Chapter 3

Regional Annual Maximum Rainfall Frequency Analysis of North East India using LQ-moment.

3.1 Introduction

In this chapter regional annual maximum rainfall analysis of North East India has been carried out using LQ-moment. Trimean based quick estimator has been used in this study. Five extreme probability distributions namely the generalized extreme value (GEV), generalized logistic (GLO), generalized Pareto (GPA), generalized log normal (GNO) and Pearson type III (PE3) are used which are also used in the previous chapter.

3.2 LQ-moment

Let $X_1, X_2, X_3, \ldots X_n$ be a sample from a continuous distribution function $F_x(.)$ with quantile function $Q_x(u) = F_x^{-1}(u)$. If $X_{1:n} \leq X_{2:n} \leq X_{3:n} \leq \cdots \leq X_n$ denote the order statistics, then the rth LQ-moments ζ_r of X proposed by Mudholkar and Hutson [34] are given by

$$\zeta_r = r^{-1} \sum_{k=0}^{k=r-1} (-1)^k \binom{r-1}{k} \tau_{p,\alpha(X_{r-k:r})}, \qquad r=1,2,\ldots \tag{3.2.1}$$

where $0 \leq \alpha \leq \frac{1}{2}$, $0 \leq p \leq \frac{1}{2}$, and

$$\tau_{p,\alpha}(X_{r-k:r}) = pQ_{x_{r-k:r}}(\alpha) + (1-2p)Q_{x_{r-k:r}}(1/2) + pQ_{x_{r-k:r}}(1-\alpha) \tag{3.2.2}$$

The linear combination $\tau_{p,\alpha}$ is a quick measure of the location of the sampling distribution of order statistic $X_{r-k:r}$. With appropriate combinations of α and p, estimators for $\tau_{p,\alpha}(.)$ can be found which are functions of commonly used estimators such as median , trimean and Gastwirth. The trimean-based estimator is defined as

$$\frac{Q_{x_{r-k:r}}\left(\frac{1}{4}\right)}{4} + \frac{Q_{x_{r-k:r}}\left(\frac{1}{2}\right)}{2} + \frac{Q_{x_{r-k:r}}\left(\frac{3}{4}\right)}{4} \tag{3.2.3}$$

The first four LQ-moments of the random variable X are given by:

$$\zeta_1 = \tau_{p,\alpha}(X), \tag{3.2.4}$$

$$\zeta_2 = \frac{1}{2}[\tau_{p,\alpha}(X_{2:2}) - \tau_{p,\alpha}(X_{1:2})], \tag{3.2.5}$$

$$\zeta_3 = \frac{1}{3}[\tau_{p,\alpha}(X_{3:3}) - 2\tau_{p,\alpha}(X_{2:3}) + \tau_{p,\alpha}(X_{1:3})], \tag{3.2.6}$$

$$\zeta_4 = \frac{1}{4}[\tau_{p,\alpha}(X_{4:4}) - 3\tau_{p,\alpha}(X_{3:4}) + 3\tau_{p,\alpha}(X_{2:4}) - \tau_{p,\alpha}(X_{1:4})]. \tag{3.2.7}$$

The coefficient of LQ-variation, LQ-skewness and LQ-kurtosis are defined by

$\eta = \frac{\zeta_2}{\zeta_1}$,

$\eta_3 = \frac{\zeta_3}{\zeta_2}$ and (3.2.8)

$\eta_4 = \frac{\zeta_4}{\zeta_2}$ respectively.

Mudholkar and Hutson [34] defined the quick location measure as

$$\tau_{p,\alpha}(X_{r-k:r}) = pQ_X[B_{r-k:r}^{-1}(\alpha)] + (1-2p)Q_X[B_{r-k:r}^{-1}(1/2)] + pQ_X[B_{r-k:r}^{-1}(1-\alpha)] \tag{3.2.9}$$

where $B_{r-k:r}^{-1}(\alpha)$ is the corresponding αth quantile of a beta random variable with parameters $r-k$ and $k+1$.

For sample order statistics Mudholkar and Hutson [34] defined the quantile estimator of $Q_X(u)$ as follows:

$$\hat{Q}_X(u) = (1-\varepsilon)X_{[n'u]:n} + \varepsilon X_{[n'u]+1:n} \tag{3.2.10}$$

where, $\varepsilon = n'u - [n'u]$ and $n' = n+1$

The rth sample LQ-moment is defined as

$$\hat{\zeta}_r = r^{-1} \sum_{k=0}^{k=r-1} (-1)^k \binom{r-1}{k} \hat{\tau}_{p,\alpha(X_{r-k:r})}, \qquad r=1,2,\ldots \tag{3.2.11}$$

where the quick estimator $\hat{\tau}_{p,\alpha(X_{r-k:r})}$ of the location of the order statistic $X_{r-k:r}$ is given by

$$\hat{\tau}_{p,\alpha}(X_{r-k:r}) = p\hat{Q}_{x_{r-k:r}}(\alpha) + (1-2p)\hat{Q}_{x_{r-k:r}}\left(\frac{1}{2}\right) + p\hat{Q}_{x_{r-k:r}}(1-\alpha)$$
$$= p\hat{Q}_X[B_{r-k:r}^{-1}(\alpha)] + (1-2p)\hat{Q}_X[B_{r-k:r}^{-1}(1/2)] + p\hat{Q}_X[B_{r-k:r}^{-1}(1-\alpha)] \tag{3.2.12}$$

where $0 \leq \alpha \leq \frac{1}{2},\ 0 \leq p \leq \frac{1}{2}$.

The first four sample LQ-moment can be written as

$$\hat{\zeta}_1 = \hat{\tau}_{p,\alpha}(X), \tag{3.2.13}$$

$$\hat{\zeta}_2 = \frac{1}{2}[\hat{\tau}_{p,\alpha}(X_{2:2}) - \hat{\tau}_{p,\alpha}(X_{1:2})], \tag{3.2.14}$$

$$\hat{\zeta}_3 = \frac{1}{3}[\hat{\tau}_{p,\alpha}(X_{3:3}) - 2\hat{\tau}_{p,\alpha}(X_{2:3}) + \hat{\tau}_{p,\alpha}(X_{1:3})], \tag{3.2.15}$$

$$\hat{\zeta}_4 = \frac{1}{4}[\hat{\tau}_{p,\alpha}(X_{4:4}) - 3\hat{\tau}_{p,\alpha}(X_{3:4}) + 3\hat{\tau}_{p,\alpha}(X_{2:4}) - \hat{\tau}_{p,\alpha}(X_{1:4})]. \tag{3.2.16}$$

The sample coefficient of LQ-variation, LQ-skewness and LQ-kurtosis are defined by

$\hat{\eta} = \frac{\hat{\zeta}_2}{\hat{\zeta}_1}$,

$\hat{\eta}_3 = \frac{\hat{\zeta}_3}{\hat{\zeta}_2}$ and (3.2.17)

$\hat{\eta}_4 = \frac{\hat{\zeta}_4}{\hat{\zeta}_2}$ respectively.

3.3 Trimean based parameter estimation of probability distributions

3.3.1 GEV Distribution

According to Mudholkar and Hutson [34] the shape parameter k can be estimated with good accuracy by using the approximation equation

$$k = 0.2985 - 0.0234\eta_3 + 0.3732\eta_3^2 - 0.1429\eta_3^3 + 0.0449\eta_3^4 \tag{3.3.1}$$

Location and scale parameters can be estimated using following relations.

$$\xi = \hat{\zeta}_1 - \alpha\left\{\frac{1}{4}Q_0\left(\frac{1}{4}\right) + \frac{1}{2}Q_0\left(\frac{1}{2}\right) + \frac{1}{4}Q_0\left(\frac{3}{4}\right)\right\} \tag{3.3.2}$$

$$\alpha = 8\hat{\zeta}_2/\{2Q_0(0.707) - 2Q_0(0.293) + Q_0(0.866) - Q_0(0.134)\} \tag{3.3.3}$$

where $Q_0(F) = \{1 - (-logF)^k\}/k$

3.3.2 GPA Distribution

The shape parameter k can be estimated with good accuracy by using the approximation equation (Bhuyan [4])

$$k = 0.9998 - 3.4965\eta_3 + 1.4681\eta_3^2 - 0.6243\eta_3^3 + 0.1535\eta_3^4. \qquad (3.3.4)$$

Location and scale parameters can be estimated using relations (3.2.19) and (3.3.5) where $Q_0(F) = \{1 - (1 - F)^k\}/k$.

3.3.3 GLO Distribution

The shape parameter k can be estimated with good accuracy by using the approximation equation (Bhuyan [4])

$$k = -1.3328\eta_3 - 0.0286\eta_3^3 + 0.0166\eta_3^5. \qquad (3.3.6)$$

Location and scale parameters can be estimated using relations (3.2.19) and (3.2.20) where $Q_0(F) = \left[1 - \left\{\frac{(1-F)}{F}\right\}^k\right]/k$.

3.3.4 GNO Distribution

The shape parameter σ can be estimated with good accuracy by using the approximation equation (Bhuyan [4])

$$\sigma = 2.3284\eta_3 - 0.0002\eta_3^2 + 0.1220\eta_3^3 + 0.0009\eta_3^4 - 0.0332\eta_3^5 \qquad (3.3.7)$$

Location and scale parameters can be estimated using following relations.

$$\zeta = \hat{\zeta}_1 - exp(\alpha)\left\{\frac{1}{4}Q_0\left(\frac{1}{4}\right) + \frac{1}{2}Q_0\left(\frac{1}{2}\right) + \frac{1}{4}Q_0\left(\frac{3}{4}\right)\right\} \qquad (3.3.8)$$

$$\mu = 8\hat{\zeta}_2/\{2Q_0(0.707) - 2Q_0(0.293) + Q_0(0.866) - Q_0(0.134)\} \qquad (3.3.9)$$

where $Q_0(F) = exp\{\sigma\emptyset^{-1}(F)\}$ and the standard parametrization may be obtained from the parametrization by setting

$$k = -\sigma, \quad \alpha = \sigma e^{\mu}, \quad \xi = \zeta + e^{\mu}.$$

3.3.5 PE3 Distribution

The shape parameter γ can be estimated with good accuracy by using the approximation equation (Bhuyan [4])

$$\gamma = 6.9839\eta_3 + 0.0001\eta_3^2 - 6.6634\eta_3^3 - 0.0035\eta_3^4 \qquad (3.3.10)$$

Location and scale parameters can be estimated using following relations.

$$\xi = \hat{\zeta}_1 - \beta\left\{\frac{1}{4}Q_0\left(\frac{1}{4}\right) + \frac{1}{2}Q_0\left(\frac{1}{2}\right) + \frac{1}{4}Q_0\left(\frac{3}{4}\right)\right\} \quad (3.3.11)$$

$$\beta = 8\hat{\zeta}_2/\{2Q_0(0.707) - 2Q_0(0.293) + Q_0(0.866) - Q_0(0.134)\} \quad (3.3.12)$$

where parameters μ (location), σ (scale) and γ (shape) are the standard parameterizations which can be obtained by setting $\alpha = \frac{4}{\gamma^2}$, $\beta = \frac{1}{2}\sigma|\gamma|$ and $\xi = \mu - \frac{2\sigma}{\gamma}$.

3.4. Regional Rainfall Frequency Analysis using LQ-moment

The procedure discussed in section 2.5 can also be employed for LQ-moment. For this purpose, L-cv, L-skewness and L-kurtosis are replaced by LQ-cv, LQ-skewness and LQ-kurtosis. For all calculations Fortran 77 programs are used.

3.4.1 Screening of data

As discussed in the section 2.5.1 Discordancy test D_i can be written as

$$D_i = \frac{1}{3}N(u_i^{LQ} - \bar{u}^{LQ})^T S_{LQ}^{-1}(u_i^{LQ} - \bar{u}^{LQ}) \quad (3.4.1)$$

where $S_{LQ} = \sum_{i=1}^{N}\left(u_i^{LQ} - \bar{u}^{LQ}\right)(u_i^{LQ} - \bar{u}^{LQ})^T$ and $u_i^{LQ} = \left[\hat{\eta}^i, \hat{\eta}_3^i, \hat{\eta}_4^i\right]^T$ for i-th station, N is the number of stations, S_{LQ} is covariance matrix of u_i^{LQ} and $\bar{u}^{LQ}$ is the mean of vector, u_i^{LQ}. Critical values of discordancy statistics are tabulated by Hosking and Wallis [25]. For $N = 12$, the critical value is 2.757. If the D-statistics of a station exceeds 2.757, its data is discordant from the rest of the regional data.

The calculated D_i values are given in the Table 3.1. It is observed from Table 3.1 that the D_i values of all the 12 stations of our study region are less than the critical value 2.757. Hence there is no data of any stations which are discordant from rest of the regional data. Therefore, all the data from 12 stations can be considered for our study.

Table 3.1: Discordancy measures of each sites of the NE region using LQ-moment

Name of sites	No. of observation	$\hat{\eta}$	$\hat{\eta}_3$	$\hat{\eta}_4$	D_i
1. Guwahati	30	0.1492	0.3960	0.1093	1.12
2. Mohanbari	30	0.1565	-0.0545	-0.0625	0.83
3. Silchar	28	0.1534	0.0931	0.2052	0.69
4. Lakhimpur	30	0.1518	0.2525	-0.0085	0.75
5. Passighat	30	0.1893	0.2302	0.2472	0.58
6. Agartala	30	0.2032	0.2278	-0.1384	2.22
7. Imphal	30	0.1744	0.2548	0.2233	0.14
8. Shillong	30	0.1779	0.2374	0.3275	0.62
9. Itanagar	26	0.1546	0.5586	0.5756	1.77
10. Dhubri	22	0.2042	0.0123	0.0151	1.25
11. Jorhat	25	0.1530	-0.1019	-0.1672	1.26
12. Lengpui	13	0.1339	0.2634	0.1863	0.77

(D_i values of all the 12 stations of our study region are less than the critical value 2.757)

3.4.2 Identification of Homogeneous Region

The heterogeneity test H, is derived for LQ-moment in the similar way as discussed in the section 2.5.2. The test can be written as follows:

$$V_1 = \sqrt{\sum_{i=1}^{N} n_i (\hat{\eta}^i - \eta^R)^2 / \sum_1^N n_i} \quad (3.4.2)$$

$$V_2 = \sum_{i=1}^{N} \{n_i [(\hat{\eta}^i - \eta^R)^2 + \left(\hat{\eta}_3^{(i)} - \eta_3^R\right)^2]^{\frac{1}{2}}\} / \sum_{i=1}^{N} n_i \quad (3.4.3)$$

$$V_3 = \sum_{i=1}^{N} \{n_i [(\hat{\eta}_3^{(i)} - \eta_3^R)^2 + \left(\hat{\eta}_4^{(i)} - \eta_4^R\right)^2]^{\frac{1}{2}}\} / \sum_{i=1}^{N} n_i \quad (3.4.4)$$

The regional average LQ-moment ratios are calculated using the following formula

$$\eta^R = \sum_{i=1}^{N} n_i \hat{\eta}^i / \sum_{i=1}^{N} n_i$$

$$\eta_3^R = \sum_{i=1}^{N} n_i \hat{\eta}_3^{(i)} / \sum_1^N n_i \quad (3.4.5)$$

$$\eta_4^R = \sum_{i=1}^{N} n_i \hat{\eta}_4^{(i)} / \sum_1^N n_i$$

where N is the number of stations and $n_i$is the record length at ith station. The heterogeneity test is then defined as

$$H_j = \frac{V_j - \mu_{V_j}}{\sigma_{V_j}} \quad , \quad j = 1,2,3 \tag{3.4.6}$$

where $\mu_{V_j}$and $\sigma_{V_j}$are the mean and standard deviation of simulated V_j values, respectively. The region is acceptably homogeneous, possibly homogeneous and definitely heterogeneous with a corresponding order of LQ-moments according as H<1, 1≤H<2 and H≥2.

From the heterogeneity measures it is found that the values of $H_1 = -1.45, H_2 = 0.87$and $H_3 = 1.77$. It has been observed from heterogeneity measures that, the values of $H_1$and H_2 are less than 1 hence the study region can be considered as a homogeneous one.

3.4.3 Choice of a Distribution

To select the best fit distribution Hosking and Wallis [25] proposed two goodness of fit measures. They are Z-statistic criteria and L-moment ratio diagram. These two tests are also applied in the similar manner for LQ-moment.

(a) Z-statistic Criteria

As in the section 2.5.3, using LQ-moments the Z-statistic for each distribution is calculated as follows

$$Z^{DIST} = (\eta_4^{DIST} - \eta_4^{R})/\sigma_4 \tag{3.4.7}$$

where DIST refers to a particular distribution, η_4^{DIST} is the LQ-kurtosis of the fitted distribution while the standard deviation of η_4^{R} is given by

$$\sigma_4 = \left[(N_{sim})^{-1} \sum_{m=1}^{N_{sim}} (\eta_4^{(m)} - \eta_4^{R})^2\right]^{1/2}$$

$\eta_4^{(m)}$is the average regional LQ-kurtosis and has to be calculated for the mth simulated region. This is obtained by simulating a large number of kappa distribution using Monte Carlo simulations. The value of the Z-statistics is considered to be acceptable at the 90% confidence level if $\left|Z^{DIST}\right| \leq 1.64$. If more than one candidate distribution is acceptable, the one with the lowest $|Z^{DIST}|$ is regarded as the best fit distribution.

(b) LQ-Moment Ratio Diagram

It is a graph of the LQ-skewness and LQ-kurtosis which compares the fit of several distributions on the same graph. As discussed in the section 2.5.3 the expression of η_4 in terms of η_3 for an assumed distribution can be written as

$$\eta_4 = \sum_{k=0}^{8} A_k \eta_3^k \qquad (3.4.8)$$

where the coefficients A_k are tabulated by Bhuyan and Borah [3]. Coefficients are given in Table A.4.

The Z-statistics values of five distribution used for the study region are given in Table 3.2. It has been observed that the Z-statistic values of GEV, GNO, PE3 and GPA distributions are less than 1.64. But that of GPA distribution is the lowest i.e. 0.46. The LQ-moment ratio diagram of our study region is shown in Figure 3.1. It has been observed from Figure 3.1 that the regional average values of LQ-skewness and LQ-kurtosis lies nearer to the GPA distribution curve. Therefore, the GPA distribution is identified as the best fit distribution for rainfall frequency analysis of the region North-East India.

Table 3.2. Z-statistics values of the distribution using LQ-moments

Sl. No.	Name of the probability distribution	Z-Statistics values
1	GLO	2.12
2	GEV	1.50
3	GNO	1.25
4	PE3	0.91
5	**GPA**	**0.43**

($|Z^{GEV}| < 1.64$, $|Z^{GNO}| < 1.64$, $|Z^{PE3}| < 1.64$ and $|Z^{GPA}| < 1.64$. But $|Z^{GPA}|$ is the lowest.)

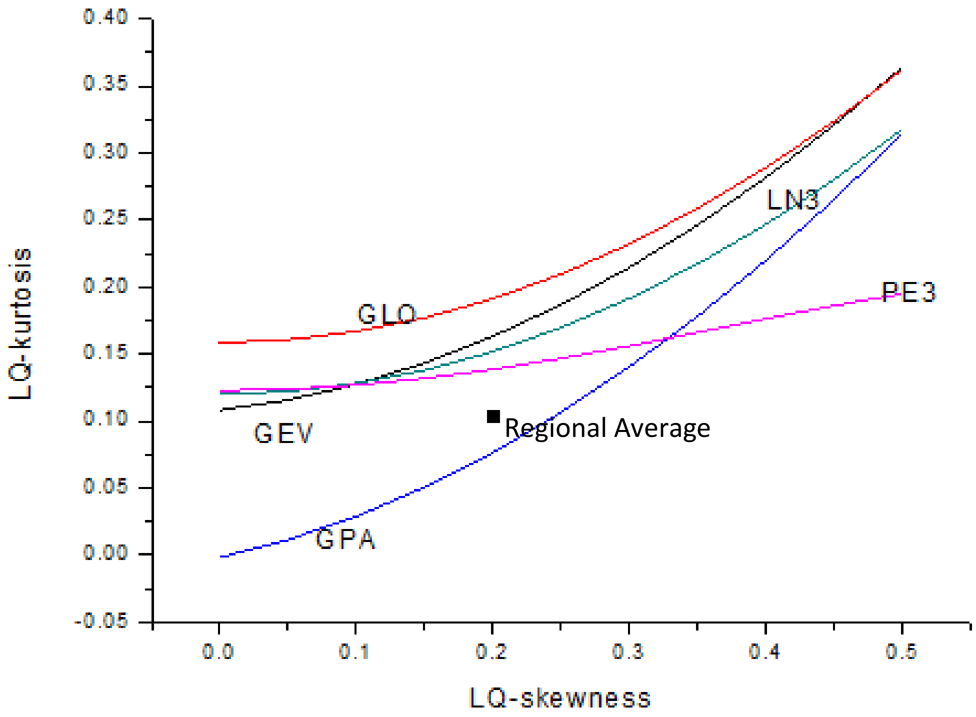

Figure. 3.1 LQ-moment ratio diagram for NE region

(Regional average of LQ-skewness and LQ-kurtosis lies nearer to GPA distribution curve)

3.4.4 Estimation of Frequency Distribution

The regional parameters of the best fit distribution GPA are calculated using the approximation expression given in section 3.3.2. Using the parameters of GPA distribution in the quantile function of GPA distribution given in section 2.3.1, growth factors are calculated.

The parameters of GPA distribution using LQ-moment and quantile estimates using GPA distribution are given in Table 3.3 and Table 3.4 respectively.

Table 3.3 Parameters of the best fitting distribution using LQ-moments

Best fit distribution	Parameters		
	Location (ξ)	Scale (α)	Shape (k)
GPA	0.668	0.511	0.357

Table 3.4 Quantile estimates by using best fitting distribution

Best fit Distribution	Return Periods (in year)				
	2	10	20	100	1000
GPA	0.982	1.471	1.609	1.824	1.979

3.5 Development of Regional Rainfall Frequency Relationship

The index flood procedure discussed in section 2.4 is used to develop regional rainfall frequency relationship. The form of regional rainfall frequency relationship or growth factor for the best fit distribution GPA can be expressed as

$$Q_T = [\xi + \frac{\alpha}{k}\{1 - (1 - F)^k\}] * \mu_i \quad (3.5.1)$$

where Q_T is the maximum rainfall for return period T, $F = 1 - 1/T$, μ_i is the mean annual maximum rainfall of the site, ξ, α and k are the parameters of the GPA distribution. The regional parameters for the GPA distributions are presented in Table 3.3. Substituting the regional values of GPA distribution based on the data of 12 gauged sites the regional rainfall frequency relationship for gauged sites of study area is expressed as:

$$Q_T = [0.668 + 1.431\{1 - (1 - \mathrm{F})^{0.357}\}] * \mu_i \quad (3.5.2)$$

For estimation of rainfall of desired non-exceedance probability for a small to moderate size gauged catchments of the study area, above regional rainfall frequency relationship may be used.

3.6 Conclusion

From discordancy test it is found that all the data of the 12 stations of the study region can be considered for the study. From heterogeneity test it is observed that the 12 stations of the study region form a homogeneous region. Z-statistic criteria and LQ-moment ratio diagram shows that GPA distribution is the best fit distribution for the study region. Parameters of GPA distribution are calculated using LQ-moments. Substituting the regional parameters of GPA distribution in the quantile function of GPA distribution growth factors at different return periods are calculated. Finally using flood index procedure and GPA distribution regional rainfall frequency relationship has been developed.

***