MAUSAM, **72**, 4 (October 2021), 835-846

551.577.36 (540.21)

# Regional frequency analysis of daily maximum rainfall in Haryana

MOHIT NAIN and B. K. HOODA*

*Department of Mathematics and Statistics, School of Chemical Engineering and Physical Sciences, Lovely Professional University, Phagwara, Punjab (India)*

**Department of Mathematics and Statistics, College of Basic Sciences and Humanities, CCS Haryana Agriculture University, Hisar – 125 004 (Haryana), India*

(*Received 4 May 2021, Accepted 14 September 2021*)

**e mail : mohitnain@hau.ac.in**

**सार** – यह पेपर एल-मोमेंट्स का उपयोग करके हरियाणा के 27 रेन गेज स्टेशनों से दैनिक अधिकतम वर्षा के क्षेत्रीय आवृत्ति विश्लेषण के लिए निर्धारित है। चूंकि हरियाणा में वर्षा का वितरण स्थानिक रूप से भिन्न होता है, 27 वर्षा गेज स्टेशनों को वार्ड की क्लस्टरिंग पद्धति का उपयोग करते हुए क्लस्टर C1, C2, और C3 नामक तीन समूहों में बांटा गया है और एल-मोमेंट्स-आधारित विषमता माप (एच) का उपयोग करके समूहों की समरूपता की पुष्टि की गई थी। अच्छाई-की-फिट माप ($Z^{DIST}$) और एल-मोमेंट अनुपात आरेख का उपयोग करते हुए, पांच उम्मीदवार वितरणों के बीच उपयुक्त क्षेत्रीय आवृत्ति वितरण का चयन किया गया था; सामान्यीकृत रसद (जीएलओ), सामान्यीकृत चरम मूल्य (जीईवी), सामान्यीकृत सामान्य (जीएनओ), सामान्यीकृत पारेतो (जीपीए), और प्रत्येक क्लस्टर के लिए पियर्सन टाइप -3 (PE3)। परिणामों से पता चला कि PE3 और GNO क्लस्टर C1 और GEV, PE3 के लिए अच्छे क्षेत्रीय वितरण थे और क्लस्टर C2 के लिए GNO फिट किए गए थे जबकि क्लस्टर C3 के लिए; GLO और GEV अच्छे फिट क्षेत्रीय वितरण थे। प्रत्येक क्लस्टर के लिए मोंटे कार्लो सिमुलेशन का उपयोग करके गणना किए गए अच्छे फिट वितरण सटीकता उपायों के बीच एक मजबूत वितरण का चयन करने के लिए। सिमुलेशन परिणाम ने दिखाया कि PE3 क्लस्टर C1 के लिए मात्रात्मक अनुमान के लिए सबसे अच्छा विकल्प था। Forcluster C2, PE3 एक बड़ी वापसी अवधि के लिए सबसे अच्छा विकल्प था और GEV एक छोटी वापसी अवधि के लिए सबसे अच्छा था। Forcluster C3, GEV मात्रात्मक अनुमान के लिए सबसे उपयुक्त वितरण था। इन मजबूत वितरणों का उपयोग करके प्रत्येक वर्षा गेज स्टेशन पर 2 से 100 वर्ष की वापसी अवधि में वर्षा मात्रा का अनुमान लगाया गया था। ये अनुमानित वर्षा मात्रा नीति निर्माताओं और संरचनात्मक इंजीनियरों द्वारा हाइड्रोलिक संरचनाओं की योजना बनाने और डिजाइन करने के लिए मोटे दिशानिर्देश हो सकते हैं।

**ABSTRACT.** This paper is sets-out for the regional frequency analysis of daily maximum rainfall from the 27 rain gauge stations in Haryana using L-moments. As the distribution of rainfall varies spatially in Haryana, the 27 rain gauge stations are grouped into three clusters namely, cluster C1, C2 and C3 using Ward's clustering method and homogeneity of clusters was confirmed using L-moments-based Heterogeneity measure (H). Using goodness-of-fit measure ($Z^{DIST}$) and L-moment ratios diagram, suitable regional frequency distributions were selected among five candidate distributions; Generalized Logistic (GLO), Generalized Extreme Value (GEV),Generalized Normal (GNO), Generalized Pareto (GPA), and Pearson Type-3 (PE3) for each cluster. Results showed that PE3 and GNO were good fitted regional distribution for the cluster C1 and GEV, PE3 and GNO fitted for cluster C2 while for cluster C3; GLO and GEV were good fitted regional distribution. To select a robust distribution among good fitted distributions accuracy measures calculated using Monte Carlo simulations for each cluster. The simulation result showed that PE3 was the best choice for quantile estimation for cluster C1. For cluster C2, PE3 was the best choicefor a large return period and GEV was best for a small return period. For cluster C3, GEV was the most suitable distribution for quantile estimation. Using these robust distributions rainfall quantiles were estimated at each rain gauge station from 2 to 100 year return periods. These estimated rainfall quantiles may be rough guideline for planning and designing hydraulic structures by policy makers and structural engineers.

**Key words** – Regional frequency analysis, Daily maximum rainfall, L-moments, Return period, Quantiles.

## 1. Introduction

Rainwater is of great importance for agricultural and other living organisms. In Haryana rainfall occurs mostly in the monsoon (June to September). Rainfall is a primary source of water in the Haryana and there are two major rivers Yamuna and Ghaggar flowing through the state. Extreme environmental events like rainfall, floods can harm agriculture as well as human society. Hence, knowledge regarding the magnitude and frequencies of extreme rainfall is essential for many reasons such as sustainable water resource management, construction of hydrologic structures, and planning for water-related emergencies (Durrans, 2004). Frequency analysis provides information regarding the magnitude of extreme events and their frequency of occurrence using suitable probability distributions (Noto and Loggia, 2009).

In practice, frequency analyses of extreme events are carried out in two ways, one is at-site and another is regional. At-site frequency methods do not provide better results due to insufficient or missing data and unequal sample length at sites/raingauge stations. The regional approach helps overcome this problem by combing the data from each sites/raingauge stations into one and also improve the quality of the estimates. The traditional statistical method like at-site has been fully used for such a long time, the method focused on regional frequency analysis is supposed to have a much better and advanced solution. Regional frequency analysis uses data from several sites, which have similar characteristics in a selected region and these data can be combined to produce a single regional frequency distribution that is applicable anywhere in the region after scaling by a site-specific scaling factor (Gabriele & Arnell, 1991).

In statistics L-moments can be used to derive estimators for the parameters of probability distributions, and the existence of higher L-moments only requires that the random variable has finite mean (Hosking, 1990). L-moment estimators are an exact analog to the method of moment estimators but are linear combinations the expected order statistics. It has been observed that L-moment estimators are often superior to method of moments and maximum likelihood for regional studies (Hosking, 1997).

Many studies around the world have used the L-moments approach in regional frequency analysis. Some of them are Devi and Choudhury (2013); Malekinezhad and Garizi (2014); Majumder *et al.* (2015) and Hussain *et al.* (2017). In Haryana past studies related to rainfall were based on at-site analysis, like Hooda (2006), Kumar (2016), Babu and Hooda (2018) and Nain and Hooda (2019a), no study has been carried using a regional approach. The study is different from the previous studies in the state in the sense that we have performed regional frequency approach using the data of daily maximum rainfall from all districts of the Haryana state. The study attempted to find a regional frequency distribution and estimation of daily maximum rainfall quantiles up to 100-year return periods at each raingauge station. The knowledge of the estimated daily maximum rainfall quantiles over raingauge stations is important for proper planning and design of various soil and water conservation structures.

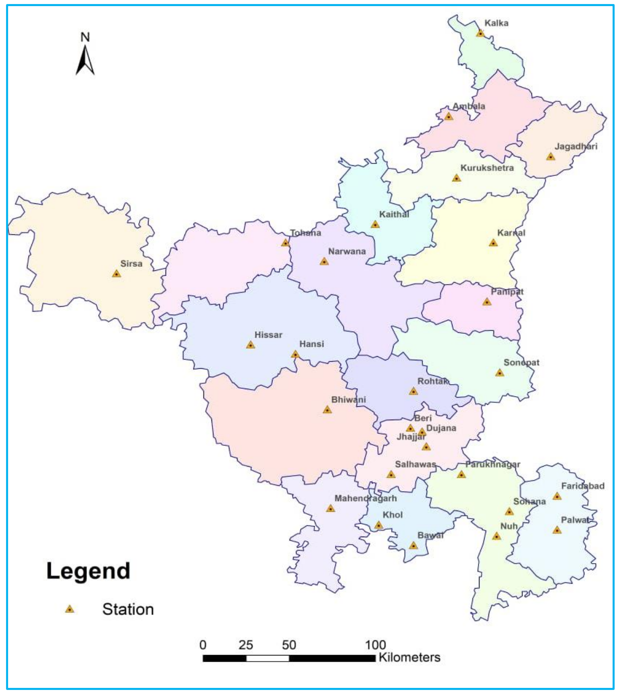



**Fig. 1.** Geographical locations of selected 27 rain gauge stations in Haryana

## 2. Materials and method

The daily rainfall data of 27 rain gauge stations from 1970-2017 were obtained from the National Data Center, India Meteorological Department (IMD) Pune. The daily maximum rainfall series were constructed for each rain gauge station for the analysis. Daily maximum rainfall usually defined as maximum rainfall in one day within each year. For example if $y_1, y_2, \ldots y_{365}$ are daily rainfall values, then the data selection point value is Max $\{y_1, y_2, \ldots y_{365}\}$; where $y_i$ is the daily rainfall of any particular year, for $i$ = 1, 2, 3…...365. The geographical locations of rain gauge stations considered for this study in Haryana have been shown in Fig. 1.

### 2.1. *The L-moments*

The probability-weighted moments (PWMs) of a random variable $X$ with a cumulative distribution function $F(x)$ were defined by Greenwood *et al.* (1979) as

$$M_{p,r,s} = E\left[X^p \{F(X)\}^r \{1 - F(X)\}^s\right] \tag{1}$$

One useful and simple functional case of the PWMs is $\beta_r = M_{1,r,0}$. Thus,

$$\beta_r = \int_0^1 x(F)F^r dF \tag{2}$$

for a distribution that has a quantile function $x(F)$.

The first four L-moments (Hosking, 1990) that are the linear combinations of the PWMs, are given by:

$$\begin{aligned} \lambda_1 &= \beta_0 \\ \lambda_2 &= 2\beta_0 - \beta_0 \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned} \tag{3}$$

L-moments ratios are given by $\tau_2 = \lambda_2/\lambda_1$, $\tau_3 = \lambda_3/\lambda_2$ and $\tau_4 = \lambda_4/\lambda_2$ Where $\tau$ is L-coefficient of variation (L-Cv), $\tau_3$ is L-coefficient of skewness (L-Cs), and $\tau_4$ is L-coefficient of kurtosis (L-Ck).

For an arranged sample $x_{1:n} \le x_{2:n} \le \ldots \le x_{n:n}$, Landwehr *et al*. (1979) defined that the statistic,

$$b_r = n^{-1} \sum_{i=1}^{n} \frac{(i-1)(i-2)\ldots(i-r)}{(n-1)(n-2)\ldots(n-r)} x_{i:n} \tag{4}$$

is an unbiased estimator of $\beta_r$.

Hence, $l_1 = b_0$, $l_2 = 2b_1 - b_0$, $l_3 = 6b_2 - 6b_1 + b_0$ and $l_4 = 20b_3 - 30b_2 - 12b_1 - b_0$ are the first four L-moments of the sample. Similarly $t = \frac{l_2}{l_1}$, $t_3 = \frac{l_3}{l_2}$ and $t_4 = \frac{l_4}{l_2}$ are the sample L-Cv, L-Cs, and L-Ck, respectively.

## 3. Regional frequency analysis procedures based on L-moments

### 3.1. *Formation of homogenous clusters, Discordance and Heterogeneity test*

For the formation of homogeneous clusters, Ward's (1963) clustering method was used. Nain and Hooda (2019b) found that Ward's method provides better results and it has been used by several other authors, Therefore, Ward's clustering method was used for clustering of rain gauge stations in this study.

L-moments based discordancy measure $D_i$ for $i^{th}$ site/station in a group of sites or stations is calculated as:

$$D_i = \frac{1}{3} N \left(a_i - \bar{a}\right)^T C^{-1} \left(a_i - \bar{a}\right) \tag{5}$$

where, $a_i = (t^{(i)} t_3^{(i)} t_4^{(i)})^T$, is a vector containing three sample L-moment ratios for $i^{\text{th}}$ site/station, $N$ is the number of sites/stations in the region, $\bar{a} = N^{-1} \sum_{i=1}^{N} a_i$ denotes the regional average of L-moments ratio and C is the sample covariance matrix as:

$$C = \sum (a_i - \bar{a})(a_i - \bar{a})^T \tag{6}$$

A site/station can be considered discordant from the group if its $D_i \ge D_{\text{Critical}}$ (Hosking and wallis, 1997). To check region's homogeneity a heterogeneity measure (H) is calculated by fitting Kappa distribution to the regional average L-moments ratios and generating 500 equivalent region data by Monte Carlo simulations. This test compares the variability of L-statistics of the real region to the simulated region. The three measures of heterogeneity, namely, $H_j$ ($j$ = 1, 2, 3), which are calculated as:

$$H_j = \frac{V_j - \mu_{v_j}}{\sigma_{V_j}} \tag{7}$$

where, $\mu_{v_j}$ is mean and $\sigma_{v_j}$ is standard deviation of the simulated $V_j$ values.

*V*-statistic as follows:

$$V_1 = \left\{ \sum_{i=1}^{N} n_i \left[ t^{(i)} - t^R \right]^2 \Big/ \sum_{i=1}^{N} n_i \right\}^{1/2} \tag{8}$$

$$V_2 = \left\{ \sum_{i=1}^{N} n_i \left[ (t^{(i)} - t^R)^2 + (t_3^{(i)} - t_3^R)^2 \right]^{1/2} \Big/ \sum_{i=1}^{N} n_i \right\} \tag{9}$$

$$V_3 = \left\{ \sum_{i=1}^{N} n_i \left[ (t_3^{(i)} - t_3^R)^2 + (t_4^{(i)} - t_4^R)^2 \right]^{1/2} \Big/ \sum_{i=1}^{N} n_i \right\} \tag{10}$$

where, and $t^R$, $t_3^R$ and $t_4^R$ are the regional average L-moments ratios. Based on this test, a cluster is acceptably homogeneous if $H_j < 1$, acceptably heterogeneous if $1 \leq H_j < 2$, and definitely heterogeneous if $H_j \geq 2$. According to Hosking and Wallis (1997), the output of cluster analysis needs not to be final, but subjective adjustments can be made to reduce region's heterogeneity by moving one or more sites/stations from one region to another or by reassigning its sites/stations to other regions.

3.2. *Selection of regional frequency distribution*

Goodness-of-fit measures ($Z^{DIST}$) and L-moment ratio diagram are used to select the regional frequency distribution. *L*-moment ratio diagram involves plotting the average sample L-moment ratios ($t_3^R, t_4^R$) as a scatter plot with theoretical L-moment ratios ($\tau_3, \tau_4$) of candidate distributions. Chose the distribution that give closest approximation to the point ($t_3^R, t_4^R$).

The value of $Z^{DIST}$ for each candidate distribution is calculated as

$$Z^{DIST} = \frac{\tau_4^{DIST} - t_4^R + B_4}{\sigma_4} \tag{11}$$

where, $\tau_4^{DIST}$ is a theoretical L-kurtosis of the fitted distribution, $t_4^R$ is an average L-kurtosis calculated from observed data in a given region, $B_4$ is the bias of $t_4^R$ and $\sigma_4$ is the standard deviation of the L-kurtosis ($t_4^R$) obtained from simulation. The value $\left|Z^{DIST}\right| \leq 1.64$ indicates that the distribution is acceptable at 90% confidence interval.

3.3. *Rainfall quantile estimation using Index Flood Procedure based on L-moments*

Index-flood procedures are a convenient way of pooling summary statistics from different data samples. Consider that there are *N* stations in a homogeneous region and station *i* having record length $n_i$. Then $Q_{i,j}$, $i$ = 1, 2,….., N and $j$ = 1, 2,….., $n_i$ denote the observed data and $Q_i(F)$, 0<F<1 be the quantile function of frequency distribution at the $i^{th}$ station. This index-flood procedure makes the following assumptions; like Observations at any given site are identically distributed, Observations at any given site are serially independent and the key assumption of an index-flood procedure is that the sites/stations form a homogeneous region, that is, that the frequency distributions of the *N* sites/stations are identical apart from a site-specific scaling factor, the index flood (Dalrymple, 1960) Formally we can write:

$$Q_i(F) = \mu_i q(F),\ i = 1, 2, \ldots, \mathrm{N}. \tag{12}$$

where $\mu_i$ is the index-flood or the scaling factor for station *i*, usually estimated by sites' mean and *q*(*F*) is the dimensionless quantile function or the regional growth curve of non-exceedance probability *F*. The sample mean at site *i* is estimated by $\hat{\mu} = \sum Q_i \big/ n$ and the dimensionless rescaled data are computed by $q_{i,j} = \dfrac{Q_{i,j}}{\hat{\mu}_i}$, $i$ = 1,2,….., N and $j$ = 1,2,…..,$n_i$, which are the basis for *q*(*F*).

Thus, quantiles at site *i* with non-exceedance probability *F* or different return can be obtained combining the estimates of $\mu_i$ and *q*(*F* ):

$$\hat{Q}_i(F) = \hat{\mu}_i \hat{q}(F) \tag{13}$$

3.4. *Assessment analysis of estimated regional growth curve*

An assessment analysis is conducted by generating a large number of reference regions using Monte Carlo simulation to estimate the accuracy of regional growth curve (Hosking and Wallis, 1997). In simulations, quantile estimates for various non-exceedance probabilities are computed. Let at the m$^{th}$ generation, the regional growth curve and the site *i* quantile estimate for non-exceedance probability $F, \hat{q}^{[m]}(F)$ and $\hat{Q}_i^{[m]}(F)$, respectively, are estimated. Then, at site *i*, the relative error of the estimated regional growth curve as an estimator of the at-site growth curve $q_i(F)$ is $\dfrac{\hat{q}^{[m]}(F) - q_i(F)}{q_i(F)}$. To approximate the bias and relative RMSE, these quantities can be averaged over all *M* repetitions, defined by

$$B_i(F) = \frac{1}{M}\sum_{m=1}^{M} \frac{\hat{q}_i^{[m]}(F) - q_i(F)}{q_i(F)} \tag{14}$$

$$R_i = \left[\frac{1}{M}\sum_{i=1}^{M} \left\{\frac{\hat{q}_i^{[m]}(F) - q_i(F)}{q_i(F)}\right\}^2\right]^{1/2} \tag{15}$$

**TABLE 1**

**L-momentsstatistics for 27 rainfall stations**

| Station name | $l_1$ | $t$ | $t_3$ | $t_4$ |
|---|---|---|---|---|
| Sirsa | 50.312 | 0.284 | 0.228 | 0.235 |
| Narwana | 90.672 | 0.310 | 0.400 | 0.331 |
| Hisar | 55.927 | 0.294 | 0.297 | 0.260 |
| Karnal | 82.508 | 0.254 | 0.211 | 0.160 |
| Ambala | 95.881 | 0.195 | 0.039 | 0.101 |
| Jhajjar | 72.850 | 0.244 | 0.169 | 0.144 |
| Hansi | 43.840 | 0.284 | 0.201 | 0.082 |
| Sonipat | 75.015 | 0.235 | 0.226 | 0.143 |
| Panipat | 68.066 | 0.162 | -0.077 | 0.063 |
| Rohtak | 56.595 | 0.252 | 0.141 | 0.121 |
| Faridabad | 79.310 | 0.346 | 0.185 | 0.083 |
| Kurukshtra | 84.747 | 0.284 | 0.181 | 0.182 |
| Kaithal | 78.669 | 0.331 | 0.310 | 0.085 |
| Bhiwani | 46.355 | 0.263 | 0.267 | 0.111 |
| Khol | 58.169 | 0.340 | 0.366 | 0.262 |
| Farukhnagar | 65.423 | 0.322 | 0.194 | 0.162 |
| Mahendragarh | 57.452 | 0.269 | 0.205 | 0.187 |
| Palwal | 63.742 | 0.245 | 0.212 | 0.283 |
| Tohana | 65.079 | 0.251 | 0.134 | 0.087 |
| Sohana | 83.321 | 0.193 | 0.050 | 0.145 |
| Bawal | 85.889 | 0.257 | 0.205 | 0.271 |
| Jagadhari | 120.617 | 0.248 | 0.131 | 0.052 |
| Dujana | 69.440 | 0.275 | 0.213 | 0.194 |
| Salhawas | 56.644 | 0.304 | 0.185 | 0.150 |
| Nuh | 97.358 | 0.273 | 0.242 | 0.228 |
| Kalka | 93.170 | 0.269 | 0.232 | 0.099 |
| Beri | 69.208 | 0.316 | 0.232 | 0.193 |

For a region, the relative bias (Rel.bias), absolute relative bias (Abs.rel.bias) and relative RMSE (rel.RMSE) are given as :

$$B^R(F)=\frac{1}{N}\sum_{i=1}^{N}B_i(F)\,,\quad A^R(F)=\frac{1}{N}\sum_{i=1}^{N}|\,B_i(F)\,| \text{ and}$$

$R^R(F)=\frac{1}{N}\sum_{i=1}^{N}R_i(F)$ respectively.

For a particular non-exceedance probability $F$ it may be found that 5% of the simulated values of $\hat{q}_i(F)/q_i(F)$ lie below some value, $L_{0.05}(F)$ , whereas 5% lie above some value $U_{0.05}(F)$ , So we can write

$$L_{0.05}(F)\le\frac{\hat{q}(F)}{q(F)}\le U_{0.05}(F) \tag{16}$$

Or inverting the expression obtained

$$\frac{\hat{q}(F)}{U_{0.05}(F)}\le q(F)\le\frac{\hat{q}(F)}{L_{0.05}(F)} \tag{17}$$

These bounds are considered as "90% error bounds" or confidence interval limit for regional growth curves.

Analysis was carried out using R Studio version 1.2.1335, for L-moment analysis, the *R* package lmom RFA was used (Hosking and Wallis, 2009).

## 4. Results and discussion

Before applying regional frequency analysis, we checked some basic assumptions of data series such as stationarity, independence, and randomness using the Mann-Kendall test, autocorrelation plot, and run test, respectively. It was observed that data fulfill the assumption and can be used for regional frequency analysis. The sample L-moments ratios for each station have been presented in Table 1.

### 4.1. *Cluster analysis and heterogeneity measures*

For the formation of the homogeneous regions/clusters, the hierarchical cluster analysis technique using Ward's method was followed. Ward's method of cluster analysis was applied on the mean monthly rainfall of the rain gauge stations and the resulting dendrogram with three clusters has been shown in Fig. 2.

**TABLE 2**

**Results of Heterogeneity measures for three clusters**

| Clusters | Before adjustment of clusters | | | After adjustment of clusters | | |
|---|---|---|---|---|---|---|
| | $H_1$ | $H_2$ | $H_3$ | $H_1$ | $H_2$ | $H_3$ |
| C1 | 0.98 | 0.98 | 0.87 | 0.98 | 0.97 | 0.87 |
| C2 | 1.48 | -0.94 | -0.83 | -0.33 | -2.55 | -1.21 |
| C3 | 4.76 | 3.31 | 3.42 | 0.94 | 0.44 | 0.57 |

The heterogeneity measures in Table 2 showed that cluster C1 is homogeneous whereas cluster C2 and cluster C3 are heterogeneous based on H criterion.

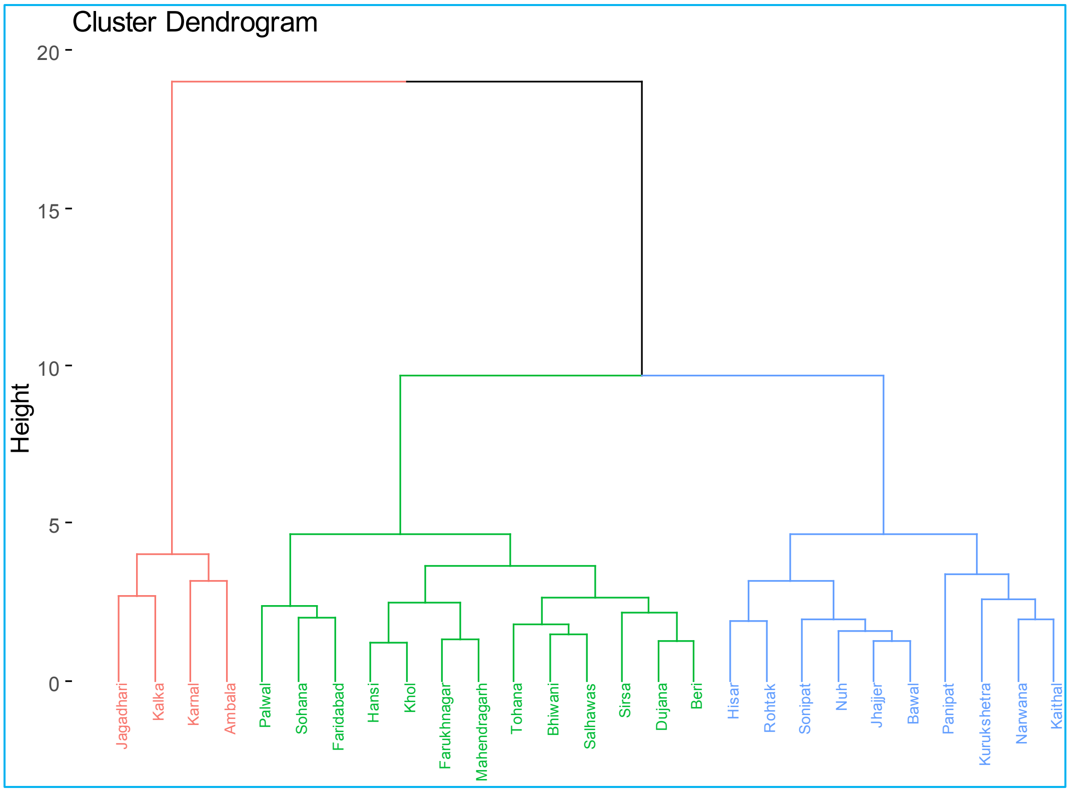



**Fig. 2.** Dendrogram by Ward's method

**TABLE 3**

**Discordancy measure value for all stations in three clusters**

| Cluster | Rain gauge stations ($D_i$) | $D_{Critical}$ |
|---|---|---|
| C1 | Ambala (1.00), Karnal (1.00), Jagadhari (1.00), Kalka (1.00) | 1.33 |
| C2 | Sirsa (0.56), Hansi (2.27), Faruknagar (0.87), Faridabad (1.48), Mahendragarh (0.20), Jhajjar (0.77), Palwal (1.57), Tohana (1.89), Dujana (0.26), Salhawas (0.39), Beri (0.76) | 2.63 |
| C3 | Hisar (0.20), Sonipat (1.09), Rohtak (0.77), Nuh (0.05), Bawal (1.57), Khol (1.20), Bhiwani (0.24), Kurukshetra (1.15), Narwana (1.48), Sohana (1.57) | 2.49 |

4.2. *Heterogeneity measures*

To reduce the heterogeneity of the clusters subjective adjustments were made and on the basis of the discordancy test result, three rain gauge stations namely Khol, Bhiwani, and Sohana in cluster C2 were removed and assigned to the cluster C3, Similarly from cluster C3, Jhajjar, Panipat, and Kaithal were removed and Jhajjar was assigned to adjacent cluster C2, while stations Panipat and Kaithal it was not possible to assign them to any cluster because assigning these stations to any cluster would cause them to be heterogeneous. So these two raingauge stations (Panipat and Kaithal) were removed from further analysis. Thus after adjustments of initial clusters, the final clusters were found acceptably homogeneous ($H_i$<1, Table 2).

4.3. *Discordancy measure, the goodness of-fit-test (Z-test) and parameter estimation*

Discordancy measure ($D_i$) was computed for each raingauge station and it was found that in each

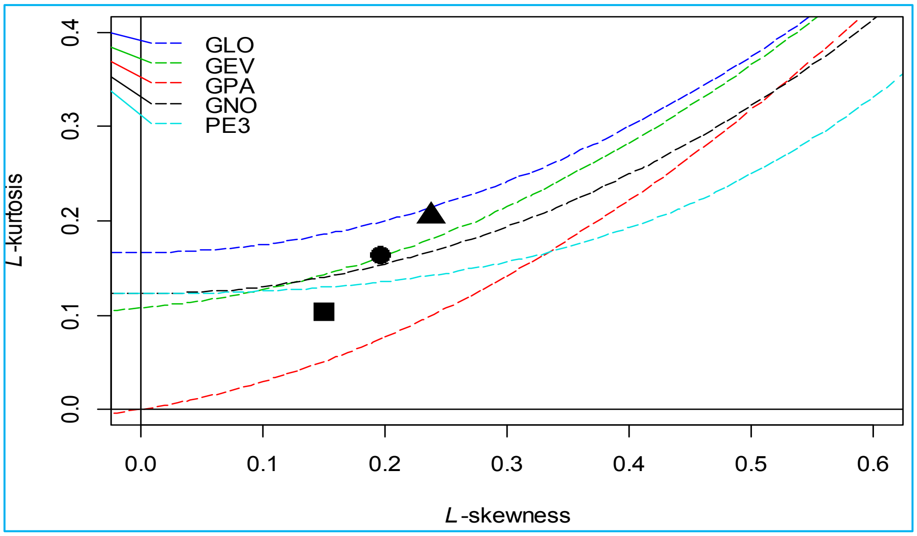



**Fig. 3.** L-moment ratio diagram with regional average L-skewness and L-kurtosis. Filled square for cluster C1; filled circle for cluster C2; Filled triangle for cluster C3

**TABLE 4**

**Goodness-of-fit measure, parameter estimation and quantile estimation**

| Cluster | Distributions | parameters | | | | quantile estimates, $\hat{q}(F)$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $Z^{DIST}$ | α | ξ | $k$ | 0.9□ | 0.95 | 0.99 | 0.995 | 0.998 |
| C1 | PE3 | 1.09** | 1.00 | 0.448 | 0.988 | 1.709 | 2.008 | 2.660 | 2.929 | 3.276 |
| | GNO | 1.57* | 0.928 | 0.415 | -0.335 | 1.601 | 1.840 | 2.351 | 2.559 | 2.826 |
| C2 | PE3 | -1.450* | 1.00 | 0.529 | 1.186 | 1.688 | 2.005 | 2.757 | 3.093 | 3.551 |
| | GEV | -0.10** | 0.755 | 0.396 | -0.041 | 1.694 | 2.005 | 2.726 | 3.044 | 3.475 |
| | GNO | -0.53* | 0.900 | 0.473 | -0.405 | 1.709 | 2.008 | 2.660 | 2.929 | 3.276 |
| C3 | GLO | 0.11** | 0.897 | 0.246 | -0.238 | 1.606 | 1.945 | 2.946 | 3.502 | 4.391 |
| | GEV | -1.32* | 0.758 | 0.351 | -0.103 | 1.647 | 1.978 | 2.823 | 3.230 | 3.813 |

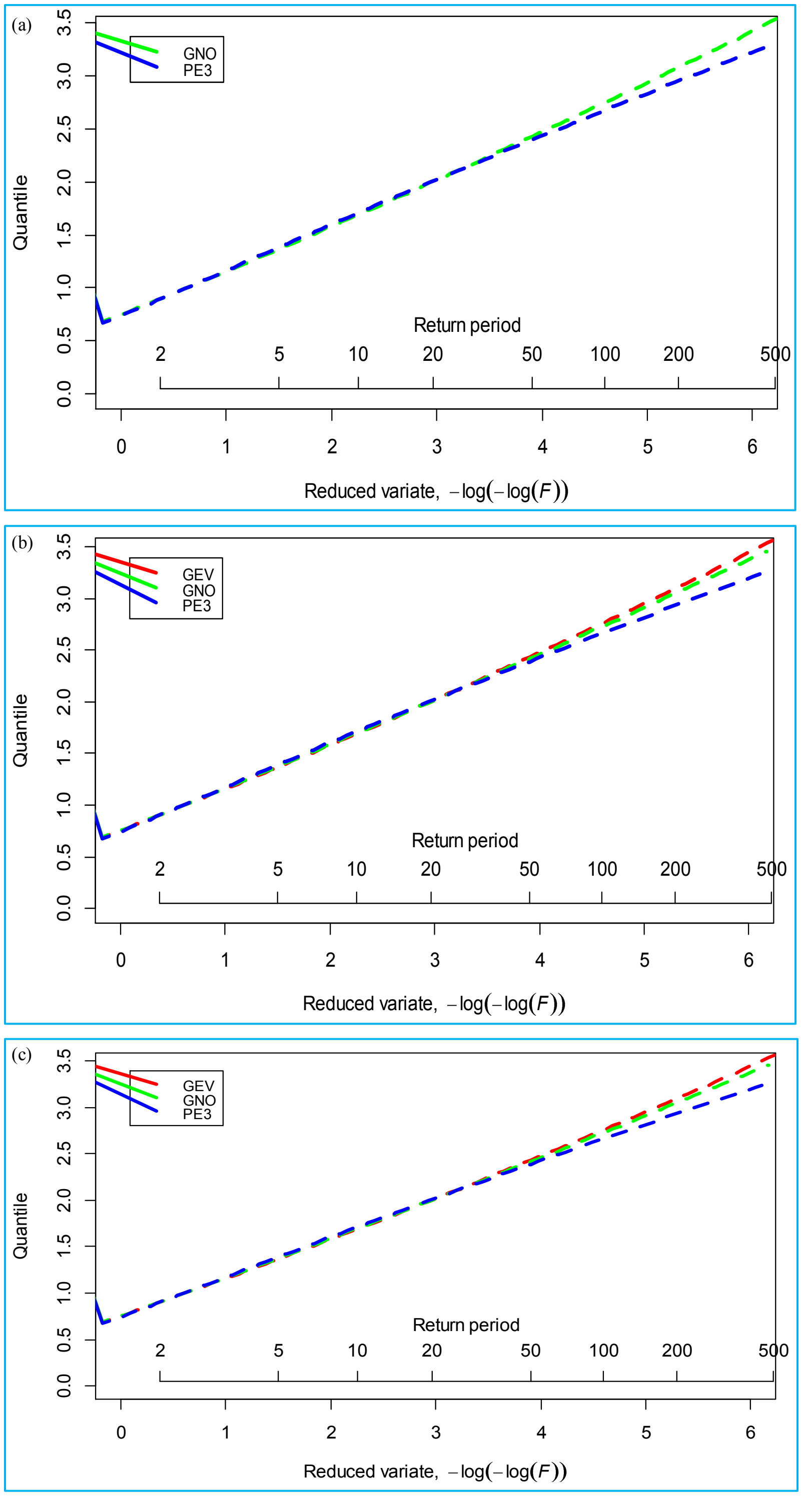
(*implies good fitted distributions, ** implies best-fit distributions and □ denote the non-exceedance probability F)

homogeneous cluster, all raingauge stations have $D_i$ values less than the critical value for the corresponding clusters and which means that there were no discordant stations in these clusters (Table 3).

Results of the goodness-of-fit test ($Z^{DIST}$), estimated parameter and estimated quantiles for all three clusters have been presented in Table 4. For cluster C1, PE3 and GNO were the good fitted distribution and the best fit as PE3 because it has the lowest $|Z^{DIST}|$ value. Forcluster C2, GLO, GEV, and GNO were good fitted distribution but the best fitted was GEV due to the smallest $|Z^{DIST}|$ value. For cluster C3, GEV and GLO have fitted distributions but GLO was best fit due to the lowest $|Z^{DIST}|$ value.

L-moment ratio diagrams for three clusters have been displayed in Fig. 2. It was observed that PE3 distribution is in close agreement with the regional average for clusterC1, for cluster C2, GEV is appropriate and for cluster C3, GLO distribution found suitable as displayed in Fig. 3.

The value of the estimated quantiles given in Table 4 can be explained as, for example for cluster C2, $\hat{q}_{GEV}(0.99)$ = 2.726 is the amount of rainfall which will




**Figs. 4(a-c).** Regional growth curves; (a) for cluster C1, (b) for cluster C2 and (c) for cluster C3

happen on an average once in 100 years and is 2.726 times larger than its average for all rain gauge stations in homogeneous cluster C2 for the given return period. We developed a regional growth curve for three homogeneous clusters. Regional frequency analysis assumes that stations in a homogeneous region/cluster have a common frequency distribution and representation of this common

**TABLE 5**

**At-station quantile estimation**

| Station | **0.5 *2 | 0.8 5 | 0.9 10 | 0.95 20 | 0.98 50 | 0.99 100 |
|---|---|---|---|---|---|---|
| Karnal | 76.5 | 110.6 | 132.1 | 151.8 | 176.3 | 194.0 |
| Ambala | 88.9 | 128.5 | 153.5 | 176.5 | 204.9 | 225.5 |
| Kalka | 86.4 | 124.9 | 149.1 | 171.5 | 199.1 | 219.1 |
| Jagadhari | 111.9 | 161.6 | 193.1 | 222.0 | 257.8 | 283.6 |
| Sirsa | 45.4 | 68.3 | 84.0 | 99.6 | 120.5 | 136.7 |
| Hansi | 39.6 | 59.5 | 73.2 | 86.8 | 105.0 | 119.1 |
| Farukhnagar | 59.1 | 88.8 | 109.3 | 129.5 | 156.7 | 177.8 |
| Mahendragarh | 51.9 | 78.0 | 96.0 | 113.7 | 137.6 | 156.1 |
| Faridabad | 71.2 | 108.1 | 133.4 | 158.4 | 191.9 | 217.8 |
| Palwal | 57.6 | 86.5 | 106.5 | 126.2 | 152.7 | 173.2 |
| Tohana | 58.8 | 88.3 | 108.7 | 128.8 | 155.9 | 176.9 |
| Dujana | 62.7 | 94.2 | 116.0 | 137.5 | 166.3 | 188.7 |
| Salhawas | 51.2 | 76.9 | 94.6 | 112.1 | 135.7 | 153.9 |
| Beri | 62.5 | 93.9 | 115.6 | 137.0 | 165.8 | 188.1 |
| Jhajjar | 65.8 | 98.9 | 121.6 | 144.2 | 174.5 | 197.9 |
| Hisar | 50.2 | 72.8 | 89.8 | 108.8 | 138.2 | 164.7 |
| Sonipat | 67.3 | 97.6 | 120.5 | 145.9 | 185.4 | 221.0 |
| Rohtak | 50.8 | 73.6 | 90.9 | 110.1 | 139.9 | 166.7 |
| Nuh | 87.4 | 126.6 | 156.4 | 189.4 | 240.6 | 286.8 |
| Bawal | 77.1 | 111.7 | 138.0 | 167.1 | 212.2 | 253.0 |
| Kurukshetra | 76.0 | 110.2 | 136.1 | 164.9 | 209.4 | 249.6 |
| Narwana | 81.4 | 117.9 | 145.7 | 176.4 | 224.1 | 267.1 |
| Bhiwani | 41.6 | 60.3 | 74.5 | 90.2 | 114.6 | 136.6 |
| Khol | 52.2 | 75.7 | 93.4 | 113.2 | 143.7 | 171.4 |
| Sohana | 74.8 | 108.4 | 133.8 | 162.1 | 205.9 | 245.4 |

**Non-Exceedance Probability F, *Return period (T) in years

distribution is the regional growth curve, which have been presented in Fig. 4.

For cluster C1, the growth curve is looking similar to the return period of 100 years and after that, there is a small difference between PE3 and GNO curves Fig. 4(a). For cluster C2, the growth curve shows the same behaviour for GEV and GNO distributions up to 500 return periods with little differences also GLO show the same behaviour up to 50 years but beyond 50 years return period GLO curve go in an upward direction, which means that quantiles estimate by GLO are little high [Fig. 4(b)]. Similarly, for cluster C3, GLO and GEV show the same pattern in quantiles up to 50 years return period, but as the return period increases up to 500 years, GLO moves in an upward direction with high quantile estimates compared to GEV [Fig. 4(c)].

4.4. *Assessment analysis of regional estimates*

*Assessment results for cluster C1* : As the GNO and PE3 are two suitable distributions for this cluster and we conducted a simulation analysis based on these two distributions. The algorithm for simulation defined by Hosking and Wallis (1997, Table 6.1) was used. For cluster C1, L-Cv values varying from 0.195 to 0.269 and average L-Cs 0.163. Rel.bias, Abs.rel.bias,rel.RMSE lower and upper bounds are calculated for the regional growth curve for different non-exceedance probabilities which are given in Table 7. For each candidate distribution, 10,000 realizations and 100 simulations are set to perform this algorithm for cluster C1. First, this procedure was performed for PE3 distribution, after it was performed for GNO distribution. In Table 6 simulation results for the cluster C1, show that the Abs.rel.bias and rel. RMSE for GNO and PE3 are almost equal performance for a return period of 2, 5, 10, 20 and 50 but when we see for large return period for example 100, 500, etc., PE3 produces low Abs.rel.bias and rel.RMSE compared to GNO. Also, error bounds of PE3 distribution are narrower than GNO at high return periods. So we can say that PE3 distribution is the best choice for quantile estimation for the large return period incluster C1.

*Assessment results for cluster C2* : In Table 7 simulation results for the regional growth have been presented. These results show that for GEV at low return periods 2, 5, 10 and 20 it's rel. RMSE and Abs.rel.bias are slightly low compared to GNO and PE3. Alternatively, for large return periods, rel. RMSE and Abs.rel.bias of PE3 is low compared to GNO and GEV distributions. Based on simulation results, it concludes that for large return periods, PE3 is the most appropriate distribution for quantile estimation and GEV is the most appropriate choice for low return periods.

*Assessment results for cluster C3* : For this region, the simulation results have been presented in table 8. These results show that GEV has relatively lo Abs.rel. bias and rel RMSE compared to GLO distribution for both low and high return periods. So it concludes that for cluster C3, GEV is the most suitable distribution for quantile estimation.

Quantiles for individual stations were estimated upto 100 years return periods by multiply the regional growth curve of robust distributions with stations' average of daily maximum rainfall (Table 5).

**TABLE 6**

**Accuracy measures results for regional growth curves incluster C1**

| | F | 0.5 | 0.8 | 0.9 | 0.95 | 0.98 | 0.99 | 0.995 | 0.998 | 0.999 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 2 | 5 | 10 | 20 | 50 | 100 | 200 | 500 | 1000 |
| PE3 | Rel.bias | 0.002 | 0.002 | 0.002 | 0.002 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| | Abs.rel.bias | 0.016 | 0.027 | 0.041 | 0.052 | 0.063 | 0.070 | 0.076 | 0.083 | 0.088 |
| | rel. RMSE | 0.018 | 0.031 | 0.048 | 0.061 | 0.077 | 0.083 | 0.090 | 0.099 | 0.104 |
| | LCL | 0.899 | 1.316 | 1.550 | 1.756 | 1.993 | 2.169 | 2.337 | 2.552 | 2.710 |
| | UCL | 0.957 | 1.373 | 1.668 | 1.953 | 2.310 | 2.587 | 2.863 | 3.225 | 3.498 |
| GNO | Rel.bias | 0.002 | 0.002 | 0.002 | 0.003 | 0.003 | 0.004 | 0.004 | 0.005 | 0.006 |
| | Abs.rel.bias | 0.016 | 0.026 | 0.040 | 0.052 | 0.064 | 0.075 | 0.084 | 0.095 | 0.103 |
| | rel. RMSE | 0.018 | 0.031 | 0.048 | 0.062 | 0.078 | 0.089 | 0.099 | 0.111 | 0.120 |
| | LCL | 0.901 | 1.301 | 1.538 | 1.731 | 1.979 | 2.144 | 2.292 | 2.490 | 2.640 |
| | UCL | 0.962 | 1.359 | 1.656 | 1.946 | 2.334 | 2.631 | 2.948 | 3.376 | 3.714 |

where Rel.bias, Abs.rel.bias and Rel. RMSE is the regional average relative bias [$B^R$ ($F$)], absolute relative bias [$A^R$ ($F$)] and relative RMSE [$R^R$ ($F$)] respectively. LEB = Lower error bound, UEB = Upper error bound

**TABLE 7**

**Accuracy measures results for regional growth curves incluster C2**

| | F | 0.5 | 0.8 | 0.9 | 0.95 | 0.98 | 0.99 | 0.995 | 0.998 | 0.999 |
|---|---|---|---|---|---|---|---|---|---|---|
| GEV | Rel.bias | 0.003 | 0.002 | 0.001 | 0.001 | -0.001 | -0.002 | -0.003 | -0.005 | -0.006 |
| | Abs.rel.bias | 0.015 | 0.027 | 0.041 | 0.051 | 0.063 | 0.071 | 0.079 | 0.089 | 0.098 |
| | rel. RMSE | 0.018 | 0.031 | 0.046 | 0.05 | 0.074 | 0.084 | 0.093 | 0.105 | 0.115 |
| | LEB | 0.879 | 1.347 | 1.653 | 1.939 | 2.301 | 2.569 | 2.832 | 3.171 | 3.425 |
| | UEB | 0.920 | 1.386 | 1.726 | 2.085 | 2.593 | 3.003 | 3.441 | 4.066 | 4.574 |
| GNO | Rel.bias | 0.003 | 0.002 | 0.001 | 0.001 | 0.001 | 0.000 | -0.001 | -0.001 | -0.002 |
| | Abs.rel.bias | 0.016 | 0.027 | 0.041 | 0.051 | 0.062 | 0.069 | 0.075 | 0.082 | 0.087 |
| | rel. RMSE | 0.019 | 0.032 | 0.048 | 0.06 | 0.073 | 0.081 | 0.089 | 0.097 | 0.104 |
| | LEB | 0.877 | 1.356 | 1.658 | 1.938 | 2.294 | 2.560 | 2.826 | 3.177 | 3.448 |
| | UEB | 0.920 | 1.391 | 1.735 | 2.086 | 2.565 | 2.942 | 3.334 | 3.879 | 4.311 |
| PE3 | Rel.bias | 0.002 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| | Abs.rel.bias | 0.016 | 0.028 | 0.042 | 0.051 | 0.06 | 0.065 | 0.07 | 0.075 | 0.078 |
| | rel. RMSE | 0.019 | 0.032 | 0.049 | 0.060 | 0.071 | 0.078 | 0.083 | 0.089 | 0.092 |
| | LEB | 0.876 | 1.372 | 1.671 | 1.943 | 2.280 | 2.523 | 2.760 | 3.065 | 3.290 |
| | UEB | 0.918 | 1.405 | 1.752 | 2.086 | 2.513 | 2.829 | 3.140 | 3.547 | 3.851 |

**TABLE 8**

**Accuracy measures results for regional growth curves incluster C3**

| | F | 0.5 | 0.8 | 0.9 | 0.95 | 0.98 | 0.99 | 0.995 | 0.998 | 0.999 |
|---|---|---|---|---|---|---|---|---|---|---|
| GLO | Rel.bias | 0.006 | 0.004 | 0.003 | 0.002 | -0.001 | -0.003 | -0.005 | -0.009 | -0.012 |
| | Abs.rel.bias | 0.022 | 0.036 | 0.058 | 0.076 | 0.096 | 0.110 | 0.122 | 0.138 | 0.150 |
| | rel. RMSE | 0.026 | 0.041 | 0.068 | 0.090 | 0.114 | 0.130 | 0.145 | 0.164 | 0.178 |
| | LEB | 0.868 | 1.277 | 1.566 | 1.870 | 2.305 | 2.677 | 3.098 | 3.757 | 4.340 |
| | UEB | 0.924 | 1.316 | 1.657 | 2.061 | 2.719 | 3.341 | 4.090 | 5.337 | 6.523 |
| GEV | Rel.bias | 0.004 | 0.003 | 0.004 | 0.003 | 0.003 | 0.002 | 0.001 | -0.001 | -0.002 |
| | Abs.rel.bias | 0.022 | 0.038 | 0.061 | 0.077 | 0.095 | 0.106 | 0.117 | 0.130 | 0.140 |
| | rel. RMSE | 0.026 | 0.044 | 0.070 | 0.090 | 0.112 | 0.126 | 0.139 | 0.155 | 0.167 |
| | LEB | 0.860 | 1.308 | 1.619 | 1.919 | 2.346 | 2.672 | 3.005 | 3.492 | 3.865 |
| | UEB | 0.903 | 1.343 | 1.686 | 2.072 | 2.665 | 3.178 | 3.752 | 4.639 | 5.405 |

## 6. Summary and conclusions

In this study, regional frequency analysis based on the L-moments approach of daily maximum rainfall using 27 rain gauge stations was conducted. All 27 raingauge were grouped into three cluster, namely, cluster C1, C2, and C3. The cluster C2 and cluster C3 did not satisfy the H-statistic criterion. After some refinement in initial clusters, the final clusters were found acceptably homogeneous. Based on the L-moments ratio diagram and $Z^{DIST}$, criteria showed that for cluster C1; PE3 and GNO were good fitted, for cluster C2; GEV, PE3 and GNO were good fitted while for cluster C3; GLO and GEV were good fitted distribution. The regional growth curve was developed for three homogeneous regions for all fitted distribution.

As each cluster has more than one fitted distribution, so an assessment analysis was conducted based on Monte Carlo simulations for assessing the accuracy of the estimated quantiles. From this simulation accuracy measures like relative bias, absolute relative bias and relative root mean square error for various return periods of quantiles was calculated for assessment in each cluster. Based on accuracy measures, it was concluded that PE3 was best one for cluster C1 and C2 for quantile estimation. For cluster C3, GLO was best choice for quantile estimation. Using these robust distributions rainfall quantiles were estimated at each station. Identifying the robust distribution amount of daily maximum rainfall data could have a wide range of applications in agriculture like crop planning and water related projects in the state.

The results obtained from the analysis may be useful for engineering planning and designing safe hydrological structures in that future year daily maximum rainfall events can be predicted. However the selected best distributions were used to predict daily maximum rainfall quantiles for the 27 stations for return periods of 2, 5, 10, 20, 50, 100 years. The results showed that RMSE values and 90% error bounds of estimated rainfall quantiles is relatively low when return periods are less than 100 years. But, for higher return periods, rainfall estimates should be treated with caution. It is recommended that 2 to 100 years are the sufficient return period for soil and water conservation measures, irrigation and drainage-related works. A policymaker, conducting a risk analysis for a 50 year plan could use the 50 year return period result of estimated daily maximum rainfall to determine risks, damage projections, etc.

*Acknowledgments*

We are extremely grateful to the National Data Center, India Meteorological Department (IMD), Pune for providing me necessary data for this study. Our special thanks to Dr. J. R. M. Hosking for technical help and assistance in data analysis.

*Disclaimer* : The contents and views expressed in this research paper/article are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

**References**


Babu, V. B. and Hooda, B. K., 2018, "Fuzzy majority approach for modeling spatial and temporal distributions of daily rainfall in western zone of Haryana", *International Journal of Agricultural and Statistical Sciences*, **14**, 1, 57-67.

Dalrymple, T., 1960, "Flood frequency analysis. Manual of hydrology : Part 3. Flood flow techniques", Water Supply Paper 1543-A. United States Geological Survey, Reston, VA, 80.

Devi, T. A. and Choudhury, P., 2013, "Extreme Rainfall Frequency Analysis for Meteorological Sub-Division 4 of India Using L-moments", *International Journal of Civil and Environmental Engineering*, **7**, 12, 945-950.

Durrans, S. R. and Kirby, J. T., 2004, "Regionalization of extreme precipitation estimates for the Alabama rainfall atlas", *Journal of Hydrology*, **295**, 1-4, 101-107.

Gabriele, S. and Arnell, N., 1991, "A hierarchical approach to regional flood frequency analysis", *Water Resources Research*, **27**, 6, 1281-1289.

Greenwood, J. A., Landwehr, J. M. N., Matalas, C. and Wallis, J. R., 1979, "Probability weighted moments : Definition and relation to parameters of several distributions expressable in inverse form", *Water Resources Research*, **15**, 5, 1049-1054.

Hooda, B. K., 2006, "Probability Analysis of Monthly Rainfall for Agricultural Planning at Hisar", *Indian Journal of Soil Conservation*, **34**, 1, 12-14.

Hosking, J. R. M. and Wallis, J. R., 1997, "Regional frequency analysis: An approach based on L-Moments", Cambridge University Press, United Kingdom.

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.

Hosking, J. R. M., 2009, "Regional frequency analysis using L-moments, lmomRFA R package", version 2.2. Available at : http://CRAN.R-project.org/package=lmomRFA.

Hussain, Z., Shahzad, M. N. and Abbas, K., 2017, "Application of regional rainfall frequency analysis on seven sites of Sindh, Pakistan", *KSCE Journal of Civil Engineering*, **21**, 5, 1812-1819.

Kumar M., Shekhar, C. and Manocha, V., 2016, "Maximum rainfall probability distributions pattern in Haryana - A case study", *Journal of Applied and Natural Science*, **8**, 4, 2029-2036.

Landwehr, J. M., Matalas, N. C. and Wallis, J. R., 1979, "Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles", *Water Resources Research*, **15**, 5, 1055-1064.

Majumder, A., Patil, S. G., Noman, M. D. and Biswas, S., 2015, "Application of L-moments for regional frequency analysis of maximum monthly rainfall in West Bengal, India", *Mausam*, **66**, 2, 273-283.

Malekinezhad, H. and Garizi, A. Z., 2014, "Regional frequency analysis of daily rainfall extremes using L-moments approach", *Atmosfera*, **27**, 4, 411-427.

Nain, M. and Hooda, B. K., 2019a, "Probability and Trend Analysis of Monthly Rainfall in Haryana", *International Journal of Agricultural and Statistical Sciences*, **15**, 1, 221-229.

Nain, M. and Hooda, B. K., 2019b, "Identification of homogeneous rainfall stations in Haryana", *JEBAS*, **7**, 5, 452-462.

Noto, L. V. and La Loggia, G., 2009, "Use of L-moments approach for regional flood frequency analysis in Sicily, Italy", *Water Resources Management*, **23**, 11, 2207-2229.

Ward Jr., J. H., 1963, "Hierarchical grouping to optimize an objective function", *Journal of American Statistical Association*, **58**, 236-244.