AT-SITE RAINFALL FREQUENCY ANALYSIS USING PARTIAL DURATION SERIES AND ANNUAL MAXIMUM SERIES: A CASE STUDY

AHMAD, I.[1,2*] – KHAN, D. A.[2] – ALMANJAHIE, IBRAHIM M.[1] – CHIKR-ELMEZOUAR, Z.[1] – LAKSACI, A.[1]

[1]Department of Mathematics, College of Science, King Khalid University, 61413 Abha, Kingdom of Saudi Arabia
(phone: +966-17-241-7734, fax: +966-17-241-7637)

[2]Department of Mathematics and Statistics, Faculty of Basic and Applied Sciences, International Islamic University, 44000 Islamabad, Pakistan

*Corresponding author
e-mail: ishfaq.ahmad@iiu.edu.pk; phone: +966-59-681-7557

Abstract. At-site Rainfall Frequency Analysis (RFA) is a crucial tool for designing of water related infrastructures. Partial Duration Series (PDS) and Annual Maximum Series (AMS) are the most popular techniques for RFA. PDS is capable of including more extreme events as compared to the AMS. Keeping in view the importance of at site RFA, in this study we identify suitable statistical model that best represent PDS and AMS extracted from daily rainfall records of 25 years in the vastly growing city, Rawalpindi, Pakistan. The most commonly used statistical distributions in RFA such as Generalized Pareto (GPA), Generalized Extreme Value (GEV), Generalized Logistic (GLO), and Pearson Type-3 (PE-3) distributions are used in this study. The parameters of these distributions are estimated by method of linear moments (LM). Anderson-Darling (A-D) testing criterion along with L-moment ratio diagram have been used to determine the best fit probability distribution. The findings of our study suggest that GEV as the best fit statistical distribution for PDS while GLO for AMS. Moreover, in this study we have also estimated quantiles for the best fitted probability distributions. The results revealed that PDS sample outperforms than AMS for various return periods. These estimates can be used in designing of water-related infrastructures such as culverts, bridges and other hydrological structures in the city.
Keywords: *frequency factor, goodness-fit test, linear moments, quintile estimation*

Introduction

Nowadays, climate change is one of the main challenge for mankind. Global warming occurs due to multiple changes in the climate such as changes in the intensity, duration, frequency and timing of extreme events including bolstering lengthening droughts, heat waves and occurrence of more precipitation. Like other countries worldwide, in Pakistan also, due to abrupt changes in the climate, several extreme changes in the weather have been observed in the form extreme rainfalls, droughts, irregular floods, glacier melting, sea-level rising, etc. (Rasul et al., 2012). In Pakistan, global warming has also putting severe impacts on hydrological parameters leading to cause an increase in precipitation for arid region including Rawalpindi city of Pakistan (Hassan et al., 2017). These heavy rains due insufficient water resources management cause flash floods in the city. Resulting disasters damage the economy of the country and create problems for citizens. In worst circumstances these disasters cause disruption of normal services like health care livelihoods and also loss of human lives. Reduction in flood risk, the mitigation of resulting damages is possible by human actions on the fluvial system, such as the

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8352 -

construction of reservoirs, culverts and other protection measures. Rainfall Frequency analysis using different probability models is one of the randomising approaches which could be used to estimates future rainfall with certain probability in order to mitigate the losses mentioned above. However, relatively small samples of annual maximum rainfall series put some limitations to the degree of statistical sophistication that can be engaged during at-site frequency analysis. The data of extreme events must reflect the variations expected in the quantity of interest. However, the values of many hydrological and hydrometeorological data vary in time giving alternating sequences of large and small values. So, usually, the larger values are greater than the mean values and make clusters in short time period followed by smaller values or null values sometimes. In the result of such temporal pattern, we get the typical configuration of a succession of exceedances over a specified threshold. The number and magnitude of such exceedances can be modeled by probability modeling. Such series are known as partial duration series (Naghettini, 2016).

For this study we have carried out RFA using two commonly known methods i.e. Annual Maximum Series (AMS) and Partial Duration Series (PDS) for the daily rainfall records of the city Rawalpindi, Pakistan. Rawalpindi is located in the north-south of the sub-mountainous region of the province Punjab. In recent decades, abrupt fluctuations in rainfall patterns are being observed particularly in northern Photohar regions. In the past years due to heavy rainfalls i.e., 1981, 1988, 1997, 2001, several tributaries have been affected badly in the city. A total of 11 major small tributaries of rain water pass through the city including the famous Lai-Nullah. The intensity amount of rainfall caused the water level of Lai-Nullah (the famous rainfall tributary in the city) particularly, other tributaries to rise remarkably and gave raise to heavy floods in the city. In the result of these floods, 74 human lives were lost, around 400,000 people were affected, 742 cattle heads were perished, 1,087 houses were completely 2,448 partially were demolished. The flash flood which drove by the monsoon rain in 2001 turn into a loss of US$ 250 million (Rasul et al., 2004). There are many factors causing this situation in these tributaries such as the illegal constructions along with these tributaries has squeezed the space for smooth flow of sewage rainwater collection of debris and also other construction materials being dumped in these rain water tributaries resulting in the rise of rainfall water to a precarious level. The accurate modeling in the form of at-site frequency analysis of these heavy rainfalls is of immense importance to mitigate future losses. Rainfall frequency analysis (RFA) with the passage of time is getting more attention worldwide. Extensive research works are being carried out in various parts of the world. According to Lang et al. (1999) two techniques, PDS and AMS are commonly used techniques in rainfall frequency analysis. AMS consists of single value, the maximum value per year of the daily records while PDS consists of all well-defined maximum peaks above a certain threshold level (Beguería, 2005). The dilemma in RFA is to select the proper series among AMS and PDS. The most regular objection with respect to use of annual maximum series is that it uses only one peak value in a year. On some occasions, the second largest event in a data may be greater than many AM events of the other years. On contrary, PDS modeling represents a complete description of extreme events (heavy rainfalls or floods), particularly when the issue of short length of data is under consideration (Lang et al., 1999). Fischer and Schuman (2014) explained that PDS modeling provide comparatively more data series per year as compared to AMS. The substitute makes possible to analyze events which are extreme still higher than annual maxima of others years and are not incorporated in the AMS technique. Lang et al. (1999) proposed to consider PDS modeling as a compromise between AMS analysis and classical

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8353 -

time series modeling. A lot of studies are available around the globe address different aspects of PDS and AMS modeling techniques (see for example Nagy et al., 2017), where they discussed the technical questions concerning the use of the partial duration series (PDS) within the domain of flood frequency analysis while the study of Chang et al. (2015) presented a comparison of AMS and PDS for derivation of rainfall intensity-duration-frequency relationships in peninsular Malaysia. Berton et al. (2018) discussed the Improving peak flow estimation across the United States by using PDS. Agilan et al. (2017) proposed non-stationary rainfall intensity-duration-frequency relationship with a comparison between AMS and PDS. Gado and Nguyen (2016) presented regional estimation of floods for ungauged sites using scaling approach partial duration series. Vrban et al. (2018) evaluated the Stormwater Infrastructure using PDS and AMS Models. Guru and Ramakar (2016) presented a study on flood estimation in Mahanadi river system, India using PDS. Olsson (2019) debated the short-duration rainfall extremes in Sweden with a regional analysis paradigm. Pham et al. (2013) discussed the statistical properties of PDS with a case study of North New Zealand. Alahmadi et al. (2014) proposed the best fit distribution for PDS of daily rainfall in Madinah, western Saudi Arabia. Karim et al. (2017) evaluated the AMS and PDS for estimating frequency of small floods for 24 gauging stations in the Great Barrier Reef (GBR) lagoon catchments in north-eastern Australia. Keeping in view the importance of PDS frequency analysis, we have compared the performance of PDS frequency analysis with AMS frequency for Rawalpindi city, Pakistan.

Objectives of the study

- To provide comparative at-site rainfall frequency analysis of Rawalpindi city of Pakistan using PDS and AMS modeling techniques.
- To determine the best fit probability distribution for PDS and AMS and estimate their parameters using robust estimation method of L-moments
- Finally, to estimate the quantiles of the best fit probability distribution under PDS and AMS paradigms.

Materials and methods

Data description

The daily rainfall data for length of 25 years of Rawalpindi City, Pakistan is being retrieved from Pakistan Meteorology department. The data is measured in millimeters (mm). Each normal year contains 365 values of daily rainfall while a leap year contains 366 values. We have total of 9131 values with minimum value of 0 mm and maximum of 592 mm. There are total 6 leap years during the study period from 1990 to 2014. Each value has been recorded during 24 h in the city. The average rainfall during 25 years is 53.52 mm with a standard deviation of 24.041 mm. The distribution of daily rainfall is positively skewed with sharp peak. The basic statistics of the study sample is given in *Table 1*.

Table 1. Summary statistics of basic data for the time period from 1990 to 2014

Number of values	Minimum value (mm)	Maximum value (mm)	Mean value (mm)	Standard deviation (mm)	Coefficient of skewness	Coefficient of kurtosis
9131	0	592	53.52	24.041	13.137	83.780

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8354 -

Preliminary trend analysis

Preliminary trend analysis is also very important to know about the nature of the data. In order to identify the random behaviour of the time series test of serial correlation could be applied. This test must be applied before applying Mann-Kedall's (MK) test. MK test is normally applied for serially uncorrelated series. If a series is found to be serially correlated, we must do prewhitening to make it serially uncorrelated. The test statistic is given in *Equation 1* and its confidence interval is given in *Equation 2*.

$$R = \frac{1/(n-1)\sum_{t=1}^{n-1}(y_t - \bar{y}_t)(y_{t+1} - \bar{y}_t)}{1/n\sum_{t=1}^{n-1}(y_t - \bar{y}_t)} \qquad (Eq.1)$$

$$\frac{\{-1 - 1.645\sqrt{n} - 2\}}{n-1} \leq R \leq \frac{\{-1 + 1.645\sqrt{n} - 2\}}{n-1} \qquad (Eq.2)$$

where R shows lag-1 correlation while $\bar{y}_t$ represents the average of the time series, n is the number of data points in the time series. The general rule is that if lag-1 correlation co-efficient R is found to be present in the interval as shown in *Equation 2*, we say that there is no significant correlation and declare that the desired series is free from serial correlation at given level of significance otherwise not. The Mann- Kendall's Test, also known as Kendall's tau statistic denoted by "S" by *Equation 3* is applied to find significance and direction of the trend. The null hypothesis will be that series at h with independent and identically distributed observations with two side alternative hypothesis. The constraints for the test are that k j ≤ n j ≠ k.

The test statistic "S" is defined as

$$S = \sum_{k=1}^{n-1}\sum_{j=k+1}^{n} \text{sgn}(y_j - y_k), \qquad (Eq.3)$$

where

$$\text{sgn}(y_j - y_k) = \begin{cases} +1 \ if \ (y_j - y_k) > 0 \\ 0 \ if \ (y_j - y_k) = 0 \\ -1 \ if \ (y_j - y_k) < 0 \end{cases}$$

where y_k y_j represent the rank of the data. For the time series with mean of 0 and identical distribution, variance of S statistic can be calculated as in *Equation 4* (Adamowski and Bougadis, 2003):

$$V(S) = \frac{\left\{n(n-1)(2n+5) - \sum_{h=1}^{n} t_h (h)(h-1)(2h+5)\right\}}{18} \qquad (Eq.4)$$

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8355 -

where t_h shows the summation of t, equal to number of tied values to the extent of c. For sample size more than 10, the above statistic follows normal distribution, as given in *Equation 5*. We can calculate the Z statistic as

$$z = \begin{cases} \dfrac{s-1}{\sqrt{\text{var}(s)}} \, if \ s \geq 0 \\ 0 \, if \ s = 0 \\ \dfrac{s+1}{\sqrt{\text{var}(s)}} \, if \ s \leq 0 \end{cases},$$ (Eq.5)

After calculating the above statistic, we can compare it with standard normal variate at specified level of significance. The S statistic not only determine the significance of the test but also the direction of the trend. For example, a positive value shows an upward trend while a negative value shows downwards trend in the data. The exact value of slope in the time series could be determined with the help of Sen's estimator, with an assumption that there is a linear trend in the data. It is also a nonparametric method used to find the change per unit time. To find the slope, first we calculate the statistic *Qi from M pairs* of the given time series as shown below in *Equation 6*:

$$Q_i = \frac{y_j - y_k}{j-k},$$ (Eq.6)

for i = 1, 2, 3,………..,M; where *"i"* indicates the *ith* pair of the data. The data point at *j*th time should be larger than the data point at *k*th time. The next step is finding the median value of these M values of *Qi*. This median value shows our required Sen's estimator. There are two cases: the first one is that if *M* is odd, the median value is given in *Equation 7*:

$$Q_{median} = Q_{\frac{m+1}{2}}$$ (Eq.7)

Second, if *M* is even, the median value is given in *Equation 8*:

$$Q_{median} = \frac{1}{2}\left[Q_{\frac{M}{2}} + Q_{\frac{M}{2}+1}\right],$$ (Eq.8)

Formation of annual maximum series (AMS) and partial duration series (PDS)

From the basic data of daily rainfall of 25 years we want to formulate two series, such as Partial Duration Series (PDS) and Annual Maximum Series (AMS) for further analysis. The AMS consists of single value, the maximum value in the whole year. It is largest value among 365 values while, PDS consists of all well-defined maximum peaks above a certain threshold level (Beguería, 2005). Selection of threshold is one of the major complexities in the PDS modeling. It turns out that there is, in the state of the art, no universal and unambiguous rule for the selection of the threshold which incorporates as much information as desirable without compromising the independence requirement between occurrences over time. The selection of threshold has direct impact on the analysis. The problem of the selection of threshold is also discussed extensively by

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8356 -

Rosbjerg and Madsen (1992). They compared different methods to define the threshold level based on certain values of frequency factor *K*. This method presents a predefined frequency factor *K*, can be written as in *Equation 9*:

$$Threshold = E\{1\} + (K).S(1) \qquad \text{(Eq.9)}$$

where E {I} and S {I} are, the mean and the standard deviation of the basic data.

To enhance the usage of PDS modeling among the practitioners, it is required to develop a practice-oriented guidelines (Rasmussen et al., 1994). Meeting the independence condition for the set of selected peaks is a prerequisite in frequency analysis of PDS. In terms of computation, PDS modeling is more complicated than the AMS modeling. According to extreme value theory, along with fundamental assumptions of independence and identical distribution, PDS modeling also requires that number of peaks above threshold to follow Poisson process while the peaks themselves follow GPD (Cunnane, 1979). Once the series satisfies the basic assumptions, the next step is to determine the best fit probability distribution. Therefore, we study most commonly used distributions for frequency analysis in this study such as GLO, GEV, GPD, PE3.

Selection of estimation method

Linear moments (LM) estimation technique has been adopted in this study to estimate the parameters of the statistical models. In case of extreme values present in the data series LM method is advantageous over the conventional moments method (Ahmad et al., 2015). Estimates of the conventional moments methods are not robust in quality and do not provide reliable estimates for the probability distributions with three or more parameters. The method of maximum likelihood (MLE) is considered as one of the efficient parameters estimation approach. It gives the least variance of the estimated parameters but it suffers from serious disadvantage of giving biased estimates with small samples especially if the numbers of parameters are large (Ahmad et al., 2016a, b, 2017). The method of LM corresponding to probability weighted moments (PWMs) are variants of the method of moments which provide a different way to summarize the data set. LM are summary statistics for probability distributions and data sample. Measure of location, dispersion, skewness, kurtosis and other descriptive statistics can be obtained through LM. To compute LM, we consider linear combination of ordered statistics. For instance, $X_1, X_2, X_3, ... X_r$, *be* the random sample of magnitude r, with cumulative distribution function $F(X)$ and quantile function $X(F)$.

Let $X(1:r) \leq X(2:r) \leq ... \leq X(r:r)$ be the order statistics of a random sample. For random variable X, the rth population L-moment as explained by (Hosking, 1990) can be written as in *Equation 10*:

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} E(X_{r-k:r})$$
$$r = 1, 2, ... \qquad \text{(Eq.10)}$$

In *Equation 10* λ_r is a linear function of the expected order statistics. The L-moment ratios are also very important. These can be defined as in *Equations 11-13*:

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8357 -

$$\tau = \frac{\lambda_2}{\lambda_1} \qquad \text{(Eq.11)}$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \qquad \text{(Eq.12)}$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} \qquad \text{(Eq.13)}$$

where τ is a measure of L-coefficient of variation (L-Cv), τ_3 τ_4 are L-skewness and L-Kurtosis, respectively. In general sample linear moments are estimated utilizing the sample obtained from the large set of data. In this study we have also obtained the sample moments using the sample of the large set of daily rainfall records of 25 years of Rawalpindi city. The r_{th} sample L-moments as suggested by Hosking (1990) are given in *Equation 14:*

$$l_r = \frac{1}{r} r \sum_{i=1}^{n} \left[\sum_{j=0}^{r-1} \frac{(-1)^j \binom{r-1}{j} \binom{i-1}{r-1-j} \binom{n-1}{j}}{\binom{n}{r}} \right] X_{i:n} \qquad \text{(Eq.14)}$$

$$r = 1, 2, 3, ...$$

The sample L-ratios are defined in *Equation 15:*

$$t = \frac{\ell_2}{\ell_1}, \ t_3 = \frac{\ell_3}{\ell_2} \ \ t_4 = \frac{\ell_4}{\ell_2} \qquad \text{(Eq.15)}$$

where $\ell 1$ is sample measure of location, *"t"* is sample measure of CV, t_3 t_4 are the sample skewness and Kurtosis based on LM respectively. These measures are used to identify the distribution from which the sample is drawn. Additionally, these measures are used in the estimation of parameters when distribution is fitted to a sample by equating the sample L-moments to corresponding population L-moments.

Goodness of fit criteria

Another fundamental step in rainfall frequency analysis after selection of efficient robust estimation method is the selection of best fit statistical models. For the selection of best fit statistical model, we have standard criteria, i.e. Akaik Information Criterion (AIC) and Bayesian Information Criterion (BIC), but do not work efficiently in the presence of small sample data. These criteria fail to produce efficient results with asymmetric probability distribution that commonly encountered in extreme events applications. Therefore, in this study we have used Anderson - Darling (A-D) test to verify whether or not a sample belongs to specific population. This test statistics is a modified form of the Kolmogorov Smirknove (K-S) test. This testing criterion is considered to be better than K-S testing criterion as it gives more weight to the tails in the highly skewed distributions. This test also considers specific distribution in the calculation process of critical values. Baldassarre and Montanari (2009) suggested using

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8358 -

A-D testing criterion as a goodness of fit test for small sample data. Similar suggestions are found in Onoz and Bayazit (1995). The mathematical form of A-D testing criterion can be defined by the formula given in *Equation 16:*

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

(Eq.16)

In *Equation 16* "n" is the size of the sample, $X_{i,s}$ are the observed data values arranged in an ascending order $F_{(X)}$, is the expected distribution function. The larger value of the test statistic will show more discrepancy between expected and empirical distribution function, hence would lead us to reject the null hypothesis. L-moment ratio diagram has also been adopted as an expedient tool for the selection of candidate distribution (Ahmad et al., 2015). These diagrams represent a theoretical relationship between the L-skewness τ_3 and the L-kurtosis τ_4 of different probability distribution. The sample ratio of L-skewness and L-kurtosis goes near to the most appropriate distribution in the diagram.

Basic assumptions of the data

PDS and AMS analysis requires that the data must fulfill the fundamental assumptions of statistical hydrology such as randomness, independence, homogeneity and stationarity. Assumption of randomness uncovers, either the given sample is drawn at random from the population. If the data fulfills the assumption, it implies that the sample values are the realization of a random variable (Naghettini, 2016). NERC (1975) proposed a nonparametric test for checking randomness by considering the number of turning points present in the sample for a given time period. A turning point may be defined as a peak or a trough in a time series plot of the data. Underlying intuitive idea behind the test is that if there are too many or few turning points in the data, then sample would be declared as possibly non-random. The numbers of turning points "p" is a purely random process of N observations with mean variance mentioned in *Equations 17* and *18:*

$$E(p) = \frac{2(N-2)}{3}$$

(Eq.17)

$$Var(p) = \frac{16N-29}{90}$$

(Eq.18)

This approximation is available in literature for large sample (Yule, 1950). It has been proposed that the resulting variable "t" follows normal distribution for larger size. Therefore, under the null-hypothesis of randomness the standardized test statistic, $(t = \frac{p-E(p)}{\sqrt{var(p)}})$, follows normal distribution.

For fulfillment of independence assumption, we used Wald and Wolfowitz test. The assumption of independence implies that not a single sample value could affect the occurrence or non-occurrence of any other sample value present in the data. Generally, the AMS is expected to follow the assumption of independence, while strong dependence is expected for PDS. In any case, the assumption of independence must be checked before further analysis. Wald and Wolfowitz (1943) suggested a non-parametric test for this purpose. Let $\{x_1, x_2, ..., x_n\}$ describe a sample of size N from X,

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8359 -

while $\{x_1', x_2', \ldots, x_N'\}$ is the sequence of differences of any ith sample value from the sample mean $\bar{x}$.

The test statistic is calculated using R, E (R) $\widetilde{Sr}$ given in *Equations 19-21:*

$$R = \sum_{I=1}^{N-1} x_i' x_{i+1}' + x_1' x_N',$$ (Eq.19)

$$E(R) = \frac{s_2}{N-1} \quad Var(R) = \frac{s_2^2 - s_4}{N-1} + \frac{s_2^2 - 2s_4}{(N-1)(N-2)} - \frac{s_2^2}{(N-1)^2}$$ (Eq.20)

where:

$$\widehat{s_r} = \sum_{i=1}^{N} (x_i')^r$$ (Eq.21)

For large sample under true null hypothesis of independence, Wald and Wolfowitz (1943) found that this test statistic follows a normal distribution as mentioned in *Equation 22*.

$$t = \frac{R - E(R)}{\sqrt{var(R)}}$$ (Eq. 22)

The assumption of homogeneity implies that all elements of the sample data belong to a single common population. However, considering mean values or total of values makes easier to detect the heterogeneities. The assumption of homogeneity is generally checked through a non-parametric test developed by Mann-Whitney (1947). For a given sample $\{x_1, x_2, x_3, \ldots, x_n\}$ of size N, one has to split the data in two sub samples of size N_1 $\{x_1, x_2, x_3, \ldots, x_{N1}\}$, N_2 $\{x_1, x_2, x_3, \ldots, x_{N2}\}$ sizes such that $N_1 + N_2 = N$ (with approximately equal subsamples). Once the splitting of the data is done the second step is to rank the entire sample in ascending order by noting the rank order say (m) of each element also from which subsample belongs to. The basic idea behind the test is that if the given sample is not homogenous, the sum of ranking orders of the elements belonging to first subsample would be consistently higher or lower than the sum of ranking orders of elements of the second subsample (Naghettini, 2016).

The Mann–Whitney test statistic can be written in the form mentioned in *Equations 23* and *24*.

$$V_1 = N_1 N_2 + \frac{N_1 (N_1 + 1)}{2} - R_1$$ (Eq.23)

$$V_2 = N_1 N_2 - V_1$$ (Eq.24)

where R_1 indicates the sum of ranks of all elements from first subsample. If both of the subsamples are larger than 20, the null hypothesis is true such that the given sample is homogenous, Mann-Whitney (1947) showed that V follows a normal distribution with mean: $E[V] = \frac{N_1 N_2}{2}$; variance: $VAR(V) = \frac{N_1 N_2 (N_1 + N_2 + 1)}{N_2}$; Hence, the require statistic for the Mann–Whitney test will be written as: $t = \frac{V - E(V)}{\sqrt{var(V)}}$, which follows a standard normal distribution.

The notion of stationarity implies that some basic properties of the observed sample as well as relevant functional form of the probability distribution and its parameters do not change with respect to time. Non-stationarity might be in form of monotonic trend

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8360 -

as well as abrupt change in the data at some time interval. It is to be noted that a non-stationary series is always non-homogenous but non- homogenous series is not necessarily stationary (Naghettini, 2016).

In order to verify the assumption of stationarity, a non-parametric test named Spearman's rank correlation coefficient is used. It is a statistical dependence between two variables, one is the ordered ranks say (m_t) of the observed sample in ascending order while the other one is the time indices $T_t = 1, 2, \ldots . N$. It reveals the monotonic trend which is not restricted to only linear as in case of conventional correlation coefficient. It also depicts the nonlinear correlation between the ranks' orders say (m_t) of the observed sample and the time indices $T_t = 1, 2, \ldots . N$. It is written as in the following form given in *Equation 25*.

$$r_s = 1 - \frac{6 \sum_{i=1}^{N} (m_t - T_t)^2}{N^3 - N} \tag{Eq.25}$$

The corresponding test statistic is written in *Equation 26:*

$$t = r_s \sqrt{\frac{N-2}{1 - r_s^2}} \tag{Eq.26}$$

Assumptions of partial duration series analysis

For PDS sample, it is required that it must fulfills the assumption of Poison process. Bogeria (2005) discovered that the original daily rainfall series did not adjust to a Poisson process at lower threshold level, however the study showed that the tendency towards the Poisson process was increased as the level of threshold increases. For at-site frequency analysis, Cunnane (1979) proposed Fisher testing criterion given in *Equation 28* to test Poisson process states that the number of peaks values exceeding certain threshold level must conforms to a Poisson Process with mean such as *Equation 27*.

$$p = \frac{e^{-\mu} \mu^x}{x!} \tag{Eq.27}$$

A test of Poisson assumption can be piloted on observed series of peaks exceeding the certain threshold level. Fisher test statistics is based on the equation given in *Equation 28*. This statistics is utilized for the verification of PD series.

$$R = \sum \frac{(ni - \lambda)^2}{\lambda} \tag{Eq.28}$$

In the equation above ni represents the number of peaks above at certain threshold in a year, λ is the mean number of peaks exceeding certain threshold level. In the literature this testing criterion has been used for the verification of Poisson assumption. Under the true null hypothesis, statistics R is distributed as chi-square distribution with the confidence limits as $\chi^2_{\alpha/2}(t-1) < R < \chi^2_{1-\alpha/2}(t-1)$. These confidence limits verify either data series to follow the Poisson distribution. If, test statistics R lies within the upper and lower confidence limits then the PDS will be considered to follow Poisson process. If the classical PDS model comprises the assumption of Poisson process then

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8361 -

the exceedances may follow exponential distribution. However, in recent years numerous researches have emphasized that the peaks in PDS modeling follow GPD (see for example Hosking Wallis, 1987; Madsen and Rosbjerg, 1994). Few other research studies have proved that peaks in PDS may also follow distributions other than GPD (for example see Rasmussen et al., 1991).

Results and discussion

Preliminary trend analysis is also very important to know about the nature of the data. In this study, we have carried out the trend analysis using monthly and overall precipitation series of 25 years of data of Rawalpindi city from 1990 to 2014. There are total thirteen variables including one for overall daily rainfall (raw) data and remaining twelve for monthly data. We implemented two non-parametric methods for trend analysis. In other words, Mann-Kendall (MK) test procedure is used to find the significance of the trend while Sen's slope estimator to find change per unit time. The MK test will help to find not only the significance of the trend but also direction of the trend with an assumption of linear trend. Before applying Mann-Kendall's test for significance of the trend, we also implemented lag-1 serial correlation test to determine either we need some prewhitening of the data or not. The data of Rawalpindi station on monthly basis as well as overall data did not show any lag-1 serial correlation among the successive values. This implies that we do not need any prewhitening of given time series. The given series can be used for further trend significance. The bold value in *Table 2* shows the presence of significant trends. All months shows insignificant trends except the month of March. Three months such as Month of January, June and November show negative insignificant trends. Overall data also indicate that the trend is insignificant positive trend. Further, we can also notice that the trend in monsoon season (July to September) is positive insignificant.

Table 2. Trend analysis in monthly rainfall and daily rainfall data from 1990 to 2014

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Overall
-0.01	0.047	**0.127**	0.081	0.05	-0.022	0.014	0.017	0.022	0.018	-0.03	0.005	0.021

Formation of partial duration (PDS) and annual maximum series (AMS)

For AMS frequency analysis, maximum value per year is considered. For the length of 25 years data we have only 25 values, whereas, in PDS, exceedances above certain threshold level are considered. For PDS different samples were selected utilizing different values of *K* (frequency factor) in objective formula.

The size of the sample in PDS is associated with the value of frequency factor (*K*). A slight increase in the value of K reduces the sample to a large extent as shown in *Table 3*. A sample of size 126 was obtained using 3.5 as frequency factor. Similarly, the sample of size 96, 58 were obtained using 4 and 5 as K values in the objective formula. For PDS sample to be used for further analysis is necessary to follow the assumption of Poisson process. Therefore, all the samples of PDS given in *Table 3* were tested for Poisson process. The results are given in *Table 4* based on 99% and 95% confidence interval.

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8362 -

Table 3. *Frequency factor and number of exceedances using objective formulation*

Frequency factor (K)	PDS sample size
K = 3.5	126
K = 4	96
K = 5	58

Table 4. *The assumption of Poisson distribution for partial duration series*

K	R	C.I (1%)	Decision	C.I (5%)	Decision
3.5	12.49	9.02 < R < 45.559	Accepted	12.401 < R < 39.36	Accepted
4.5	5.82	9.02 < R < 45.559	Rejected	12.401 < R < 39.36	Rejected

Table 4 shows that a sample of size 126 was obtained using value of 3.5 for *K* in the objective formula which fulfills the assumption of Poisson process both at 99% confidence interval 95% confidence interval. The value of the statistic falls in the confidence interval which supports that the PDS sample from this frequency factor is following Poisson process assumption. Whereas, rest of the sample do not fulfills the assumption of Poisson process. For further rainfall frequency analysis we will consider only the sample with frequency factor = 3.5. For AMS analysis we have 25 values while for PDS analysis we have 126 values for further analysis.

Fundamental assumptions are essential before carrying out formal RFA. Therefore, we have tested both samples, i.e. PDS and AMS for randomness, Independence, homogeneity and stationarity. *Table 5* presents that the values of test statistics *t* for different assumptions at 5% level of significance. The calculated values of the said test statistic are less than the critical value of 1.96, therefore it can be suggested that both series are random, independent, homogenous and stationary. Thus it is concluded that both series are suitable for further RFA.

Table 5. *Results of fundamental assumptions for PDS AMS*

Series	Randomness*	Independence*	Homogeneity*	Stationarity*
PDS	$\hat{t} = 1.80$; Since $\lvert\hat{t}\rvert < Z_{1-\frac{\alpha}{2}} = 1.96$ Do not reject H_0	$\hat{t} = 1.48$; Since $\lvert\hat{t}\rvert < Z_{1-\frac{\alpha}{2}} = 1.96$ Do not reject H_0	$\hat{t} = -0.153$; Since $\lvert\hat{t}\rvert < Z_{1-\frac{\alpha}{2}} = 1.96$ Do not reject H_0	$\hat{t} = -0.3279$; Since $\lvert\hat{t}\rvert < t_{1-\alpha/2,(N-2)} = 0.687$ Do not reject H_0
AMS	$\hat{t} = 1.29$; Since $\lvert\hat{t}\rvert < Z_{1-\frac{\alpha}{2}} = 1.96$ Do not reject H_0	$\hat{t} = -0.412$; Since $\lvert\hat{t}\rvert < Z_{1-\frac{\alpha}{2}} = 1.96$ Do not reject H_0	$\hat{t} = -1.22$; Since $\lvert\hat{t}\rvert < Z_{1-\frac{\alpha}{2}} = 1.96$ Do not reject H_0	$\hat{t} = 1.419$; Since $\lvert\hat{t}\rvert < t_{1-\alpha/2,(N-2)} = 1.66$ Do not reject H_0

*5% level of significance

To find the best distribution for both samples, we considered most commonly used distributions for at-site frequency analysis. The goodness of fit of these distributions is evaluated using the most common goodness of fit criteria, i.e., AD and L-moment ratio diagram. Based on the AD testing criterion and the L-moment ratio diagram, the best fitted probability distribution for PDS sample was the GEV while for the AMS sample the GLO distribution. In case of AD testing criterion, the smallest test statistics values

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8363 -

would be considered as the best fitted and will be ranked one. As shown in *Table 6,* the best fitted probability distribution is GEV that ranked one in case of PDS sample whereas, GLO in case of AMS. The ratio of sample L-skewness and L-Kurtosis is shown with symbol "+". The closeness of symbol "+" to the curve in the diagram declares the distribution as best fitted for sample at hand. It is apparent from *Figures 1* and *2* that GEV distribution is the best fitted for PDS while GLO for AMS. The outputs of the graphs validated the mathematical results obtained from AD testing criterion.

Table 6. Determination of best fit probability distribution for PDS AMS

Distributions	PD Series (n = 126)		AM series (n = 25)	
	AD	Rank	AD	Rank
GEV	0.29438	1st	0.49	2nd
PE-3	5.5266750	4th	6.11	4th
GLO	3.57	2nd	0.47	1st
GPD	3.646	3rd	2.92	3rd

The next step in RFA after declaration of the best fitted probability distribution is the quantiles estimation. Quantile estimation provides estimates to ensure a resilience hydrological structure. Therefore, we constructed a *Table 7* that consists of quantiles estimates for various return periods. These estimates are computed for PDS and AMS samples separately and further their standard errors are given in parenthesis. These standard errors will assure the robustness and efficiency of PDS and AMS samples. The relationship between return periods and occurrence of an extreme event is established through notion of a geometric random variable (waiting time until an extreme event occurs) using mathematical expression: $E(X) = T = 1/p$, which is in fact the mean of the geometric random variable (average time for occurrence of an extreme event, for example exceeding of water flow above certain level of a hydrological structure), where *p* is the probability of occurrence or exceedance for T year return period. Suppose we have a return period of ten-years, the corresponding probability of exceedance is calculated using above equation as 0.1 and the probability of non-exceedance as 0.9. The corresponding quantile estimates based on the best fitted distribution (GEV for PDS) is 129.64 mm, which implies that in next ten-year, the daily rainfall in the city would exceed 129.64 mm once on the average with a probability of 0.1. These quantiles could be used for planning of water-related emergencies, sustainable water resource management, construction of better flood design estimates in the form of culverts, bridges other hydrological structures in the city. The quantile estimation for AMS series showed significant variation in magnitude as compared to PDS for various return periods. Further, to identify the best technique between AMS and PDS, we calculated the standard errors of these quantiles estimates at different return periods. These standard errors are given in parenthesis as shown in *Table 7*. The standard errors of PDS sample are significantly smaller than AMS sample. It implies that PDS sample outperforms than AMS sample at different return periods which strengthens the use of PDS in the current study. Comparisons of AMS and PDS extreme events have also indicated that for all return periods PDS give more consistent, efficient and robust predictions of extreme rainfall than AMS. The predicted magnitude based on PDS are also more closer to the historical records as compared to AMS. The quantiles estimates

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8364 -

for different return periods are helpful to construct a variety of hydrological infrastructures such as bridges, culverts and others in the city

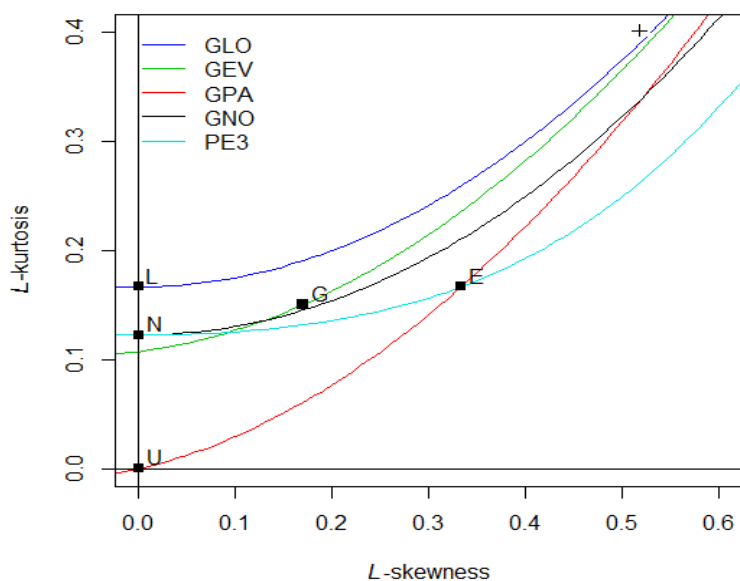
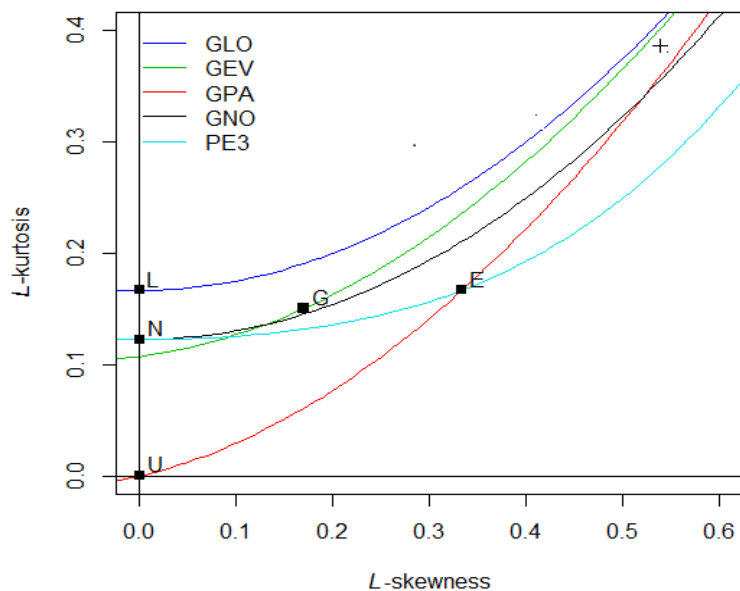*Figure 1. L moment ratio diagram for AMS sample*

Figure 2. L moment ratio diagram for PDS sample

Table 7. Quantile estimates for different return periods

	Best fitted distribution	**5**	**10**	**100**	**1000**
		**80%	**90%**	**99%**	**99.9%**
PDS	GEV	98.072 (0.346)	129.64 (0.245)	308.77(0.175)	720.52 (0.179)
AMS	GLO	172.0268 (0.742)	230.08 (0.641)	594.74 (0.442)	1591.07 (0.318)

**Probability of non-exceedances

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8365 -

Conclusions

In this study, we conducted rainfall frequency analysis in the fast growing city Rawalpindi, Pakistan by extracting two series such as PDS and AMS from daily rainfall data measured in millimeters for a period of 25 years (1990-2014). The selection of best sample between PDS and AMS had always been a dilemma in rainfall frequency analysis. This study also provided a comparative analysis of these two techniques. After preliminary trend analysis, for PDS sample extraction we used objective method. The size of the sample for PDS was higher than AMS. After extraction of the required samples, we implemented the most common probability distributions used in statistical hydrology to check the best fit. For this purpose two GOF criteria such as Anderson Darling (A-D) test and L-moment ratio diagram were implemented. The analysis revealed that GEV was best fit for PDS while GLO for AMS. In this study we also found quantiles estimates using the parameters of the best probability distributions for both series. The predicted values based on quantile estimation for AMS series showed significant variation in magnitude as compared to PDS for various return periods. The standard errors of quantile estimates based on these techniques revealed that PDS was better than AMS. PDS would be a more reasonable and effective choice for such type of frequency analysis. Our results are also analogous to the outcomes of earlier studies (e.g., Fischer and Schuman, 2014; Alahmadi et al., 2014; Karim et al., 2017) where PDS technique outperformed than AMS.

Thus we can conclude that the results obtained using PDS sample were generally preferable for at-site quantile estimation. The findings from this study could be implemented in selection of better design criteria for flood management, particularly flood protection measures in the city. This study could also be extended for rainfall frequency analysis in the other areas of Pakistan with similar geomorphologic hydrological characteristics.

Conflict of interests. Authors have no conflict of interests on publication of this paper in Applied Ecology and Environmental Research (AEER) journal.

REFERENCES

[1]   Adamowski, K., Bougadis, J. (2003): Detection of trends in annual extreme rainfall. – Hydrological Processes 17(18): 3547-3560.
[2]   Agilan, V., Umamahesh, N. V. (2017): Non-stationary rainfall intensity-duration-frequency relationship: a comparison between annual maximum and partial duration series. – Water Resources Management 31(6): 1825-1841.
[3]   Ahmad, I., Fawad, M., Abbas, A., Saghir, A. (2016a): Probability modelling of low flows at different sites of Indus basin in Pakistan using L-moments and TL-moments. – Pakistan Journal of Science 68(1): 86-93.
[4]   Ahmad, I., Fawad, M., Mahmood, I. (2015): At-site flood frequency analysis of annual peak stream flows in Pakistan using robust estimation methods. – Polish Journal of Environmental Studies 24(6).

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8366 -

[5]   Ahmad, I., Fawad, M., Akbar, M., Abbas, A., Zafar, H. (2016b): Regional frequency analysis of annual peak flows in Pakistan using linear combination of order statistics. – Polish Journal of Environmental Studies 25(6): 1-10.

[6]   Ahmad, I., Yasin, M., Fawad, M., Saghir, A. (2017): Regional frequency analysis of Low flows using L. moments for Indus Basin, in Pakistan. – Pakistan Journal of Science 69(1): 75-83.

[7]   Alahmadi, F., Rahman, N. A., Abdulrazzak, M. (2014): Evaluation of the best fit distribution for partial duration series of daily rainfall in Madinah, western Saudi Arabia. – Proceedings of the International Association of Hydrological Sciences 364: 159-163.

[8]   Baldassarre, G. D., Montanari, A. (2009): Uncertainty in river discharge observations: a quantitative analysis. – Hydrology Earth System Sciences 13(6): 913-921.

[9]   Beguería, S. (2005): Uncertainties in partial duration series modeling of extremes related to the choice of the threshold value. – Journal of Hydrology 303(1): 215-230.

[10]   Berton, R., Rahmani, V. (2018): Improving peak flow estimation across the United States by using partial duration series. –AGU Fall Meeting Abstracts #H41M-2279.

[11]   Chang, K. B., Lai, S. H., Othman, F. (2015): Comparison of annual maximum and partial duration series for derivation of rainfall intensity-duration-frequency relationships in peninsular Malaysia. – Journal of Hydrologic Engineering 21(1): 0501-513.

[12]   Cunnane, C. (1979): A note on the Poisson assumption in partial duration series models. – Water Resources Research 15(2): 489-494.

[13]   Fischer, S., Schumann, A. (2014): Comparison between Classical Annual Maxima and Peak over Threshold Approach Concerning Robustness. – Universitätsbibliothek Dortmund.

[14]   Gado, T. A., Nguyen, V. T. V. (2016): Regional estimation of floods for ungauged sites using partial duration series and scaling approach. – Journal of Hydrologic Engineering 21(12): 04016044.

[15]   Guru, N., Jha, R. (2016): Flood estimation in Mahanadi river system, India using partial duration series. – Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 10(2): 135-145.

[16]   Hassan, I., Ghumman, A. R., Ghazaw, Y., Abdel-Maguid, R. H., Samreen, B. (2017): Climate change impact on precipitation in arid areas of Pakistan. – Int. J. Water Resour. Arid Environ. 6: 80-88.

[17]   Hosking, J., Wallis, J. (1987): Parameter quantile estimation for the generalized pareto distribution. – Technimetrics 29(3): 339-349.

[18]   Hosking, J. R. M. (1990): L-moments: Analysis and estimation of distributions using linear combinations of order statistics. – Journal of the Royal Statistical Society: Series B (Methodological) 52(1): 105-124.

[19]   Karim, F., Hasan, M., Marvanek, S. (2017): Evaluating annual maximum and partial duration series for estimating frequency of small magnitude floods. – Water 9(7): 481-488.

[20]   Lang, M., Ouarda, T. B. M. J., Bobee, B. (1999): Towards operational guidelines for over-threshold modeling. – Journal of Hydrology 225(3): 103-117.

[21]   Madsen, H., Rosbjerg, D., Harremoës, P. (1994): PDS-modelling regional Bayesian estimation of extreme rainfalls. – Hydrology Research 25(4): 279-300.

[22]   Mann, H. B, Whitney, D. R (1947): On the test of whether one of two random variables is stochastically larger than the other. – Ann Math Stat 18: 50-60.

[23]   Naghettini, M. (2016): Fundamentals of Statistical Hydrology. – Springer, Berlin.

[24]   Nagy, B. K., Mohssen, M., Hughey, K. F. D. (2017): Flood frequency analysis for a braided river catchment in New Zealand: comparing annual maximum and partial duration series with varying record lengths. – Journal of Hydrology 547: 365-374.

[25]   NERC (1975): Flood Studies Report. Vol 1. – National Environmental Research Council, London.

Ahmad et al.: At-site rainfall frequency analysis of partial duration series and annual maximum series using L-moments in
Rawalpindi city of Pakistan
- 8367 -

[26] Olsson, J., Södling, J., Berg, P., Wern, L., Eronn, A. (2019): Short-duration rainfall extremes in Sweden: a regional analysis. – Hydrology Research. https://doi.org/10.2166/nh.2019.073.

[27] Önöz, B., Bayazit, M. (1995): Best-fit distributions of largest available flood samples. – Journal of Hydrology 167(1-4): 195-208.

[28] Pham, H. X., Shamseldin, A. Y., Melville, B. (2013): Statistical properties of partial duration series: Case study of North Island, New Zealand. – Journal of Hydrologic Engineering 19(4): 807-815.

[29] Rasmussen, P., Ashkar, F., Rosbjerg, D., Bobee, B. (1994): The POT Method for Flood Estimation: A Review. – In: Hiper, K. W. (ed.) Stochastic and Statistical Methods in Hydrology and Environmental Engineering. Springer, Dordrecht.

[30] Rasul, G., Sixiong, Z., Qingcun, Z. (2004): A diagnostic study of record heavy rain in twin cities Islāmābad-Rāwalpindi. – Advances in Atmospheric Sciences 21(6): 976-988.

[31] Rasul, G., Mahmood, A., Sadiq, A., Khan, S. I. (2012): Vulnerability of the Indus delta to climate change in Pakistan. – Pakistan Journal of Meteorology 8(16): 89-107.

[32] Rosbjerg, D., Madsen, H. (1992): On the choice of threshold level in partial duration series. – Nordic Hydrological Conference Alta, NHP-Report 30: 604-615.

[33] Vrban, S., Wang, Y., McBean, E. A., Binns, A., Gharabaghi, B. (2018): Evaluation of stormwater infrastructure design storms developed using partial duration and annual maximum series models. – Journal of Hydrologic Engineering 23(12).

[34] 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.

[35] Yule, G. Y., Kendall, M. G. (1950): An Introduction to the Theory of Statistics. – Charles Griffin and Co., London.