Regional Rainfall Frequency Analysis on Seven Sites of Punjab, Pakistan using L-moments

Uzma Nawaz¹, Zamir Hussain¹, Tooba Nihal¹, and Saira Usman¹
¹Department of Statistics, The Women University, Multan, Pakistan
*Corresponding Authors: uzma_69@yahoo.com

Abstract
The hydro-meteorological variables of extreme rainfall are not easy to explain due to unexpected changes in climate and varied usage of water with growing population. Regional rainfall frequency analysis is the one such method that is useful for the requirement of more accurate estimates of rainfall yearly or desineally for the regions having lack of fresh water resources. The series of Annual Maximum Monthly Rainfall Totals (AMMRT) has been used for the seven sites of northern Punjab, Pakistan using L-moments. The results of different test, the run test, lag-1 correlation and Mann-Whitney U test illustrate that the data series of the seven sites of northern Punjab were found random and independently and identically distributed and have no serial correlation. Heterogeneity measure exposed that the region is homogeneous and discordancy measure gives the evidence that no site is discordant among the seven. The result of goodness of fit test including L-moment Ratio diagrams, ZDIST statistic and Mean Absolute Deviation Index exposed the Pearson Type III (PE3), Generalized Normal (GNO) and Generalized Extreme Value (GEV) are best suitable of the regional distribution for the quantiles estimation. The quantiles estimates obtained for different return periods. A linear regression model was developed with good fit between the at site characteristics and the mean of the AMMRT of the sites. The estimates of the study may be used for the estimation of the rainfall quantiles of the seven sites for different return periods. The estimates will be useful to design future preventive measures for the harmful impact of hydro meteorological events at these sites in Punjab Pakistan.

Keywords: L-moments, Regional Rainfall Frequency Analysis, Discordancy Measures, Heterogeneity Measure, quantiles, Linear Model.

Introduction
Natural disaster like floods, rainstorms, droughts, earth quakes river discharge flow having severe influence on the human society. These are termed as civilization. The extreme events are not very the usual routine events. The occurrence is for a short time interval resulting often in short data records. Therefore the significance of these events lies infact in the frequency of its
occurrences. The estimation of such frequencies is another issue of not so easy paradigm. The causes of such events processes are tagged with numerous uncertainties. Therefore estimation purpose will be one that acknowledge the existence of uncertainty and enables its effect to be quantified. One such approach for the analysis of extreme data is Regional Frequency Analysis (RFA). The main objective of RFA is that the resulting estimates have been used successfully for the estimation of quantiles at ungauged sites.

The application of RFA completes in four steps.

i. Initial screening of the data series of AMMRT by using different statistical tests.
ii. Identification of a homogeneous region by using Discordancy Measure and Heterogeneity Measure based on L-moments of the seven sites.
iii. Choice of an appropriate frequency distribution using L-Moment Ratio Diagram and ZDIST criteria and Mean Absolute Deviation Index (MADI).
iv. Quantile estimation and Inference of suitable distribution for different return periods.

Data Collection

The rainfall series of seven sites of Punjab were collected from the Pakistan Meteorological Department (PMD) Karachi. The data in a study is a 36 year record of rainfall. The seven sites were one cluster as these were at higher elevation in the northern region of Punjab Pakistan. The occasional rainfall in Punjab in the monsoon period depends on the annual monsoon winds of the surroundings. The floods may happen due to continuous heavy rainfall in the monsoon season. Therefore, the monthly maximum magnitude of rainfall each year is valuable to provide information for the anticipation of natural tragedies. AMMRT is the variable of study for the RFA. The seven sites located geologically on the map of Punjab, Pakistan is presented in (Fig. 1).

![Map of Punjab](image)

Fig. 1. The Geology of the Seven Sites of Punjab under study

(a) The Run Test on the AMMRT data

| S. No. | Sites  | No. | ONR  | P-values |
|-------|--------|-----|------|----------|
| 1     | Faisalabad | 36  | 19   | 0.610    |
| 2     | Islamabad | 36  | 17   | 0.602    |
| 3     | Jhelum   | 36  | 22   | 0.310    |

Table 1: Values of Run Test
We see that the corresponding p-values of ONR gives insignificant results providing randomness in the AMMRT series of each sites.

(b) Mann-Whitney U Test

Table 2: Values of Mann-Whitney U Test

| S.No. | Sites     | v   | Groups | W   | P-values |
|-------|-----------|-----|--------|-----|----------|
| 1     | Faisalabad| 36  | 18,18  | 286.0 | 0.141    |
| 2     | Islamabad| 36  | 18,18  | 331.0 | 0.962    |
| 3     | Jhelum    | 36  | 18,18  | 335.0 | 0.962    |
| 4     | Lahore    | 36  | 18,18  | 324.0 | 0.788    |
| 5     | Murree    | 36  | 18,18  | 336.0 | 0.937    |
| 6     | Sargodha  | 36  | 18,18  | 318.0 | 0.646    |
| 7     | Sialkot   | 36  | 18,18  | 338.0 | 0.886    |

*W = Mann-Whitney U Test Statistics

We see that the corresponding p-values of the test statistic (W) give insignificant results that the two groups are independently and identically distributed for each site.

(c) Lag-1 Correlation Coefficient

Table 3: Values of Lag-1 Correlation Test

| S. No. | Sites     | r₁   | LBQ Statistics | P-values |
|-------|-----------|------|----------------|----------|
| 1     | Faisalabad| 0.016| 0.011          | 0.918    |
| 2     | Islamabad| -0.074| 0.214          | 0.644    |
| 3     | Jhelum    | -0.145| 0.826          | 0.363    |
| 4     | Lahore    | 0.057| 0.127          | 0.722    |
| 5     | Murree    | -0.051| 0.103          | 0.748    |
| 6     | Sargodha  | -0.047| 0.085          | 0.770    |
| 7     | Sialkot   | -0.201| 1.580          | 0.209    |

*LBQ Statistics = Ljung Box Q Statistics

We see that the corresponding p-values gives insignificant results that there is no serial correlation in the data at lag one.

L-Moments

L-moments are the linear combination of the probability weighted moments (PWM), introduced by Hosking (1990). The PWM were introduced by Greenwood et al. (1979), mathematically defined as,
Where \( v \) is the number of data for each site, \( i \) is the \( i^{th} \) site and \( j \) is the \( j^{th} \) rank data inter \( j^{th} \) site.

The \( r^{th} \) L-moments defined by Hosking (1990) are,

\[
\lambda_{r+1} = \sum_{i=0}^{r} \beta_i (-1)^{r-k} \binom{r}{i} \left( \tau_i \right)
\]

The first four L-moment ratios are given below:

Location: \( \lambda_1 = \text{Mean of each site} \)

Scale, L-CV (\( \tau \)): \( \tau = \lambda_2 / \lambda_1 \)

L-Skewness (\( \tau_3 \)): \( \tau_3 = \lambda_3 / \lambda_2 \)

L-Kurtosis (\( \tau_4 \)): \( \tau_4 = \lambda_4 / \lambda_2 \)

where

\[
\lambda_1 = \beta_0 \\
\lambda_2 = 2\beta_1 - \beta_0 \\
\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \\
\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0
\]

and the data series \( U_j \) ranked in ascending order.

**Discordancy Measure**

A successful RFA of the data series require that no any site be discordant. Discordancy measures rely on L-moments. Discordancy measurement (\( D_i \)) of a site is calculated by the following formula,

\[
D_i = \frac{V}{3(V-1)} \left( \mu_i - \overline{\mu} \right)^T M^{-1} \left( \mu_i - \overline{\mu} \right)
\]

Where \( V \) is the total no. of sites,

\[
\mu_i = \begin{bmatrix} \tau_i & \tau_3^i & \tau_4^i \end{bmatrix}^T,
\]

\[
\overline{\mu} = V^{-1} \sum_{i=1}^{V} \mu_i,
\]

and \( M \) is the variance-covariance matrix of \( \mu_i \).

If any of the \( D_i \) value of a site exceeds the critical value of discordancy statistics as determined by Hosking and Wallis (1997) the site is considered discordant.

**Table 4**: Summary Statistics of L-Moments and Discordancy Measure of each Site
Regional Rainfall Frequency Analysis on Seven Sites of Punjab, Pakistan using L-moments

| S.No. | Sites      | $v$ | $\lambda_1$ | $\tau_1$ | $\tau_3$ | $\tau_4$ | Di  |
|-------|------------|-----|--------------|----------|----------|----------|-----|
| 1     | Faisalabad | 36  | 99.7         | 0.23     | 0.29     | 0.13     | 1.66|
| 2     | Islamabad | 36  | 203.7        | 0.18     | 0.19     | 0.15     | 0.17|
| 3     | Jhelum     | 36  | 168.3        | 0.16     | 0.07     | 0.10     | 1.05|
| 4     | Lahore     | 36  | 101.5        | 0.18     | 0.16     | 0.06     | 0.63|
| 5     | Murree     | 36  | 280.2        | 0.17     | 0.07     | 0.06     | 0.60|
| 6     | Sargodha   | 36  | 114.3        | 0.22     | 0.38     | 0.30     | 1.64|
| 7     | Sialkot    | 36  | 162.9        | 0.18     | 0.25     | 0.16     | 1.25|

Table 4 illustrates the results of L-moment ratios $(\lambda_1, \tau_1, \tau_3, \tau_4)$ along with the discordancy measure $D_i$. The critical value of $D_i$ for the seven sites is 1.917 (Hosking and Wallis 1997) in the table 3 at page 47. All sites have $D_i$ values < 1.917 exploring no site as discordant.

**Heterogeneity Measure of the Region**

Hosking and Wallis (1993) presented a heterogeneity measure that compares the observed dispersion of L-moments with the simulated dispersion of L-moments calculated through simulations. The four parameter Kappa distribution is used to generate 500 simulated data regions. These regions are homogeneous have no autocorrelation, and sites have the equal record lengths as real data.

The heterogeneity measure is given by:

$$H = \frac{V - \mu_v}{\sigma_v}$$

where $V$ for each simulated sample with $\mu_v$ and $\sigma_v$.

The weighted standard deviation of L CVs is,

$$V = \frac{\sum_{i=1}^{N} v_i(t'_i - \bar{T})^2}{\sum_{i=1}^{N} v_i}$$

The region is

"Acceptably homogeneous" if $H < 1$ ,

"Possibly heterogeneous" if $1 < H < 2$ ,

"Absolutely heterogeneous" if $H \geq 2$ .

The computed heterogeneity measure ($H=-0.79<1$) form an acceptably homogenous region of each site in the northern Punjab.
**L-moment Ratio Diagram**

L-moments ratio diagram is a graph of L-skewness and L-kurtosis. The curve shows the imaginary relation among L-skewness and L-kurtosis, which is used for selection of suitable frequency distribution. L-moments ratio diagram are widely used for the choice of a probability distribution in RFA. L-moment ratios are also called standardized L-moments. The diagram illustrates the most appropriate distribution for the region and used for the estimation of regional quantiles.

![Fig.1: L-Moment Ratio Diagram](image)

In Fig.1 the average \( (\tau_1, \tau_4) \) lies exactly on the curve of the GNO and between the curve of PE3 and GEV distributions. The three distributions are the best choice for the homogeneous region under study.

**Z_Dist Criteria**

It is used to check the goodness-of-fit of all the proposed distribution in the L-Moment Ratio diagram. Hosking and Wallis (1997) give a valuable measure that is defined as,

\[
 Z_{D}^D = \frac{(\tau_{4}^{D} - \tau_{4}^{R} + B_{4})}{\sigma_{4}}
\]

where \( \tau_{4}^{D} \) is the L-kurtosos of the fitted Kappa distribution, \( \tau_{4}^{R} \) s the L-kurtosis, B is the bias of \( \tau_{4}^{R} \) and \( \sigma_{4} \) is the simulated standard deviation of \( \tau_{4}^{R} \).
This fit is good if $|Z^{DIST}| \leq 1.64$, the most suited distribution is selected which has small $|Z^{DIST}|$ measure. Simulation and calculations are covered using FORTRAN routines available in international mathematical and statistical libraries (IMSL).

**Table 5: Results of Goodness-of-Fit Test**

| S. No. | Distributions | $|Z^{DIST}|$ |
|--------|---------------|-------------|
| 1      | PE3           | 0.07        |
| 2      | GNO           | 0.68        |
| 3      | GEV           | 1.05        |
| 4      | GPA           | 2.33        |
| 5      | GLO           | 2.52        |

Table illustrates the results of Goodness of Fit measure we see that PE3, GNO and GEV has the smallest value of the $Z^{DIST}$ criteria computed as 5% level of significance far <1.64. The results are similar to the L-Moment Ratio Diagram.

**The Mean Absolute Deviation Index**

The Mean Absolute Deviation Index ($MADI$) is another measure of goodness of fit that has been used for the comparison between the probability distributions for fitting the data. Jain and Singh was used this method. The main objective of this method is to check whether the given distribution fits the data closely and to select from a number of candidate distributions, the one that gives the best fits to the data. The $MADI$ is calculated by using following formula:

$$MADI= \frac{1}{v} \sum_{(i=1)}^{v} |(u_i-w_i)/u_i|$$

Where $u_i$ is the observed values for each site, $w_i$ is the predicted values of L-moments mean ($l_1$) and $v$ is the number of observation for each site.

The smallest the value of $MADI$ indicates that this distribution is more appropriate for the actual data.

**Table 6: Values of MADI for Different Distributions using L-moments**

| Sr. No. | Station Name | PE3 | GNO | GEV |
|---------|--------------|-----|-----|-----|
| 1       | Faisalabad   | 0.459 | 0.362 | 0.400 |
| 2       | Islamabad   | 0.303 | 0.291 | 0.268 |
| 3       | Jhelum      | 0.327 | 0.260 | 0.321 |
| 4       | Lahore      | 0.236 | 0.285 | 0.253 |
| 5       | Murree      | 0.238 | 0.278 | 0.276 |
| 6       | Sargodha    | 3.464 | 0.323 | 0.263 |
| 7       | Sialkot     | 0.325 | 0.259 | 0.277 |
Table 7: Ranking of Distributions by MADI using L-moments

| Distribution | 1 | 2 | 3 |
|--------------|---|---|---|
| GEV          | 2 | 5 | 0 |
| GNO          | 3 | 2 | 2 |
| PE3          | 2 | 0 | 5 |

The results of MADI for seven stations are given in the above table 6 are also ranked for order 1 to 3 given in table 7. Table 7 represents PE3 distribution receives rank 1 for two sites, GNO receives rank 1 for three sites and GEV receives rank 1 for two sites. So it can be concluded that by using L-moments GNO receives high rank as compare to PE3 and GEV. Therefore, GNO is the best fitted distribution for the region according to MADI.

Estimation of Quantiles and Statistical Inference

The parameters estimates of the best fitted frequency distributions are as follow.

Table 8: Parameter Estimates

| Distribution | μ   | α   | λ   |
|--------------|-----|-----|-----|
| PE3          | 1.000 | 0.355 | 1.241 |
| GNO          | 0.930 | 0.314 | -0.424 |
| GEV          | 0.834 | 0.262 | -0.054 |

Where μ is the location parameter, α is the scale parameter and λ is the shape parameter of the following distribution.

Table 9: Regional Quantiles Estimates of the Fitted Distributions

| F   | PE3  | GNO  | GEV  |
|-----|------|------|------|
| 0.0100 | 0.495 | 0.466 | 0.450 |
| 0.0500 | 0.564 | 0.558 | 0.555 |
| 0.1000 | 0.618 | 0.620 | 0.620 |
| 0.2000 | 0.701 | 0.708 | 0.711 |
| 0.5000 | 0.928 | 0.930 | 0.931 |
| 0.9000 | 1.476 | 1.465 | 1.461 |
| 0.9500 | 1.680 | 1.677 | 1.678 |
| 0.9800 | 1.938 | 1.958 | 1.972 |
| 0.9900 | 2.127 | 2.175 | 2.202 |
| 0.9950 | 2.312 | 2.397 | 2.440 |
| 0.9975 | 2.493 | 2.624 | 2.687 |
| 0.9980 | 2.551 | 2.699 | 2.768 |
Table 9 provides the quantiles estimates at different return periods (nonexceedance probabilities) for the best fitted distribution. The return period is also called as recurrence interval, is an estimation of the probability of an event to occur, such as rainfall, flood, storms etc. The return period \( T \) is computed as \( T = \frac{1}{P} \) where P correspond to the probability of exceedance (the chance of occurrence of some events over a given time period), and is related to \( T \) as, \( P = \frac{1}{T} \). The non-exceedance probability is computed as \( F = 1 - P \). Let suppose for a ten years return period \( T=10 \), \( P=0.1 \) and the non-exceedance probability is calculated as \( F=1-0.1=0.900 \). The quantiles estimates can be interpreted as, for PE3 distribution, the non-exceedance probability \( F=0.9000 \), \( q(F)=PE3(0.9000)=2.127 \) is the magnitude of rainfall which will happen once on the average in 10 years, for a non-for GNO distribution, the non-exceedance probability \( F=0.9900 \), \( q(F)=GNO(0.9900)=2.175 \) is the magnitude of rainfall which will happen once on the average in 100 years, similarly for GEV distribution, the non-exceedance probability \( F=0.9980 \), \( q(F)=GEV(0.9980)=2.768 \) is the magnitude of rainfall which will happen once on the average in 500 years. All these estimates of quantiles for given distribution can be interpreted in the similar manner.

Table 10: At Site Quantiles Estimates of the Seven Sites

| F   | Faisalabad | Islamabad | Jhelum | Lahore | Murree | Sargodha | Sialkot |
|-----|------------|-----------|--------|--------|--------|----------|---------|
| 0.010| 49.3       | 100.8     | 83.3   | 50.3   | 138.7  | 56.6     | 80.6    |
| 0.050| 56.2       | 114.9     | 94.9   | 57.3   | 158.0  | 64.5     | 91.9    |
| 0.100| 61.6       | 125.9     | 104.0  | 62.8   | 173.2  | 70.6     | 100.7   |
| 0.200| 69.9       | 142.8     | 117.1  | 71.2   | 196.4  | 80.1     | 114.2   |
| 0.500| 92.5       | 189.0     | 156.2  | 94.2   | 260.0  | 106.1    | 151.2   |
| 0.900| 147.1      | 300.7     | 248.4  | 149.9  | 413.6  | 168.7    | 240.4   |
| 0.950| 167.5      | 342.2     | 282.8  | 170.6  | 470.7  | 191.1    | 273.7   |
|      | 193.2      | 394.8     | 326.2  | 196.8  | 543.0  | 221.5    | 315.7   |
Table 10 provides the at site quantiles estimates for different nonexceedance probabilities of PE3 distribution. The values of quantiles given in table can be interpreted as, for a site Faisalabad, the nonexceedance probability $F=0.9900$, $q(F)=PE3(0.9900)=211.1$ is the magnitude of rainfall which will happen once on the average in 100 years.

**Results**

On the basis of the best regional distribution the quantile estimates present the expected occurrence of rainfall in the coming years to occur. These may be used for hydrological planning and management of the water resources particularly in the monsoons spells in the upper Punjab in the years to come. The estimated rainfall quantiles are used for the preventive measure of flood disaster, agricultural water management and in the improvement projects for the rehabilitation and modernization of major barrages.

**Recommendations**

The RRFA was first developed as a statistical series of applications for the data of extreme events in 1993. Hosking and Wallis basically worked on it. In Pakistan some statistical work exits in literature focusing on the sites of Sindh and Punjab for the whole year. The novelty of the study lies here in particular the monsoons winds which are the sources of major water related issues.

There is a need to work out the regional quantiles in the rainy season of the meteorological sites in the provinces of Pakistan for detailed comparative study.

**References**

Hirsch, R. M., & Slack, J. R. (1984). A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*, 20(6), 727-732.

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29(2), 271-281. doi: 10.1029/92wr01980

Hussain, Z., Shahzad, M. N., & Abbas, K. (2016). Application of regional rainfall frequency analysis on seven sites of Sindh, Pakistan. *KSCE Journal of Civil Engineering*, 1-8.
Regional Rainfall Frequency Analysis on Seven Sites of Punjab, Pakistan using L-moments

Khalifa, A. H. N., & Al-Omran, A. T. CONSUMPTION ADAPTIVE VOLTAGE CONTROLLED OSCILLATOR USING FPAA.

Khattak, M. S., & Ali, S. (2015). Assessment of temperature and rainfall trends in Punjab province of Pakistan for the period 1961-2014. *Journal of Himalayan Earth Science, 48*(2).

Koutsoyiannis, D., & Baloutsos, G. (2000). Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences. *Natural Hazards, 22*(1), 29-48.

Lee, C. (2005). Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in Southern Taiwan. *Crop Environ. Bioinf, 2*, 31-38. Mekonnen, S. (2013). *Rainfall Intensity-Duration-Frequency Analysis for Selected Areas of Wollega Zone, Ethiopia*. Haramaya University.

Salma, S., Rehman, S., & Shah, M. (2012). Rainfall trends in different climate zones of Pakistan. *Pakistan Journal of Meteorology, 9*.

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).

Van Gelder, P., & Neykov, N. M. (1998). *Regional frequency analysis of extreme water levels along the North-Sea Coasts*. Paper presented at the European Geophysical Society, XXIV General Assembly, The Hague, The Netherlands.

Vivekanandan, N. (2014). Rainfall Frequency Analysis Using L-moments of Probability Distributions. *International Journal of Computer Application and Engineering Technology, 3*(3), 248-256.

Zahid, M., & Rasul, G. (2011). Frequency of extreme temperature and precipitation events in Pakistan 1965–2009. *Sci. Int.(Lahore), 23*(4), 313-319.