International Journal of Advanced Statistics and Probability, *1 (3) (2013) 97-101*
©Science Publishing Corporation
www.sciencepubco.com/index.php/IJASP

A review: regional frequency analysis of annual maximum rainfall in monsoon region of Pakistan using L-moments

Amina Shahzadi

Department of Statistics, GC University, Lahore
E-mail: aminashahzadi@gmail.com

Abstract

A regional frequency analysis has been conducted for annual maximum rainfall in Monsoon region of Pakistan using Hosking's L-moments technique. Monsoon region was divided into three regions based on at-site characteristics. It was found generalized normal distribution to be best robust for quantile estimation of large return periods and generalized extreme value distribution for quantiles of low return periods for all three regions.

Keywords: *Monsoon rainfall, L-moments, Regional frequency analysis*

1 Introduction

Extreme events like extreme rainfalls, floods and tsunamis can take thousands of life and cause billions of property in damage. Hydrologists are always interested in the estimation of likelihood of such events for the construction of hydraulic structures like dams, reservoirs, spillways and storm surge barriers etc. so that the water can be stored and lives and money can be saved. Physical laws are inappropriate to deal with the inherit randomness and uncertainty of extreme events. Statistical methods acknowledge the existence of uncertainty and enable its effects to be quantified (Hosking and Wallis [11]).
Frequency analysis is the estimation of magnitude and frequency of rare events. According to Chow et al. [3], the frequency analysis is to relate magnitude of extreme events to their frequency of occurrence using probability distributions. The availability of data is an important aspect in frequency analysis. If there is sufficiently a long record length of extreme rainfall or flood data, then its frequency distribution can be specified. But in many situations, the data available is usually insufficient to define the risk precisely. And the estimates obtained from small samples are unrealistic, especially when distribution has three or more parameters (Rao and Hamed [18]).
When one has 30 years of data on a variable to estimate the extreme event expected to occur in 100 or 1000 years, then extrapolation is required. In such situation, data of a particular phenomenon from several sites are pooled for estimation with greater reliability. This approach is known as regional frequency analysis. This is the substitution of space for time by using the data from different locations in a region to compensate the short record lengths (National Research Concil, [16]; Stendiger et al. [21]). This approach is easier for rainfall because of uniform characteristics over large areas but difficult for flood because of effects of topography and geology (Stendiger at el. [21]). Regional frequency analysis for rainfall is usually based on annual maximum rainfall (AMR) or partial duration series for extrapolation. A bulk of literature is available on regional frequency analysis of rainfall, e.g. see Section 1 Shahzadi et al. [20].

2 Purpose and area of study

Modeling of rainfall has immense importance for construction of hydraulic structures as well as for agricultural planning and electricity purposes in Pakistan. Because of topographic conditions, there is extreme variation in climate of Pakistan. Pakistan receives rainfall in summer due to monsoon and in winter due to western disturbances. Northern areas receive more rainfall and rainfall decreases in southern part of Pakistan. Monsoon causes 60% of rainfall in Pakistan and 40% because of other causes.
An index-flood regional frequency analysis of AMR has been conducted by Shahzadi et al. (2013) in Monsoon climatic zone of Pakistan. The data was collected from Regional Meteorological Centre (RMC) Lahore. *L*-moments statistics were used for characterizing variability, skewness and kurtosis of data sets, measuring the homogeneity and helping in

identification of regional growth curves. The main objective was to estimate the quantile on regional data basis, so that the at-site quantiles or design events can be obtained with greater reliability.

3 Index-Flood Regional Frequency Analysis (RFA)

There are several ways of regionalization (Buishand [1], Cunnane [4], GREHYS [7]). The successful and simple regionalization technique proposed by Dalrymple [5] is index flood regionalization. Since the technique has its first application to floods, so it has been known as index-flood RFA. However, it can be applied to any kind of data.

Suppose that there are N sites in a homogeneous region, with site i having record length of n_i and observed data $Q_{ij}, j = 1,2, \dots, n_i$. Let $Q_i(F),\ 0 < F < 1$ be the quantile function of frequency distribution at site i. Then the index flood procedure states that the distribution of N sites are identical except for a scale or index flood parameter. So we may write:

$Q_i(F) = \mu_i q(F)$

where μ_i is the mean at site i usually employed as index flood. However, any location parameter e.g. median or trimmed mean can be used as index flood (Hosking [9]). The$q(F)$, common to each site, is the dimensionless quantile function of regional frequency distribution of rescaled data $q_{ij} = \frac{Q_{ij}}{\mu_i}, j = 1,2, \dots, n_i, i = 1,2, \dots, N$. British flood studies report [17] called $q(F)$ the regional growth curve.

The estimated quantile $Q_i(F)$ are used as estimates of extreme events or design events in Hydrology with non-exceedance probability *F*. These events are usually interpreted in terms of return period. A return period is defined as average time interval between the occurrences of two extremes. The quantile of return period *T*, Q_T, ia an event magnitude so extreme that it may be exceeded any single event once in T years.

i.e. $Q_T = Q(1 - \frac{1}{T})$

A return period is often specified rather than nonexceedance probability (1-1/*T*), e.g., annual maximum rainfall exceeded with 1 percent probability in any year or chance of 1 in 100 years is called 100-year rainfall event (Riggs [19] and Chow et al. [3]).

4 Estimation method: L-moments

When rainfall, flood flows or any water-quality variables are well described by some family of distributions, the tasks of hydrologists is to estimate the parameters of distribution so that the quantiles can be calculated with the fitted model. Several estimation methods have been suggested in literature e.g. method of moments, maximum likelihood method, Bayesian approach etc. Method of moments is simple but moment estimators are sensitive to extreme values. Maximum likelihood estimates have very good large sample properties but maximum likelihood estimators cannot be reduced to simple formulas, so estimates must be calculated using numerical techniques (Kite [14]). Bayesian approach has also been used which combines prior information and regional hydraulic information with likelihhod function with available data (e.g. see references in Stedniger et al. [21] & GREHYS [7]). Probability weighted moments (PWM) have also been used as parameter estimation method (Greenwood et al. [6]). However, PWMs are difficult to interpret directly as measures of scale and shape of distribution (Hosking [9]). Since the decade of 1990's, *L*-moments have been used for regional frequency analysis in Hydrology. *L*-moments are the linear combinations of PWMs and also defined by linear combinations of order statistics (Hosking [9]). The rth *L*-moment and its unbiased estimator as linear combination of sample order statistics are respectively;

$$\lambda_r = \sum_{k=0}^{r-1} \beta_k (-1)^{r-k-1} r^{-1} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} E(X_{r-j:r})\ ; \quad \text{provided, } E|X| \text{ exists.}$$

$$l_r = \sum_{k=0}^{r-1} p^*_{r-1,k}\, b_k, \quad r = 1,2, \dots, n$$

where $p^*_{r-k} = (-1)^{r-k}\binom{r}{k}\binom{r+k}{k}$ and $b_r = n^{-1}\binom{n-1}{r}^{-1} \sum_{j=r+1}^{n} \binom{j-1}{r} x_{j:n}$ is unbiased estimator of rth PWM $\beta_r = E[X\{F(X)\}^r], \quad r = 0,1$ ....

The sample *L*-moment ratios, i.e., *L*-C_v (coefficient of variation), *L*-C_s (coefficient of skewness) and *L*-C_k (coefficient of kurtosis) are dimensionless meaures given by

$$t = l_2/l_1, \quad t_r = l_r/l_2, \quad r = 3,4$$

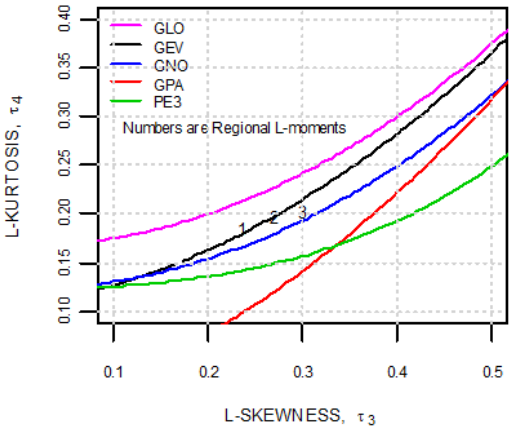
These are analog of ordinary moment ratios and measure the shape of distribution independently of its scale of measurement. *L*-moments have superior abilities which make them suitable for small samples and heavy tailed distributions (e.g. Hosking [9]).

5 L-moments in RFA

A unified theory of RFA using *L*-moments has been outlined by Hosking and Wallis [11]. The practical implementation of regional frequency analysis involves various steps. Initially RFA requires the checking of any inconsistencies or gross errors in data sets. This can be done using discordancy measure (*D*) which is based on *L*-moments ratios (Chapter 3, Hosking and Wallis [11]. The delineation of homogeneous region requires careful attention and may involve some subjective judgement, but it can be tested by objective measures e.g. *L*-moments based heterogeneity statistics defined in Chapter 4 of Hosking and Wallis [11]. The heterogeneity measure (H) compares the between-site variation in the L-moment ratio with what would be expected for a homogeneous region.
When dealing with extreme values, three or more parametric should be considered in order to produce the less biased estimates in tails of distribution. If we look at literature, Hosking [9], Stendiger et al. [21], Vogel and Fennessey [22] and Hosking and Wallis [11] suggested the L-moment ratio diagram for the selection of regional distribution. A goodness of fit statistics, Z^{DIST}-statistic, can also be used when regional average point (t_3^R, t_4^R) falls between two curves. This measure compares the regional *L*-C_k (t_4) with its distribution when a homogeneous region has specified frequency distribution (Chapter 5, Hosking and Wallis, [11]).
A regional *L*-moment algorithm (pg.89, Hosking and Wallis [11]) involves the estimation of chosen regional growth curve *q(F)* which can be used for at-site quantiles. Hosking and Wallis [11] has presented an algorithm for simulation of regional L-moment algorithm in Table 6.1 (Chapter 6) which is used to calculated the some accuracy measures e.g. Bias, Absolute Bias, Relative Root Mean Square Error (RMSE) and 90% confidence limits for estimated regional growth curves and at-site quantiles.

6 Results and conclusions

A regional analysis was conducted for AMR for 24 hours duration in Monsoon climatic zone of Pakistan. A total of 23 gaged stations were included in the study having record lengths from 24 to 60 years (Figure 1, Shahzadi et al. [20]). Assumptions of stationarity, identically and independently distributed were satisfied by Mann-Whitney test, Kendall's tau test and Ljung-Box Q-statistic. Site 17, Multan showed significant autocorrelations (Table 1) at various lags. It was because of many drought years for this site. But this site had moderate values of *L*-C_v, *L*-C_s and *L*-C_k. Site 20, Khanpur, was having large value of *L*-C_v, so appeared to be discordant with D_i= 3.0203.
Site 2, Saidu Sharif, was not discordant but it has also relatively large *L*-C_v (Table 2, [20]). Sites 2 and 20 did not form any homogeneous region with their nearest sites, so they were discarded from analysis. Remaining 21 gaged sites were made of 3 homogeneous regions, region 1 having 11 sites, region 2 and 3 each having 5 sites. The three regions satisfied the heterogeneity statistics (Table 4, [20]).



***L*-moment Ratio Diagram**

Z^{DIST}-Statistic for Various Distributions

Region	Z^{GLO}	Z^{GEV}	Z^{GNO}	Z^{GPA}	Z^{PE3}
1	*1.16**	*-0.49**	*-1.14**	*-4.53*	*-2.31*
2	*0.61**	*-0.23**	*-0.74**	*-2.44*	*-1.65*
3	*0.92**	*0.18**	*-0.42**	*-1.87*	*-1.66*

*shows acceptance at 10% level of significance

L-moment ratio diagram plotted above is depicting the regional average *L*-C_s and *L*-C_k of final three regions. The point (t_3^R, t_4^R) for region1 and 2 falls on GEV curve while for region 3 it is between GEV and GNO curve. In table above, the Z^{DIST}-statistic accepted GEV, GLO and GNO distributions at 10% level of significance for all three regions.

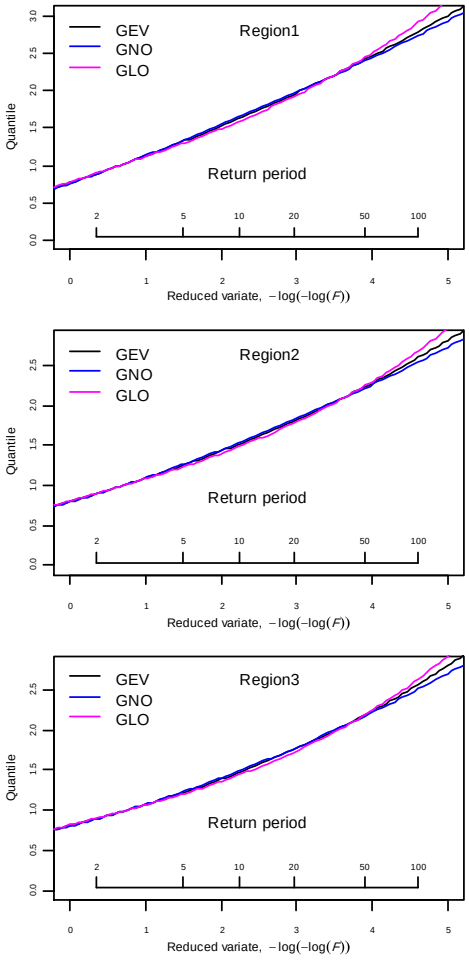
Using L-moments regional algorithm, regional growth curves of GEV, GNO and GLO for all regions were estimated for various return periods (Table 8, [20]) plotted here. The growth curves are approximately equal for low return periods but after return period 50, regional quantiles become large from GLO distribution for all three regions.

Since the actual objective of RFA is to obtain the best robust distribution giving the robust estimates. For this purpose an assessment analysis were performed in Table 8(a,b,c). GLO distribution produced quantiles with lager values of BIAS, absolute BIAS, relative RMSE.

The quantiles from GNO distribution are robust for large return periods of 50, 100, 200, 500 and 1000 on the basis of less absolute BIAS, less RMSE while for low return periods of 1,2, 5, 10 and 20 GEV distribution behaved well.

Thus GEV was found to be best for low return periods and GNO for large return periods in Monsoon climatic zone of Pakistan. Same conclusion have been found for annual peak flows in Pakistan (Hussain and Pasha [13]).

At-site quantiles at any site of Monsoon climatic zone can be calculated by multiplying site mean with the corresponding regional growth curve.



Regional Growth Curves

7 Recommendations for future study

1. The study can be conducted using other estimation methods like maximum likelihood and trimmed *L*-moments, then estimated quantiles can be compared.
2. The at-site quantiles can be estimated using frequency factor (Cheng et al. [2]) method from GEV and GNO distribution and then can be compared obtained by RFA.
3. The four or five parametric distributions like Kappa or Wakeby distribution can be used for quantile estimates which can be compared with those obtained from three parametric distributions

References

[1] Buishand, T. A. (1990). Bias and variance of quantile estimation from a partial duration series. Journal of Hydrology, 120, 35-49.
[2] Cheng, K.-S., Chiang, J.-L. & Hsu, C.-W. (2006). Simulation of probability distributions commonly used in hydrological frequency analysis. Hydrological Processes, DOI: 10.1002/hyp.6176.
[3] Chow, V. T., Maidment, D. R. & Mays, L.W. (1988). Applied Hydrology. McGraw-Hill, New York
[4] Cunnane, C. (1988). Methods and merits of regional frequency analysis. Journal of Hydrology, 100, 269-290.
[5] Dalrymple, T. (1960). Flood frequency analyses. Water Supply Paper 1543-A, U.S. Geological Survey, Reston, Va.
[6] Greenwood, J.A., Landwehr, J.M., Matalas, N.C. & Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. Water Resources Research, 15(5), 1049-1054
[7] GREHYS (1996). Presentation and review of some methods of regional flood frequency analysis. Journal of Hydrology, 186, 63-84.
[8] Hirsch, R.M., Helsel, D.R., Chon, T. A. & Gilory, E. J. (1993). Statistical analysis of Hydrological data, In: D.R. Maidment (ed.) Handbook of Hydrology, McGraw Hill, New York.
[9] Hosking, J. R. M. (1991). Approximations for use in constructing L-moment ratio diagram. Res. Rep. RC 16635, IBM Research Division, Yorktown Heights, New York.
[10] Hosking, J.R.M. & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. Water Resources Research, 29, 271-281.

[11] Hosking, J.R.M. & Wallis, J. R. (2005). Regional Frequency Analysis: An approach Based on L-moments. Cambridge University Press, New York.
[12] Hosking, J.R.M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal Royal Statistical Society, Series B, 52, 105-124.
[13] Hussain, Z. & Pasha, G. R. (2009). Regional Flood Frequency Analysis of the Seven Sites of Punjab, Pakistan, Using L-Moments. Water Resources Management, 23(10), 1917-1933.
[14] Kite, G. W. (1988). Frequency and Risk Analysis in Hydrology. Water Resources Publications, Fort Collins, Co.
[15] Ljung, G. M. & Box, G. E. P. (1978). On a measure of a lack of fit in time series models, Biometrika, 65, 297-303.
[16] National Research Council, (1988). Committee on techniques for estimating probabilities of extreme floods, "Estimating Probabilities of Extreme Floods, Methods and Recommended Research", National Academy Press, Washington, D. C
[17] Natural Environmental Research Council (1975). Flood Studies Report. Vol. 1, Hydrological Studies, London.
[18] Rao, A. R. & Hamed, K. H. (2000). Flood Frequency Analysis. CRC press, New York.
[19] Riggs, H. C. 1961. Frequency of Natural Events. Journal of Hydraulic Division, ASCE, 15-26.
[20] Shahzadi, A., Akhter, A. S. and Saf, B. (2013). Regional frequency analysis of annual maximum rainfall in Monsson region of Pakistan using L-moments. Journal of Statistical and Acturial Sciences, IX(1), 111-136.
[21] Stendinger, J.R., Vogel, R.M. & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events, In: D.R. Maidment (ed.) Handbook of Hydrology, McGraw Hill, New York
[22] Vogel, R.M. & Fennessey, N.M. (1993). L-moment diagrams should replace product moment diagrams, Water Resources Research, 29(6), 1745-1754.