n-1)(n-2)\ldots.(n-r)} x_{j:n}, \quad (35)$$

To determine the relation between the first four sample L-moments with probability weighted moments, you should use the following formulas:

$$l_1 = b_0, \quad (36)$$

$$l_2 = 2b_1 - b_0, \quad (37)$$

$$l_3 = 6b_2 - 6b_1 + b_0 \quad (38)$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0. \quad (39)$$

The sample L-moments ratios are:

$$t = \frac{l_2}{l_1}, \quad (40)$$

$$t_3 = \frac{l_3}{l_2}, \quad (41)$$

$$t_4 = \frac{l_4}{l_2}, \quad (42)$$

where t is the quantile L-moments coefficient of variance t_3 and t_4, the L-kurtosis and L-skewness is useful for relating probability distributions respectively.

Goodness of fit tests

The specific PD depends on a number of factors, among which comparison of PD, the availability of maximum rainfall data, and the method of parameter estimation. In the current study, the Kolmogorov-Simonov test, the Anderson-Darling test, and chi-square test to variability criteria were applied. These tests, namely AD, KS and chi-square test statistics, help to describe how well the data conforms to a given distribution. Goodness of fit describes the difference between the theoretical values and the actual data sequence calculated from the tested distribution. In addition, to assess the fitness visually, various techniques were employed, including a quantile-quantile (Q-Q) plot, L-moment ratio plot, a PP plot and extreme value plots. The advantage of the above graphical test and fit test have been applied to maximum rainfall data (Yuan et al., 2018; Kunz et al., 2010; Lawan et al., 2015; Khan et al., 2021). In the context of climate change, goodness of fit is consistently used to select the best fit distribution for modelling.

Kolmogorov-Smirnov (KS) test

An empirical distribution function (EDF)-based KS test is used to determine if a sample is from hypothetical continuous distribution (Khan et al., 2022). Suppose a random sample $x_1, x_2, x_3, \ldots, x_n$ from a distribution, which EDF is give as:

$$F_{n(x)} = \frac{1}{n}[\text{number of observation } \leq x]. \quad (43)$$

The KS test depends on the maximum vertical distance between the EDF and the theoretical PD. The test statistics are given as:

$$D = \max_{1 \le i \le n} \left[F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right]. \quad (44)$$

In expression (44), $F(x_i)$ is the cumulative distribution function, x_i is the i^{th} order, while n is the sample size.

Test of Anderson-Darling (AD)

The test of Anderson-Darling (AD) evaluates the goodness of fit of a distribution. The AD test assigns a more significance to the tail of the distribution, since this is the main property of modelling extreme phenomena (Huang et al., 2018). The AD test statistic A^2 can be expressed as:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \times$$

$$\times \left[\ln F(x_i) + \ln\left(1 - F(X_{n-i+1})\right) \right], \quad (45)$$

where A^2 corresponds to the test result, X is the variable, $F(X_i)$ is the distribution function, and n is the sample size.

RESULTS AND DISCUSSION

The initial step is to check the homogeneity, independence, stationarity and frequency analysis of AMR series. Stationarity indicates that, in addition to random changes, the annual maximum rainfall series over time is unchanged. Non-stationary characteristics are characterized by trends, jumps and oscillations. Trends in climatic conditions may be due to regular changes, while the long-term climate fluctuations may follow cyclical patterns. Jumps are mainly associated with changes in flood frequency, often due to abrupt variations in river systems. Independence means that each observation in the sequence is not influenced by any other observations. In practice, the degree of dependence can vary depending on the spacing between consecutive elements in the sequence. Typically, the dependence is weaker between annual maximum values, while the dependence between consecutive values is often stronger.

Homogeneity, in this context, signifies that all observations in the data originate from the same population. In the case of extreme events, floods, rainfall and snowmelt, if heterogeneity is too high, it can be challenging to interpret the trends. Although in this particular case, it can be accepted as uniform depending on the test results obtained. At the same time, using annual series, heterogeneity is easier to detect. To identify and assess heterogeneity, tests such as the ADF-test, MK-test, the homogeneity Mann-Whitney U and the independent WW-test were performed. All workstations in the annual maximum precipitation series of the above tests were passed at the level of 5 % significance. The results of the ADF-test, MK-test results uniformity test, MWU-test and WW-test are presented in Table 2. In cases where the assumptions about the AMRF series are not seen, more complicated statistical methods that consider observed data and their correlation change over time must be employed. Additionally, when investigating monotonic trends, the results of the MK-test are consistent with the patterns observed in time series plots. Reference (Durrans & Kirby, 2004) applied a consistent methodology to examine the modulation in the probability of extremes in rainfall and temperature, and they identified statistically significant long-term increases in extreme maximum temperatures but with marked regional and seasonal variations. To analyse these changes, they used Wilcoxon test, and the results were summarized using boxplots for four AEP: 50 %, 10 %, 5 % and 1 %. They considered stationary and non-stationary GEV distributions in the analysis.

Table 1. Data analysis for the studied stations

Variable	SD	sKENESS	Kurtosis	Mean	Max rainfall	Latitude	Longitude	Altitude
Sialkot	157.127	0.794	3.097	403.262	839.60	32°30'	74°31'	256 m
Peshawar	69.962	1.466	6.563	147.000	409.00	34°0'	71°35'	359 m
Multan	46.065	1.081	3.772	90.389	217.30	30°11'	71°28'	123 m
Lahore	109.824	1.359	5.472	259.000	640.00	31°33'	74°19'	215 m
Faisalabad	69.384	2.224	9.516	144.881	435.300	31°25'	73°44'	184 m
DI Khan	64.040	2.018	9.004	111.508	376.00	31°49'	70°54'	164 m
Sibbi	36.119	0.824	4.299	69.332	188.10	29°09'	68°29'	130 m
Karachi	76.094	0.812	2.705	98.586	270.40	24°54'	67°4'	4 m

Table 2. Testing the assumption of independence, stationarity and homogeneity of annual rainfall sequence of eight stations

Station	MK-test		Spearman test		WW-test		MWU-test	
	T Stat	P-value	T Stat	P-value	T State	P-value	T State	P-value
Sialkot	-0.0601	0.610	-0.766	0.221	0.808	0.209	-2.187	0.014
Peshawar	0.0376	0.753	0.358	0.361	1.342	0,089	-0.091	0.463
Multan	0.045	0.704	0.485	0.313	0.542	0.293	-0.121	0.451
Lahore	0.048	0.6851	0.466	0.320	1.727	0.042	-0.182	0.427
Faisalabad	0.137	0.2391	1.221	0.110	1.278	0.100	-1.245	0.106
DI Khan	0.114	0.3266	1.019	0.153	-0.373	0.354	-1.033	0.150
Sibbi	0.183	0.1135	1.615	0.053	0.547	0.292	-1.154	0.124
Karachi	0	1	-0.019	0.492	0.590	0.277	-0.151	0.439

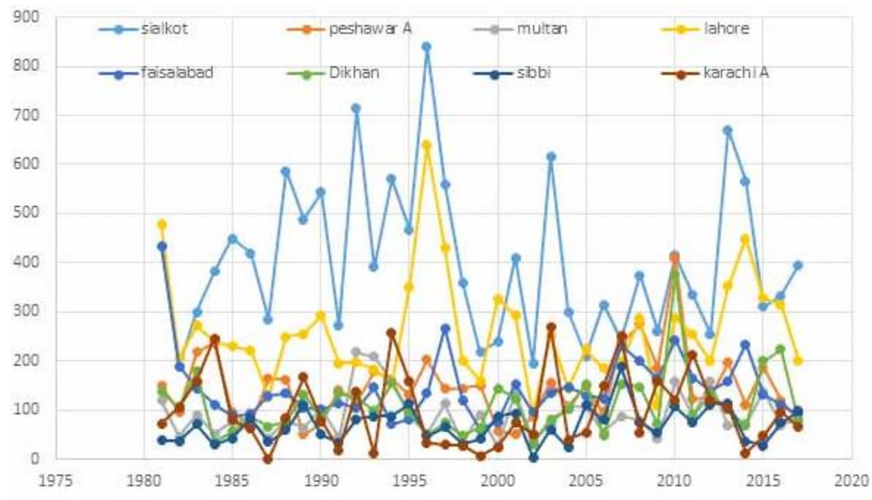


Figure 2. Time series charts for the stations under study

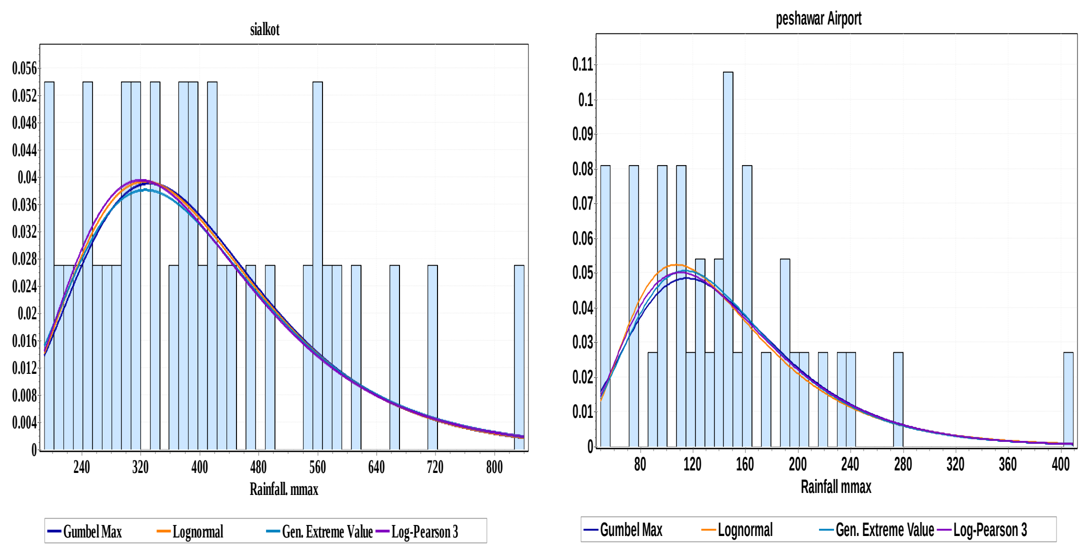
Probability distribution function (PDF)

The annual maximum rainfall for the selected stations is analysed using PDFs. Figure 3 shows the annual maximum rainfall series and probability distribution analysis for the representative stations Sialkot, Peshawar, Multan, Lahore, Faisalabad, DI Khan, Sibbi, and Karachi airport. The results show that the probability distribution of these PDFs varies from one station to another.

Probability distributions (PDs) selection

Modelling of extreme rainfall for design of rainfall farms. To accurately determine the PD of maximum precipitation is important for measuring climate change in a region. Primarily considered are five PDs in the study, such as GEV, LN, LP3, GUM (Max) distribution. The goodness of fit was used to evaluate these PD, for example Kolmogorov-Simonov test, Anderson-Darling test and chi-square test. The results of statistical data for best fit distribution based on KS, AD and chi-square test are recorded in Table 4 to Table 6. The best fit distribution for each station is based on the goodness of fit test between all PDs measured in the study and the distribution selected at the 5 % significance level according to the AD and KS tests. The main result of each PD is given in Table 4 to Table 6. L skewness, L-coefficient of variation and L kurtosis in Table 3 gives the calculation results for all sites. Table 3 shows that the Sibbi station has lower L kurtosis than other sites, while Sialkot and Multan stations have moderate, and the others are higher kurtosis. The skewness of the Sialkot, Peshawar and Sibbi stations is moderate, and the others exhibit high skewness. The remaining stations display lower L-moments skewness and kurtosis. Overall, all stations show lower L-CV. The results of the AD test indicate that in Peshawar and Lahore, LP3 is the most suitable distribution. For Karachi and Sibbi stations, LN is the most suitable distribution. The GUM distribution is found to be the most suitable for others sites, namely Sialkot, Multan, Lahore, Faisalabad, DI Khan, see Table 4. According to the KS test results, GUM (Max) is found to the best-fitted distribution for five stations, namely Sialkot, Multan, Lahore, Faisalabad and DI Khan. Meanwhile, LN is suitable distribution for Peshawar, Sibbi and Karachi stations as shown in Table 5. Similarly, the chi-square test is used to find the most appropriate PDF from the selected stations. Here we calculate and find the most suitable PDF for the selected stations, and the results are shown in Table 6.

The results show that the best-suited distribution of the eight selected stations is the LN distribution, which includes Peshawar, Faisalabad, DI Khan, Karachi. However, the generalize extreme value (GEV) distribution is found to be the most fitting for Multan, Lahore and Sibbi, while for Sialkot, GUM (Max) distribution is the best fit. Therefore, we conclude that the LN, GUM (Max) and GEV are the best PDF for chi-square, used to predict extreme precipitation.



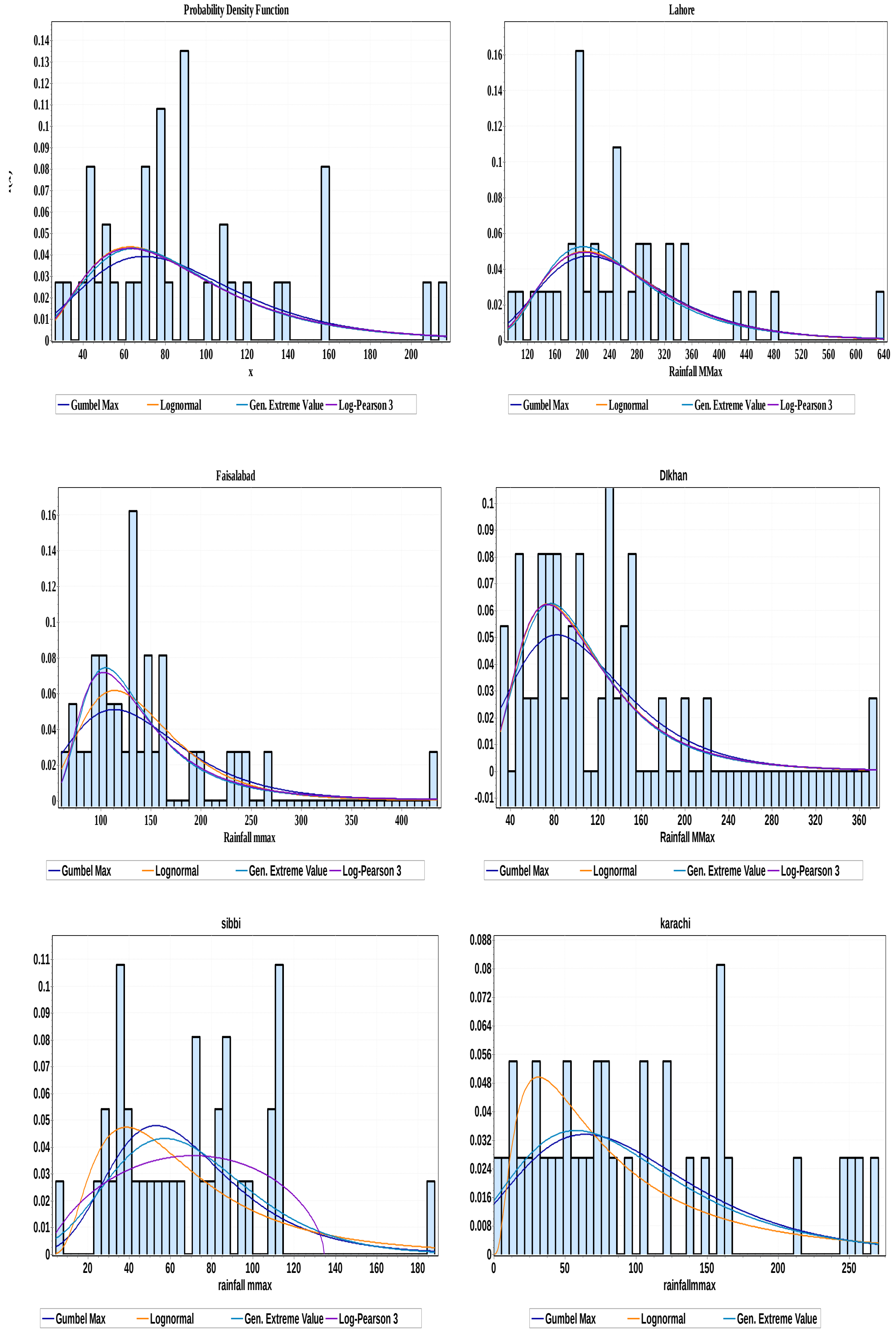


Figure 3. PDF of annual maximum temperature (AMT) for eight stations

Table 3 Sample L-Ratios for different stations

Location	L CV	L-SKEWNESS	L-KURTOSIS
Sialkot	0.2191	0.1800391	0.1024245
Peshawar	0.2521	0.1860525	0.2076851
Multan	0.2780	0.232522	0.1658959
Lahore	0.2250	0.2251192	0.2246678
Faisalabad	0.2342	0.3220915	0.2805230
DI Khan	0.2903	0.2545115	0.2115040
Sibbi	0.2888	0.107069	0.089558
Karachi	0.2521	0.211365	0.116301

Table 4 Goodness of fit augmented Dickey-Fuller (AD) test

Locations	AD test for goodness of fit				Best fit PDF
	GEV	LN	LP3	GUM (max)	
Sialkot	0.16158	0.18408	0.16336	0.19884	GUM
Peshawar	0.18509	0.20753	0.21518	0.19478	LP3
Mulan	0.19083	0.19466	0.18781	0.25958	GUM
Lahore	0.21372	0.22118	0.22968	0.25718	GUM
Faisalabad	0.20413	0.35649	0.22786	0.81658	GUM
DI Khan	0.20452	0.21365	0.21821	0.43271	GUM
Sibbi	0.40534	0.88048	0.9204	0.5088	LN
Karachi	0.24778	0.4288		0.33533	LN

Table 5 Goodness of fit Kolmogorov-Simonov (KS) test

Locations	Ks test				Best fit PDF
	GEV	LN	LP3	GUM (max)	
Sialkot	0.07434	0.07782	0.07677	0.0797	GUM
Peshawar A	0.08321	0.0864	0.08418	0.07732	LN
Multan	0.07241	0.07261	0.07539	0.09811	GUM
Lahore	0.0912	0.08757	0.09068	0.0931	GUM
Faisalabad	0.08582	0.1039	0.07985	0.13032	GUM
DI Khan	0.08311	0.07943	0.08187	0.10393	GUM
Sibbi	0.08649	0.12716	0.08347	0.10583	LN
Karachi	0.0601	0.08104		0.07393	LN

Table 6. Goodness of fit chi-square test

Stations	Chi-square test				Best fit PDF
	GEV	LN	LP3	GUM (max)	
Sialkot	0.0704	0.50577	0.16731	0.5433	GUM
Peshawar	3.417	2.0654	3.3963	3.4471	LN
Mulan	0.16469	0.21794	0.20552	0.39477	GEV
Lahore	0.19422	1.1696	1.177	0.91276	GEV
Faisalabad	3.5969	1.0149	3.1896	4.4716	LN
DI Khan	0.25369	0.19945	0.20422	1.7277	LN
Sibbi	1.3688	2.2866	2.110	2.5952	GEV
Karachi	1.9151	1.1799	1.9332	1.4956	LN

The results of the KS test indicate that GUM (Max) distribution is THE most appropriate for five stations, while the LN distribution is the best for the remaining three stations. However, for the AD test, the GUM (max) distribution is found to be the best fitted distribution for four stations, while the LP3 and LN distributions are the best for two stations each. In the chi-square test, GEV distribution is the best fit for three stations, LN distribution for four and GUM (Max) distribution for one station. All selection periods fit the GEV distribution and estimate the indicator. The likelihood ratio test shows that the best model involves a linear increase in the location parameter, while the indicators shape and amount remain constant. Model diagnosis including quantile plots, probability plots and density plot showed a good degree of fit. The GOF test, such as AD and KS tests, shows that modelling approach yields nearly identical fitting results. Reference (Rahman et al., 2013) observed two parameter distribution performance, particularly in the case of analysing Gamma and LN2 probability distribution function. A reference (Beskow et al., 2015) suggested that for maximum rainfall series, the number of adjusted series for GLO distribution of EV1 is the least, which is considered to be the most suitable PDF for only one sequence. Although the presentation in the Rio Grande was not satisfactory, some authors informed permissible GLO distribution indicators in regional studies (Beskow et al., 2015; Khan et al., 2021), because GEV has made appropriate adjustments to more than 96 % of the series. In this study, compared with other PDFs, GEV offers better modifications for shorter sequences and better modification for longer average sequences. This feature is very important for areas where hydrological observation often lacks an extended historical sequence. Reference (Ng et al., 2020) focused on a suitable probability distribution to examine the annual maximum atmospheric falls in hot regions of the Malaysian Peninsular. They explored the fitting of Gumbel, Gamma, log-Pearson type-III and GEV distributions to describe the maximum annual atmospheric falls in the river basin. They evaluated the performance of these distributions by using GOF tests and found that the GEV distribution is the best fitting distribution to characterize the precipitation sequence, because it has the advantages of high flexibility and efficiency.

Analysis of maximum annual precipitation shows that a subpopulation is present in the observations, and therefore indicates that, apparently due to high water vapor content, the weather mechanism that causes heavy rainfall has changed, temperature variation between continuous and incoming air, duration of low pressure in a given area, local conditions and thermodynamic balance of atmospheric conditions (Młyński et al., 2018). In another work, (Villarini, 2012) proposed that log Gamma and Weibull distribution are the most suitable distributions for central Poland. Another study by (Wdowikowski et al., 2016) proved that the Generalized Pareto and exponential distributions are the best method to estimate P maxp% in the Odra Basin. Furthermore, Yuan et al. (2018) believes that the GEV function is the best distribution for estimation of the highest annual average precipitation. Due to the high mid-tall, there is a specific probability of excess in central European countries. In spite of such studies, it should be noted that the form of the probability distribution, especially its indicators, is tightly interconnected to the meteorological, physical and geographic characteristics of the area that affect atmospheric fall. The means that the determination of the precise probability of exceeding maximum rainfall is highly dependent on these factors.

Relative absolute error

Extreme precipitation's prediction plays an important role in turbine design, society and engineering design. A reference (Beskow et al., 2015) emphasizes the importance of examining the effect of PD selection on estimating the number of countless related to a predefined return period. The calculation of the Relatively Absolute Error (RAE) involves estimating the design annual maximum rainfall using an acceptable fitting probability distribution (PD) and comparing it to the design maximum rainfall estimated by the most suitable PD. The RAE has been determined with application the following equation as suggested in works (Beskow et al., 2015; Cassalho et al., 2018):

$$\mathrm{RAE} = \left|\frac{\mathrm{A-B}}{\mathrm{B}}\right| \times 100,$$

where A represents the quantile estimated using a specific PD, B represents the quantile estimations obtained from the best-suited fitted PD.

Table 7 summarizes the relative absolute error related with each station using the most acceptable distribution based on the AD and KS tests, as well as chi-square criteria. This emphasizes the importance of considering multiple distributions in the analysis.

Table 7. Relative Absolute Error of quantile estimates (%)

Stations	Best fit	Acceptable fit	Return period								
			2	5	10	25	50	100	200	500	1000
Sialkot	GUM	GEV	0.569	1.760	1.333	0.075	1.523	3.178	4.972	7.401	9.270
		LN	3.856	1.238	0.494	2.478	3.871	5.232	6.653	8.511	9.935
		PE3	0.266	0.847	0.642	0.075	0.856	1.671	2.522	3.659	4.524
Peshawar	LN	GUM	0.569	1.760	1.333	0.075	1.523	3.178	4.972	7.401	9.270
		PE3	3.856	1.238	0.494	2.478	3.871	5.232	6.653	8.511	9.935
		GEV	0.266	0.847	0.622	0.075	0.856	1.671	2.522	3.659	4.524
Multan	GUM	GEV	0.569	1.760	1.333	0.075	1.523	3.178	4.972	7.401	9.270
		LN	3.856	1.238	0.494	2.478	3.871	5.232	6.653	8.511	9.935
		PE3	0.266	0.847	0.422	0.075	0.856	1.671	2.522	3.659	4.524
Lahore	GEV	GUM	0.569	1.760	1.333	0.075	1.523	3.178	4.972	7.401	9.270
		LN	3.856	1.238	0.494	2.478	3.871	5.232	6.653	8.511	9.935
		PE3	0.266	0.847	0.622	0.075	0.856	1.671	2.522	3.659	4.524
Faisalabad	GUM	GEV	0.569	1.760	1.333	0.075	1.523	3.178	4.972	7.401	9.270
		LN	3.856	1.238	0.494	2.478	3.871	5.232	6.653	8.511	9.935
		PE3	0.266	0.847	0.622	0.075	0.856	1.671	2.522	3.659	4.524
DI Khan	GUM	GEV	0.569	1.760	1.333	0.075	1.523	3.178	4.972	7.401	9.270
		LN	3.856	1.238	0.494	2.478	3.871	5.232	6.653	8.511	9.935
		PE3	0.266	0.847	0.642	0.075	0.856	1.671	2.522	3.659	4.524
Sibbi	LN	GUM	0.569	1.760	1.333	0.075	1.523	3.178	4.973	7.401	9.270
		PE3	3.856	1.238	0.494	2.478	3.871	5.232	6.653	8.511	9.935
		GEV	0.267	0.847	0.642	0.075	0.856	1.671	2.522	3.659	4.524
Karachi	LN	GUM	0.569	1.760	1.333	0.075	1.523	3.178	4.972	7.401	9.270
		PE3	3.856	1.238	0.494	2.478	3.871	5.232	6.653	8.511	9.935
		GEV	0.266	0.847	0.642	0.075	0.856	1.671	2.522	3.659	4.524

Table 7 shows the quantiles estimated by the PD that is not properly fitted compared to the more satisfactory quantile estimates obtained from the probability distributions. It is important to analyse the different distributions, because even if some of these distributions pass the goodness-of-fit tests, there are still some significant differences in quantile estimation. These differences can have a significant impact on planners, policy makers, and decision makers. As described in Table 7, some distributions have the smallest RAE for each station, indicating that they provide estimates that are closer to the actual values.

RAE results are based on the most appropriate GUM distribution, revealed that the Sialkot, Multan, Faisalabad and DI Khan stations produce very low errors (0.266, 0.847, 0.075, 0.856, 1.671, 2.522, 3.659, 4.524) which is used for P3 distributions among each of the other distributions for all return periods. While LN has the advantage of a return period of 10 years, with an RAE of 0.494. However, from table it is evident that the most suitable distribution is LN distribution for Peshawar, Sibbi and Karachi Stations. The GEV distribution also performs well with small errors for various return periods (0.266, 0.847, 0.075, 0.856, 1.671, 2.522, 3.659, 4.524), corresponding to the return periods 2, 5, 25, 50, 100, 200, 500 and 1000 years. In contrast, P3 has the advantage of a return period of 10 years for all stations, as they all produce similar results for this particular return period. Moreover, for the Lahore station as indicated in Table 7, the P3 distribution yields a small margin of error for return periods ranging from 2 years to 1000 years (RAE values: 0.266, 0.847, 0.075, 0.856, 1.671, 2.522, 3.659, 4.524), while LN has a 10-year return period.

CONCLUSION

Initially, the homogeneity, independence and stationary of the eight stations were tested. For this purpose, time series diagrams, Mann-Whitney U test, Wald-Wolfowitz test and Dickey-Fuller test were applied. Eight stations successfully completed these tests, allowing further analysis of the AMR series at these eight sites.

The L-moments method was applied for parameter evaluation of probability distribution, and various tests and plots, including the AD test, KS test, chi-square test, Q-Q plots, L scale chart, P-P plots and RAE, were utilized to find the best appropriate distribution between all PDs. The goodness-of-fit indicates that GUM (Max), LN, GEV, and P3 are appropriate distributions for different frequency analysis stations.

The KS and AD test results are consistent with the graphical L-ratio graph outcomes.

The chi-square test was also employed, indicating that LP3 distribution is appropriate for most stations.

Smaller errors in certain distributions indicate that they provide close and accurate estimates, making them the preferred choices for estimating extreme rainfall quantiles.

For engineers looking at renewable energy, structural design, and climatology, the RAE and quantile estimates are important for preparing precipitation-based energy development. This result is very useful for the design of rainfall farms, and agricultural applications.

The current research provides valuable insights into estimating extreme rainfall at stations where rainfall data are available.

In the future, the focus will expand to encompass the entire nation, considering all stations for both maximum and minimum rainfall in Pakistan. The aim will be to determine the most suitable probability distribution for site-specific extreme rainfall analysis across the country.

Acknowledgements

We thank our respected reviewers for their valuable comments and suggestions that helped us to improve this paper.

Declaration of conflicting interest

The authors declare no competing interests.

Contributions

All authors contributed to the study's conception and design.

Conceptualization: T.K.; Data curation: T.K.; Formal Analysis: T.K., M.A.; Investigation: F.M.; Methodology: F.M.; Software: M.A., O.K., W.U.; Visualization: M.A.; Writing – original draft: T.K.; Writing – review & editing: M.A., O.K., W.U.

Funding

Not applicable

Data availability statement

All the authors of this manuscript confirmed that the data supporting the findings of this study are available in the article. All the required data is available and easily accessible.

Additional information

Publisher's note

Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Alam, F., Salam, M., Khalil, N. A., khan, O., & Khan, M. (2021). Rainfall trend analysis and weather forecast accuracy in selected parts of Khyber Pakhtunkhwa, Pakistan. *SN Applied Sciences*, 3, 1–14. https://doi.org/10.1007/s42452-021-04457-z.

Alam, J., Muzzammil, M., & Khan, M. K. (2016). Regional flood frequency analysis: comparison of L-moment and conventional approaches for an Indian catchment. *ISH Journal of Hydraulic Engineering*, 22(3), 247–253.https://doi.org/10.1080/09715010.2016.1177739.

Almazroui, M., Nazrul Islam, M., Athar, H., Jones, P. D., & Rahman, M. A. (2012). Recent climate change in the Arabian Peninsula: annual rainfall and temperature analysis of Saudi Arabia for 1978–2009. *International Journal of Climatology*, 32(6), 953–966. https://doi.org/10.1002/joc.3446.

Back, Á. J., Henn, A., & Oliveira, J. L. R. (2011). Equações de chuvas intensas para o estado de Santa Catarina. *Revista Brasileira de Ciência do Solo*, 35, 2127–2134.

Beskow, S., Caldeira, T. L., de Mello, C. R., Faria, L. C., & Guedes, H. A. S. (2015). Multiparameter probability distributions for heavy rainfall modeling in extreme southern Brazil. *Journal of Hydrology: Regional Studies*, 4, 123–133. https://doi.org/10.1016/j.ejrh.2015.06.007.

Cassalho, F., Beskow, S., de Mello, C. R., de Moura, M. M., Kerstner, L., & Ávila, L. F. (2018). At-site flood frequency analysis coupled with multiparameter probability distributions. *Water Resources Management*, 32, 285–300. https://doi.org/10.1007/s11269-017-1810-7.

Cooley, D., Nychka, D., & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479), 824–840. https://doi.org/10.1198/016214506000000780.

Darnthamrongkul, W., & Mozingo, L. A. (2021). Toward sustainable stormwater management: Understanding public appreciation and recognition of urban Low Impact Development (LID) in the San Francisco Bay Area. *Journal of Environmental Management*, 300, 113716. https://doi.org/10.1016/j.jenvman.2021.113716.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531.

Durrans, S. R., & Kirby, J. T. (2004). Regionalization of extreme precipitation estimates for the Alabama rainfall atlas. *Journal of Hydrology*, 295(1–4), 101–107. https://doi.org/10.1016/j.jhydrol.2004.02.021.

Fawad, M., Ahmad, I., Nadeem, F. A., Yan, T., & Abbas, A. (2018). Estimation of wind speed using regional frequency analysis based on linear-moments. *International Journal of Climatology*, 38(12), 4431–4444. https://doi.org/10.1002/joc.5678C.

Hosking, J. R. M., & Wallis, J. R. (1997). Regional Frequency Analysis, by J. R. M. Hosking and James R. Wallis, pp. 240. ISBN 0521430453. Cambridge, UK: *Cambridge University Press*, April 1997.

Huang, K., Chen, L., Zhou, J., Zhang, J., & Singh, V. P. (2018). Flood hydrograph coincidence analysis for mainstream and its tributaries. *Journal of Hydrology*, 565, 341–353. https://doi.org/10.1016/j.jhydrol.2018.08.007.

Khan, S. A., Hussain, I., Hussain, T., Faisal, M., Muhammad, Y. S., & Mohamd Shoukry, A. (2017). Regional frequency analysis of extremes precipitation using L-moments and partial L-moments. *Advances in Meteorology*, 2017.

Khan, T., Ahmad, I., Wang, Y., Salam, M., Shahzadi, A., & Batool, M. (2022). Comparison approach for wind resource assessment to determine the most precise approach. *Energy & Environment*, 0958305X221135981. https://doi.org/10.1177/0958305X22113598.

Khan, T., Wang, Y., & Anwar, M. (2021). Analysis of Annual Maximum Rainfall for Frequency Distribution to Determine the Best-fitted Probability Distribution for Different Sites in Pakistan. *17 August 2021, PREPRINT (Version 1) available at Research Square*. https://doi.org/10.21203/rs.3.rs-780243/v1.

Koh, D. K., Choo, T. H., Maeng, S. J., & Trivedi, C. (2008). Regional frequency analysis for rainfall using L-moment. *The Journal of the Korea Contents Association*, 8(3), 252–263.

Kunkel, K. E., Andsager, K., & Easterling, D. R. (1999). Long-term trends in extreme precipitation events over the conterminous United States and Canada. *Journal of Climate*, 12(8), 2515–2527. https://doi.org/10.1175/1520-0442(1999)012<2515:LTTIEP>2.0.CO;2.

Kunz, M., Mohr, S., Rauthe, M., Lux, R., & Kottmeier, C. (2010). Assessment of extreme wind speeds from Regional Climate Models–Part 1: Estimation of return values and their evaluation. *Natural Hazards and Earth System Sciences*, 10(4), 907–922. https://doi.org/10.5194/nhess-10-907-2010.

Lawan, S. M., Abidin, W. A. W. Z., Chai, W. Y., Baharun, A., & Masri, T. (2015). Wind Energy Potential in Kuching Areas of Sarawak for Small-Scale Power Application. *International Journal of Engineering Research in Africa*, 15, 1–10. https://doi.org/10.4028/www.scientific.net/JERA.15.1.

Mann, H., & Whitney, D. (1947). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Annals of Mathematical Statistics*, 18(1), 50–60.

Mello, C. D., Viola, M. R., Beskow, S., & Norton, L. D. (2013). Multivariate models for annual rainfall erosivity in Brazil. *Geoderma*, 202, 88–102. https://doi.org/10.1016/j.geoderma.2013.03.009.

Młyński, D., Petroselli, A., & Wałęga, A. (2018). Flood Frequency Analysis by an Event-based Rainfall-Runoff Model in Selected Catchments of Southern Poland. *Soil & Water Research*, 13(3). https://doi.org/10.17221/153/2017-SWR.

Ng, J. L., Yap, S. Y., Huang, Y. F., Noh, N. M., Al-Mansob, R. A., & Razman, R. (2020, April). Investigation of the best fit probability distribution for annual maximum rainfall in Kelantan River Basin. In *IOP Conference Series*: *Earth and Environmental Science* (Vol. 476, No. 1, p. 012118). IOP Publishing. https://doi.org/10.1088/1755-1315/476/1/012118.

Salam, M., Alam, F., Hossain, M. N., Saeed, M. A., Khan, T., Zarin, K., ... & Khan, O. (2021). Assessing the drinking water quality of educational institutions at selected locations of district Swat, Pakistan. *Environmental Earth Sciences*, 80, 1–11. https://doi.org/10.1007/s12665-021-09595-6.

Salam, M., Shahzadi, A., Zheng, H., Alam, F., Nabi, G., Dezhi, S., ... & Bilal, M. (2022). Effect of different environmental conditions on the growth and development of Black Soldier Fly Larvae and its utilization in solid waste management and pollution mitigation. *Environmental Technology & Innovation*, 28, 102649. https://doi.org/10.1016/j.eti.2022.102649.

Santos, G. G., Figueiredo, C. C. D., de Oliveira, L. F., & Griebeler, N. P. (2009). Intensidade-duração-frequência de chuvas para o Estado de Mato Grosso do Sul. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 13, 899–905. https://doi.org/10.1590/S1415-43662009000700012.

Senapeng, P., & Busababodhin, P. (2017). Modeling of maximum temperature in Northeast Thailand. *Burapha Science Journal*, 92–107.

Shahzadi, A., Akhter, A. S., & Saf, B. (2013). Regional frequency analysis of annual maximum rainfall in monsoon region of Pakistan using L-moments. *Pakistan Journal of Statistics and Operation Research*, 9(1), 111–136. https://doi.org/10.18187/pjsor.v9i1.461.

Villarini, G. (2012). Analyses of annual and seasonal maximum daily rainfall accumulations for Ukraine, Moldova, and Romania. *International Journal of Climatology*, 32(14), 2213–2226. https://doi.org/10.1002/joc.3394.

Wagesho, N., & Claire, M. (2016). Analysis of rainfall intensity-duration-frequency relationship for Rwanda. *Journal of Water Resource and Protection*, 8(07), 706. https://doi.org/10.4236/jwarp.2016.87058.

Wald, A., & Wolfowitz, J. (1943). An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics*, 14(4), 378–388.

Wdowikowski, M., Kaźmierczak, B., & Ledvinka, O. (2016). Maximum daily rainfall analysis at selected meteorological stations in the upper Lusatian Neisse River basin. Meteorology Hydrology and Water Management. *Research and Operational Applications*, 4(1), 53–63.

Yuan, J., Emura, K., Farnham, C., & Alam, M. A. (2018). Frequency analysis of annual maximum hourly precipitation and determination of best fit probability distribution for regions in Japan. *Urban Climate*, 24, 276–286. https://doi.org/10.1016/j.uclim.2017.07.008.

Yue, S., & Hashino, M. (2007). Probability distribution of annual, seasonal and monthly precipitation in Japan. *Hydrological Sciences Journal*, 52(5), 863–877. https://doi.org/10.1623/hysj.52.5.863.

Zakaria, Z. A., Shabri, A., & Ahmad, U. N. (2012). Regional frequency analysis of extreme rainfalls in the west coast of Peninsular Malaysia using partial L-moments. *Water Resources Management*, 26, 4417–4433. https://doi.org/10.1007/s11269-012-0152-8.