# MODELLING OF EXTREME RAINFALL IN PUNJAB: PAKISTAN USING BAYESIAN AND FREQUENTIST APPROACH

AHMAD, I.[1, 2, 4*] – AHMAD, T.[3] – ALMANJAHIE, IBRAHIM M.[1,2]

[1]*Department of Mathematics, College of Science, King Khalid University, Abha 62529, Kingdom of Saudi Arabia*

[2]*Statistical Research and Studies Support Unit, King Khalid University, Abha 62529, Kingdom of Saudi Arabia*

[3]*Department of Statistical Sciences, University of Padova, Padova 35121, Italy*

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

*\*Corresponding author*
*e-mail: ishfaq.ahmad@iiu.edu.pk*

**Abstract.** In this study we compared the efficiency of frequentist approach with Bayesian approach by carrying out extreme value analysis of Annual Maximum Daily Rainfall (AMDR). For frequentist frequency analysis of AMDR, we used the data of one station i.e. Lahore in Punjab province, Pakistan while for Bayesian analysis we used the data of three other neighboring stations as prior information. During frequentist approach, Generalized Extreme Value (GEV) was found to be a best-fit distribution on the data. In frequentist method, the parameters of GEV distribution were estimated using Maximum Likelihood Estimation (MLE), while in the Bayesian framework the Markov Chain Monte Carlo (MCMC) simulation technique along with Metropolis-Hasting algorithm and Gibbs sampler were implemented. Findings of this study indicate that despite the asymptotic requirements of the MLE, its performance can be enhanced by adopting the Bayesian MCMC paradigm. In order to acquire the posterior densities of GEV parameters, non-informative and informative priors based on the historical data of surrounding weather stations were employed. The result of Bayesian MCMC might be affected by the choice of priors. In addition, the performance of the parameters estimation methods was verified by employing several robustness measures. Robustness measures results proved that the Bayesian MCMC method performed better than MLE in estimating GEV parameters and future return levels. Therefore, the findings of these analyses could be helpful in adopting proper flood protection measures and designing infrastructures of culverts, buildings, bridges and dams in the region.
**Keywords:** *annual maximum daily rainfall, generalized extreme value, the Markov Chain Monte Carlo, maximum likelihood estimation, Bayesian paradigm, priors specification*

**Abbreviations:** AMDR: Annual Maximum Daily Rainfall; GEV: Generalized Extreme Value; MLE: Maximum Likelihood Estimation; MCMC: Markov Chain Monte Carlo; UNFCC: United Nations Framework on Climate Change; FFC: Federal Flood Commission; POT: Peak over a Threshold; GP: Generalized Pareto; GLO: Generalized Logistic; L-moment: Linear moment; mm: millimeters; SE: Standard Error; RRMSE: Relative Root Mean Square Error; RASE; Relative Absolute Square Error; PPCC: Probability Plot Correlation Coefficient; CV: Coefficient of Variation; SD: Standard Deviation; FA: Frequency Analysis; CI: Confidence Intervals; GB: Gibbs Sampler; M-H: Metropolis-Hasting; MLEs: Maximum Likelihood Estimates

## Introduction

Extreme rainfall events are frequently connected with climatic fluctuations which may cause a series of natural disasters such as heavy floods and landslides. According to the United Nations Framework on Climate Change (UNFCC) the human lives, animal lives, and water resources, agricultural lands, food security, and coastal zones will be heavily affected

due to climatic changes in Asia. Pakistan is situated in the western part of the Indian subcontinent, with Afghanistan and Iran on the western borders, India on the eastern borders, and the Arabian Sea to the south. It is an agricultural country. The country with 70% of its population depends on agriculture and its products (Faisal and Sadiq, 2009). Most of the agricultural lands and other living organisms depend on rain-water. Rainfall occurs in Pakistan, not only in summer but also in winter. In summer, rainfall happens frequently during the monsoon period (from early July to September). The top-months of the monsoon rainfall in Pakistan are July and August (Ahmad et al., 2014, 2016). Rainfall is a major source of water for agriculture and energy production. According to Adnan and Khan (2009), the 60% of the total annual rainfall in Pakistan is due in summer. All Kharif crops are more dependent on monsoon rainfalls. These crops are also known as summer or monsoon crops. The major Kharif crops are rice, maize, cotton, jowar, bajra etc. Furthermore, winter rains are very beneficial for Rabi crops in the country (Shamshad, 1988; Ahmad et al., 2016). Rabi crops are sown at the beginning of winter season. The major Rabi crops are wheat, gram, peas and barley etc. Rabi crops are also known as winter or spring crops. In Pakistan, rainwater plays an important role in hydroelectricity and agricultural production. Agriculture is considered as the backbone of Pakistan's economy, but unfortunately, heavy rainfalls cause loss of lives, crops and infrastructures. This fact is undeniable that the society and the economy of any country can be affected due to extreme environmental events (Shabri et al., 2011). Therefore, the suitable modelling of extreme events can be done through statistical analysis.

The main purpose of the statistical analysis of extreme values is to enumerate the stochastic behavior of extreme environmental events (Smith, 1989; Coles et al., 2001). The most meteorological variables, such as rainfall, wind speed, temperature, floods and precipitation, are of under consideration because they have a terrific influence on humans and thus it is crucial to portray their behavior statistically due to their extreme positions.

Due to monsoon season in Pakistan; as heavy rainfalls are anticipated to take place frequently every year. According to the Federal Flood Commission (FFC) Pakistan, it is estimated that about cumulative financial loss US\$ 38.171 million during the past 70 years. Around 616,598 km$^2$ of the total land area of Pakistan is prone to flood, some 197,275 villages were damaged and 12,330 people have lost their lives due to 24 major floods. Being a developing country like Pakistan, natural disasters unquestionably cripples the country's output. It is evident that any extreme environmental event is unpredictable. Nevertheless, the influence of heavy rainfall events may be abridged by precautionary measures based on the results from statistical analysis of Annual Maximum Daily Rainfall (AMDR) series, as suggested by Zin et al. (2010), Khamkong (2012) and Diriba et al. (2014). The major objective of statistical modelling of extreme rainfall was to estimate the quantiles that might occur for the next years to come.

There are two main statistical approaches to model AMDR series, namely the block maxima and peak over a threshold (POT). The block maxima approach models the highest value of each year collected from the large sample, whereas (POT) approach requires the observations that exceed a certain level of threshold in the collected data (Pickands, 1975; Davison and Smith, 1990). However, according to Madsen et al. (1997) and Eastoe and Tawn (2012) the block maxima is preferred since the method in selecting the appropriate threshold in (POT) procedure sometimes can be complex.

The AMDR series needs to be modelled by appropriate probability distributions that provide the better inferences of the future behaviour of extreme rainfalls. Various kinds of statistical distributions often are able to be used in AMDR. They are Generalized Extreme

Value (GEV), Gumbel, Generalized Pareto (GP), Generalized Logistic (GLO) and the lognormal distribution. Because of their suitability to modelling of extreme value, these probability distributions have numerous applications for environmental data (e.g., An and Pandey, 2005; Furrer and Katz, 2008; Chikobvu and Sigauke, 2013; Jonathan and Ewans, 2013; Diriba et al., 2015; Ahmad et al., 2016). Several studies based on frequentist approaches have been conducted for finding the best fit distribution and modelling the AMDR data in Pakistan by Rasul et al. (2004), Haroon and Rasul (2009), Rasul et al. (2009) and Ahmad et al. (2014, 2016). Ahmad et al. (2016) suggested that the (GEV) and (GP) are the most suitable distribution for the AMDR series in the different regions of Pakistan.

In this study, the five extreme value statistical distributions (such as GP, GEV, GLO, Pearson Type-III and Log-Normal three parameter) are applied to the extreme rainfall data of Lahore station of the Punjab province by using the frequentist procedure via linear moment (L-moment) estimation method. However, the results demonstrate the tendency that the GEV distribution among all others fits the data in a more appropriate way, attention is therefore limited to the GEV distribution.

Meanwhile, extreme data are sporadic by their nature, the statistical inference on extremes may be improved by the addition of other sources of information through a prior distribution. Unlike frequentist approaches (maximum likelihood and L-moments), a Bayesian analysis of extreme values is also not dependent on the fundamental assumptions that are required by the asymptotic theory of frequentist approaches (Smith, 1985; Coles et al., 2001; Smith, 2002). In addition, according to Hamed and Rao (2000) the MLE method that was unable to get an appropriate estimate with a small sample. The Bayesian method can be employed in estimating the parameter of GEV distribution. The basic theory of the Bayesian paradigm about extreme values has been discussed by many authors in a number of outstanding articles and books (Coles and Tawn, 1996, 2005; Beirlant et al., 2006; Naghettini, 2017, p. 505). In this study, the researcher has focused on its applications by using the GEV distribution. Additionally, the primary purpose of this article is the comparison of both estimation methods by using the robustness measures. The first one is the MLE method and the second one is the Bayesian method. In Bayesian analysis, the non-informative priors and informative priors based on the historical data of neighboring stations in term of extreme quantiles are formulated. In the previous study, Ahmad et al. (2014) discussed the behaviour of monsoon rainfalls in Pakistan through kappa distribution by employing L-moments estimation method. In another study Ahmad et al. (2016) used the data of extreme rainfall of 28 stations in Pakistan and found the best fit distribution among five extreme value distributions. After finding the best fit distribution the researcher carried out at-site frequency analysis on the same data by using L-moments and Trimmed L-moment estimation methods. However, they did not use the Bayesian paradigm for the modelling of extreme rainfall in Pakistan. The main purpose of this paper, therefore, is to study the extremes of rainfall by applying the GEV distribution to the Lahore station data, the province of Punjab, Pakistan. The frequentist approach, i.e. MLE and the Bayesian method were applied (Coles and Tawn, 1996; Diriba et al., 2017). The effects of non-informative and informative priors on the estimate of GEV parameters and return level are investigated and compared with MLE results. Despite the application of the Bayesian approach is progressively prevalent in many areas, a challenge when using this method is the computational problems. This can be solved by the application of Markov Chain Monte Carlo (MCMC) simulations (Coles et al., 2001). Moreover, the previous studies conducted on Bayesian Paradigm by using MCMC were Coles and Tawn (1996), Coles et al. (2003), Smith (2005) and Eli et al. (2012). The rest paper is organized as follows. We commence with description of data used in this study. Next, the probability

distribution and parameters estimation methods used in this study will be described. After that, the results and discussion section of our analysis will be presented. Later on, the conclusion, recommendations concerning the plan for future studies will be discussed in final section.

## Materials and methods

### Data description

The modeling of extreme rainfalls in the province Punjab, Pakistan, the data of daily rainfall of Lahore station had been acquired from Pakistan Metrological Center Karachi for the period of 1985 to 2014, which is recorded in millimeters (mm). Data have been selected on the following standard criteria like as length of the data, variability, quality, urbanization and climate change. After that, the Annual Maximum Daily Rainfall (AMDR) series were extracted from the daily rainfall series. AMDR series is a single maximum value for any specific year and station among the all values of recorded daily rainfall. The AMDR series for 30 years of the same length were considered. For construction of informative priors for the Bayesian analysis, extreme rainfall characteristics of three sites at various distances from Lahore station were selected, namely Faisalabad, Sialkot and Jhelum.

The selected stations for informative priors were located to the south-west, north-east and north-west of Lahore station. Additionally, they are 118 km, 105 km, 164 km from Lahore station, respectively. The plots under spatial network comprise of selected weather stations studied in present research are shown in *Figure 1*. This study have 25 years of data at the same time points from 1990-2014 for each station considered for construction of informative priors.



***Figure 1.*** *Spatial plots of network weather stations used in study for the period of 1985-2014*

### Summary statistics of the selected stations

The statistical parameters for the amount of AMDR data of four selected stations are briefed in *Table 1*. Where the minimum, maximum, mean, standard deviation (SD), coefficient of variation (CV), skewness and kurtosis of AMDR series of the all included stations in this study are given. The AMDR data designates that the annual daily rainfall

was strongly positively skewed and highly leptokurtic for Jhelum station. It is also noted that AMDR data for Sialkot station is positively skewed and platykurtic. The CV of AMDR around 42-45% in the province Punjab, Pakistan.

*Table 1. Summary statistics of the annual maximum daily rainfall (AMDR) data for the selected stations*

| Study locations | Stations | | | |
|---|---|---|---|---|
| | Lahore | Faisalabad | Jhelum | Sialkot |
| Minimum (mm) | 29.40 | 25.00 | 54.90 | 45.00 |
| Maximum (mm) | 189.70 | 136.00 | 242.20 | 207.50 |
| Mean | 90.45 | 63.43 | 99.40 | 110.10 |
| Standard deviation | 38.32 | 27.66 | 44.56 | 47.44 |
| Coefficient of variation | 42.35 | 43.61 | 44.83 | 47.44 |
| Coefficient of skewness | 0.91 | 1.05 | 1.52 | 0.71 |
| Kurtosis | 3.35 | 4.02 | 5.29 | 2.34 |

*The fundamental assumptions*

Primarily, it is compulsory to check the basic assumptions of data before conducting the further analysis of an annual maximum series in the field of statistical hydrology; because the final results may be ambiguous without fulfilling the assumptions. The fundamental assumption of hydrological data are randomness, independence, homogeneity and stationarity. We tested these assumptions using different tests such as NERC test for randomness, Wald-Wolfowitz for independence, Mann-Whitney test for homogeneity with respect to location parameters and Spearman rank correlation test for stationarity. The results of these tests given in *Table 2* are calculated by using (ALEA 2.0) software.

*Table 2. Results of different tests for basic assumptions at 5% level of significance*

| Study locations | NERC (P-value) | Wald-Wolfowitz (P-value) | Mann-Whitney (P-value) | Spearman Rank (P-value) |
|---|---|---|---|---|
| Lahore | -0.7445 (0.2283) | 1.3505 (0.0884) | -0.3940 (0.3468) | 0.7895 (0.2149) |
| Faisalabad | 1.9358 (0.1264) | 0.5230 (0.3005) | -2.0117 (0.1021) | 2.4021 (0.0820) |
| Jhelum | 1.0423 (0.1486) | 1.5812 (0.0569) | -1.8043 (0.0556) | -1.3981 (0.0810) |
| Sialkot | -0.7445 (0.2283) | 0.5039 (0.3072) | -0.4770 (0.3167) | 0.2767 (0.3910) |

The probability value of these tests for each stations is more than 5% level of significance showed that the relevant assumptions of the data are fulfilled, and the available data can be used for further frequency analysis (FA).

### Generalized extreme value distribution

Jenkinson (1955) introduced the GEV distribution for the three limiting cases for maxima. Furthermore, the GEV distribution is very extensively used for the design structure of extreme events. Engineers, hydrologist and business analyst are most commonly used GEV distribution in the recent decade. Because, the risk always linked with the extreme type events such as risk in business, the stock market, and also risk is related to extreme floods, heavy rainfall, high wind speed and extreme temperature etc. The GEV probability density function and distribution function of a random variable w is given by:

$$f(w_i, \xi, \mu, \sigma) = \frac{1}{\sigma}\left[1 + \xi\frac{(w_i - \mu)}{\sigma}\right]^{-\left(1+\frac{1}{\xi}\right)} \exp\left[-\left\{1 + \xi\frac{(w_i - \mu)}{\sigma}\right\}^{-\frac{1}{\xi}}\right] \qquad \text{(Eq.1)}$$

The range of w is: $-\infty < w \leq \mu + \frac{\sigma}{\xi}$ if $\xi > 0$; $-\infty < w \leq \infty$ if $\xi = 0$; $\mu + \frac{\sigma}{\xi} < w \leq \infty$ if $\xi < 0$.

$$F(w, \xi, \mu, \sigma) = \exp\left[-\left\{1 + \xi\frac{(w - \mu)}{\sigma}\right\}^{-\frac{1}{\xi}}\right] \qquad \text{(Eq.2)}$$

where $\{w : 1 + \xi(w - \mu)/\sigma\}$ and $\mu, \sigma$ and $\xi$ are location, scale and shape parameters of GEV distribution (Beirlant et al., 2006). The shape parameter affects the nature of upper tail of the GEV distribution. Additionally, GEV distribution is the mixture of three limiting extreme value distributions i.e., Gumbel distribution, Freshet distribution and Weibull distribution. If $\xi \to 0$ the GEV distribution in *Equation 2* relates to the Gumbel distribution.

$$F(w, \xi, \mu, \sigma) = \exp\left[-\exp\left(-\frac{(w - \mu)}{\sigma}\right)\right]; -\infty < w < \infty \qquad \text{(Eq.3)}$$

It means that the probability of obtaining the maximum high observation come to be very small exponentially. For $\xi > 0$ the GEV distribution in *Equation 2* converts to Frechet distribution, which is considered by a heavy tail that declines polynomially. Hence, obtaining higher values of maximum are larger probability in comparison with a Gumbel distribution. For $\xi < 0$ the distribution in *Equation 2* becomes negative Weibull distribution and this is bounded above (Coles et al., 2003; Karim, 1995; Gingras and Adamowski, 1992). Consequently, there is a finite value that the maximum cannot exceed.

### Parameter estimation of GEV distribution

Maximum Likelihood Estimation (MLE) method was primarily articulated by a German mathematician Johann Carl Friedrich Gauss. But later as a general method of

estimation was initially developed by a British statistician Sir Ronald Aylmer Fisher in his series of papers discussed by Gupta (1982). MLE method involved a function that maximizes the parameters of the distribution, the function is known as the likelihood function. Likelihood function reaches to the maximum value when equating it to the conditions of unknowns and an identical number of equations in the system, whose solution produce the MLE estimators for unknown distribution parameters. The author was used MLE method to estimate the GEV distribution parameters. The density function of GEV distribution is given in *Equation 1*. Therefore the joint likelihood function associated with $w_1, w_2, \ldots, w_m$ follows from *Equation 1*.

$$L(\mu, \sigma, \xi; w_1, w_2, \ldots, w_m) = \frac{1}{\sigma^m} \prod_{i=1}^{m} \left[ 1 + \xi \frac{(w_i - \mu)}{\sigma} \right]^{-\left(1 + \frac{1}{\xi}\right)} \times \exp \left[ -\sum_{i=1}^{m} \left\{ 1 + \xi \frac{(w_i - \mu)}{\sigma} \right\}^{-\frac{1}{\xi}} \right] \quad \text{(Eq.4)}$$

The maximum likelihood estimates of the parameters $\mu, \sigma$ and $\xi$, say $\hat{\mu}, \hat{\sigma}$ and $\hat{\xi}$, are calculated by maximizing the logarithm of *Equation 4*, that is, maximizing

$$l(\mu, \sigma, \xi; w_1, w_2, \ldots, w_m) = m \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m} \log \left[ 1 + \xi \frac{(w_i - \mu)}{\sigma} \right] - \sum_{i=1}^{m} \left\{ 1 + \xi \frac{(w_i - \mu)}{\sigma} \right\}^{-\frac{1}{\xi}} \quad \text{(Eq.5)}$$

With respect to unknown parameters, say $\mu, \sigma$ and $\xi$. In other words, the log-likelihood function i.e., $\ln[L(\mu, \sigma, \xi; w_1, w_2, \ldots, w_m)] = l(\mu, \sigma, \xi; w_1, w_2, \ldots, w_m)$ is maximized instead of $L(\mu, \sigma, \xi; w_1, w_2, \ldots, w_m)$, which simplifies the complication and gives the easy solution to the system. In a continuous parameter space the shape parameter of Gumbel distribution, $\xi = 0$ is a single point, it will not be estimated by maximizing this log likelihood with probability 1. In practice the maximization of log likelihood functions is done by numerically iterations e.g., with a quasi-Newton procedure (Diriba et al., 2014, 2015).

### Return level estimation for GEV distribution

The attention of extreme climatic events analysis does not sometimes lie on the estimates of GEV parameters, however the application of the fitted model to estimate the other measures. For illustration, it is very important the forecast of future extreme rainfall and heavy floods to make a proper plan to reduce its negative impact (for example landslides, human lives, animals and destruction of infrastructures such as power stations, buildings and roads) and return level or quantiles are used for such forecast. In other word, return level estimates play a dynamic role in rainfall frequency analysis to finding the future risk connected with a return period corresponding to a fitted model. The return level for GEV model corresponding the return period $T = \frac{1}{p}$ is acquired through quantile function, which is the inverse of *Equation 2* given by Coles (2001), denoted by $w_p$ where $F(w_p) = 1 - p$ and $0 < p < 1$.

$$w_p = \begin{cases} \mu - \dfrac{\sigma}{\xi}\left[1-\{-\log(1-p)\}^{-\xi}\right], \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\}, \xi = 0 \end{cases} \qquad \text{(Eq.6)}$$

The return level $w_p$ is a quantile of considered GEV distribution corresponding to the upper tail probability $p$. The ML estimate of the return level $w_p$, denoted by $\hat{w}_p$ is found by substituting the ML estimates $\hat{\mu}, \hat{\sigma}$ and $\hat{\xi}$. The standard error (SE) of $\hat{w}_p$ are acquired via delta method (Rao, 1973).

### Bayesian paradigm

The Bayesian approach has been extensively used in extreme rainfall frequency analysis. Bayesian statistical analysis depends on Bayes Theorem, which describes us how to update prior information about parameters and hypotheses using the additional information of data, to produce posterior beliefs suggested (Jackman, 2009). According to Naghettini (2017) and Tang and Ang (2007) difference between Bayesian and frequentist approaches are based on the parameters of given probability distribution.

In a classical approach, the parameters have fixed, but unknown, they are estimated from given sample by maximizing the likelihood function. But in Bayesian approach the parameters are treated as random, they have their own probability distribution. The parameters are updated on the basis of prior knowledge and sample information, the updated information is called the posterior information of the parameters.

As in MLE method, suppose AMDR $w = (w_1, w_2, \ldots, w_m)$ are independently and identically distributed and their distribution belongs to a parametric family i.e. GEV. However, the parameters of GEV distribution ($\mu, \sigma$ and $\xi$) are treated as random variables for which the author describes the prior distributions. Let $\theta = (\mu, \sigma, \xi)$ and suppose the prior information about parameter $\theta$ can be articulated by a probability density function $g_\theta(\theta)$ with no reference to the actual data. After multiplication of likelihood and prior information using Bayes theorem to get the posterior density of $\theta$ has the form:

$$\pi(\theta/w) = \frac{L(\theta/w)g_\theta(\theta)}{\int_\Theta L(\theta/w)g_\theta(\theta)d\theta} \propto L(\theta/w)g_\theta(\theta) \qquad \text{(Eq.7)}$$

where $L(\theta/w)$ is the likelihood function of GEV distribution given in *Equation 4* and $\pi(\theta/w)$ is the posterior distribution for $\theta$. Furthermore, the integral is taken over the parameter space $\Theta$. In this paper, both the non-informative and informative priors were used. The description about construction of these prior are given in previous sections.

### Priors specification

#### Non-informative priors

The main demur in the use of Bayesian frame work is the requirement for postulating a prior knowledge $g_\theta(\theta)$. When the minimal or no information is available, the

Bayesian analysis is performed by using objective or non-informative and Jeffery's priors. The non-informative priors (also known as diffuse, flat or vague priors) used to indicate a set of principles about the concerned parameters of GEV distribution, because to betoken that the meaningful information of extreme rainfall is quite inaccessible at the moment. The AMDR series have a GEV distribution i.e., $W_i \sim GEV(\mu, \sigma, \zeta)$, the likelihood function is given in *Equation 4*. The parametrization $\varphi = \log \sigma$ is used in the place of $\sigma$, for the purpose of easier work to the specification of prior, $\sigma$ is inhibited to be positive. The tri-variate normal distribution based on $\theta' = (\mu, \varphi, \xi)$ is used with prior density as follows:

$$g_\theta(\theta) \propto \frac{1}{\sigma} \exp\left\{ -\frac{1}{2}(\theta' - u)^T \sum\nolimits^{-1}(\theta' - u) \right\} \tag{Eq.8}$$

where the mean vector $u$ is symmetric positive definite $(3 \times 3)$ covariance matrix and $\sum$ must be quantified. The prior density is preferred to be

$$g(\mu, \varphi, \xi) = g(\mu)g(\varphi)g(\xi) \tag{Eq.9}$$

The marginal priors of $g_\mu(.), g_\varphi(.)$ and $g_\xi(.)$ are:

$$\mu \sim N(0, 10000)$$
$$\varphi \sim LN(0, 10000)$$
$$\xi \sim N(0, 100)$$

These are known as independent normal priors with large variances. The larger variances are preferred enough to create the distribution nearly flat, comparable to prior obliviousness (Eli et al., 2012; Fawcett and Walshaw, 2008).

*Informative priors*

Since the nonexistence of expert knowledge on extreme rainfalls in Lahore station, we have articulated the prior knowledge for the Bayesian analysis from rainfall characteristics of nearby weather stations in the same province. Hence, the informative priors were constructed by employing the approach given by (Coles and Tawn, 1996). Prior information were elicited in term of extreme quantiles. The brief description of the technique given in the following paragraph.

Remember that return level $w_p [= q_p$ Coles and Tawn (1996) notation] for GEV distribution corresponding with the return period of $\frac{1}{p}$ is given in *Equation 6,* therefore $F(q_p) = 1 - p$, and where $F(.)$ is the cumulative distribution function of GEV in *Equation 2*. Let $q_{p_i}$, $i = 1, 2, 3$ with $p_1 > p_2 > p_3$, be the quantiles calculated from the past extreme rainfall data of the nearby weather stations in the province Punjab other than Lahore. So, the independent quantiles $q_{p_i}$ are calculated for each nearby station by means of ML estimates of GEV parameters in *Eq. 2*. A joint prior distribution for given

probabilities $p_1 > p_2 > p_3$ could be elicited from the quantiles $(q_{p_1}, q_{p_2}, q_{p_3})$ discuss by Coles and Tawn (1996). One slight compilation with this approach is that parameters must be ordered $q_{p_i}$, $i = 1, 2, 3$, i.e. $q_{p_1} < q_{p_2} < q_{p_3}$. Hence, the suggest quantile differences, because of the assumption of independent priors on $q_{p_i}$, $i = 1, 2, 3$ would not be valid.

$$\left.\begin{aligned} \tilde{q}_1 &= q_{p_1} - e_1 \\ \tilde{q}_2 &= q_{p_2} - q_{p_1} \\ \tilde{q}_3 &= q_{p_3} - q_{p_2} \end{aligned}\right\} \tag{Eq.10}$$

where $e_1$ is a physical lower end point for the process variable (rainfall) or naturally considered to be $e_1 = 0$. Perceive that the quantile differences confirm the ordering of quantiles. The marginal priors on the quantile differences are now supposed to independent Gamma distribution with parameters $(\alpha_i^*, \beta_i^*), i = 1, 2, 3,$ we can take in the form

$$\tilde{q}_{p_i} \sim gamma(\alpha_i^*, \beta_i^*), \ \alpha_i^* > 0, \beta_i^* > 0; \ i = 1, 2, 3. \tag{Eq.11}$$

From *Equations 10* and *11* we can get the joint prior for the $(\alpha_i^*, \beta_i^*)$, $i = 1, 2, 3$ in following form:

$$g(q_{p_1}, q_{p_2}, q_{p_3}) \propto \tilde{q}_{p_1}^{\alpha_1^* - 1} \exp(-\beta_1^* \tilde{q}_{p_1}) \times \prod_{i=2}^{3} \tilde{q}_{p_i}^{\alpha_i^* - 1} \exp(-\beta_i^* \tilde{q}_{p_i}) \tag{Eq.12}$$

provided that $q_{p_1} < q_{p_2} < q_{p_3}$. Then utilize *Equation 6* in *Equation 12* and multiplying by the Jacobean (J) of the transformation from $(q_{p_1} < q_{p_2} < q_{p_3})$ to $\theta = (\mu, \sigma, \xi)$, it gives to an expression for the prior in term of the GEV distribution parameters $\theta$. Further, this can be written as

$$g_\theta(\theta) \propto \prod_{i=1}^{3} \tilde{q}_{p_i}^{\alpha_i^* - 1} \exp(-\frac{\tilde{q}_{p_i}}{\beta_i^*}) \times J \tag{Eq.13}$$

provided that $q_{p_1} < q_{p_2} < q_{p_3}$. The Jacobean used in *Equation 13* is given by

$$J = \begin{cases} \dfrac{\sigma}{\xi^2} \left| \sum_{i<j} (-1)^{i+j} (y_i \times y_j) \log\left(\dfrac{y_j}{y_i}\right) \right|, [i, j \in \{1, 2, 3\}; \xi \neq 0 \\ \dfrac{\sigma}{2} \left| \sum_{i<j} (-1)^{i+j} \log y_i \times \log y_j \times \log\left(\dfrac{y_j}{y_i}\right) \right|, [i, j \in \{1, 2, 3\}; \xi = 0 \end{cases} \tag{Eq.14}$$

where $y_i = -\log(1 - p_i), i = 1, 2, 3$. The posterior density $\pi(\theta / w)$ is found by replacing the likelihood and prior density given in *Equations 5* and *12,* respectively, in *Equation 7*. However, the analytical calculation for *Equation 7* is completely intractable, because of the involvement of integrals in denominator. Therefore, the features of posterior distribution $\pi(\theta / w)$ of $\theta$ is to be estimated by using Markov Chain Monte Carlo (MCMC) method with the combination of Gibbs Sampler and Metropolis-Hastings algorithm.

In Bayesian analysis, to estimate the return levels, usually, the vectors of simulated values are found by employing the marginal posterior distributions of GEV parameters. Return level $q_p$ is obtained for a specified return period $\dfrac{1}{p}$ by substituting the simulated samples of GEV parameters from posterior distribution. Furthermore, the summary statistics can easily be acquired.

### Robustness measures

On the basis of robustness measures, the authors will also compare the performance between the classical (MLE) and Bayesian MCMC methods (both for non-informative and for informative priors) in estimating GEV parameters and return levels for AMDR data of Lahore weather station in the province Punjab, Pakistan.. The selected robustness measures are relative root mean square error (RRMSE), relative absolute square error (RASE) and probability plot correlation coefficient (PPCC). The first two measure comprises the assessment on the difference between the observed and the estimated values under the presumed distribution, while the third one is measure the relationship between the ordered values and associated estimated values (Zawiah et al., 2009). The formulae are given as:

$$RRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{w_{i:n} - \hat{q}(F_i)}{w_{i:n}}\right|\right)^2} \tag{Eq.15}$$

$$RASE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{w_{i:n} - \hat{q}(F_i)}{w_{i:n}}\right| \tag{Eq.16}$$

$$PPCC = \frac{\sum_{i=1}^{n}(w_{i:n} - \bar{x})(\hat{q}(F_i) - \bar{q}(F_i))}{\sqrt{\sum_{i=1}^{n}(w_{i:n} - \bar{x})^2 \sum_{i=1}^{n}(\hat{q}(F_i) - \bar{q}(F_i))^2}} \tag{Eq.17}$$

where $w_{i:n}$ represents the observed sample values of *ith* order statistics of a random sample of size $n$, and $\hat{q}(F_i)$ are estimated quantiles linked with the *ith* Weibull plotting position, $F$. The Weibull plotting is defined as follows:

$$F = \frac{i}{n+1}$$

where $i$ denote the ranks of the observations arranged in ascending order. The minimum values of RRMSE and RASE will show the best method. In contrast, the value of PPCC that is closest to -1 or 1 will be reflected as the best method for elucidation the behavior of extreme rainfalls in Punjab, Pakistan.

## Results and discussion

### *GEV distribution using MLE*

To analyze the AMDR data, GEV distribution is fitted to the extreme rainfall values. The parameters of the fitted model are estimated by MLE method. Maximum likelihood estimates with their SE in (parentheses) and the 95% confidence interval (CI) in [square brackets] for fitted model at Lahore Station are given in *Table 3*. It can be seen that the entire results of the fitted model (i.e. GEV) supports the AMDR data.

***Table 3.*** *Parameter estimates of location, scale and shape of GEV distributions, their standard errors and confidence interval*

| Study locations | Location ($\hat{\mu}$) (SE) [CI] | Scale ($\hat{\sigma}$) (SE) [CI] | Shape ($\hat{\xi}$) (SE) [CI] |
|---|---|---|---|
| Lahore | 72.8995 (5.9325) [61.2686, 84.5208] | 28.6298 (4.3780) [20.0469, 37.2029] | 0.0317 (0.1461) [-0.2544, 0.3180] |

Furthermore, before calculating the return levels, we should have to check the goodness of fit of our model. The adequacy of the chosen model can be checked using the quantiles, probability and return level plot. The plots for GEV distributions fitted to the data of Lahore station are given in *Figure 2*. The quantiles and probability plots shows that the assumptions required for the GEV distributions have been fulfilled.

The return levels for best fitting model (i.e. GEV) could be estimated for next '$T$' year after base year using *Equation 6*. Therefore, the return levels estimates will be time varying. An estimate of the rainfall that the authors might assume to see at Lahore station in the next '$T$' years by employing the MLEs of the parameters in GEV distribution. Before estimating $\hat{w}_p$, the number of years should be converted to the probability (i.e. $p = \dfrac{1}{T}$). The predicted amount of maximum rainfall, that is, in context of return level, corresponding to return periods in *Figure 2* suggests that the return level falls below the 20-year return periods. Furthermore, it can be noted that the 95% confidence bounds are becoming very distant from the straight line for the return periods beyond the 20 years. Return levels for different return periods with corresponding the 95% CIs were estimated using the profile likelihood methods are given in *Table 4*.

Therefore, the return levels for AMDR increased slowly for the higher return periods. It can be seen that from *Table 4,* the profile likelihood intervals were gradually wider as return period increased.

This supports with the results showed in the return level plots given in *Figure 2*. However, since the profile likelihood method takes into account the skewness of the parameter distributions, it gives more accurate results for longer return periods.
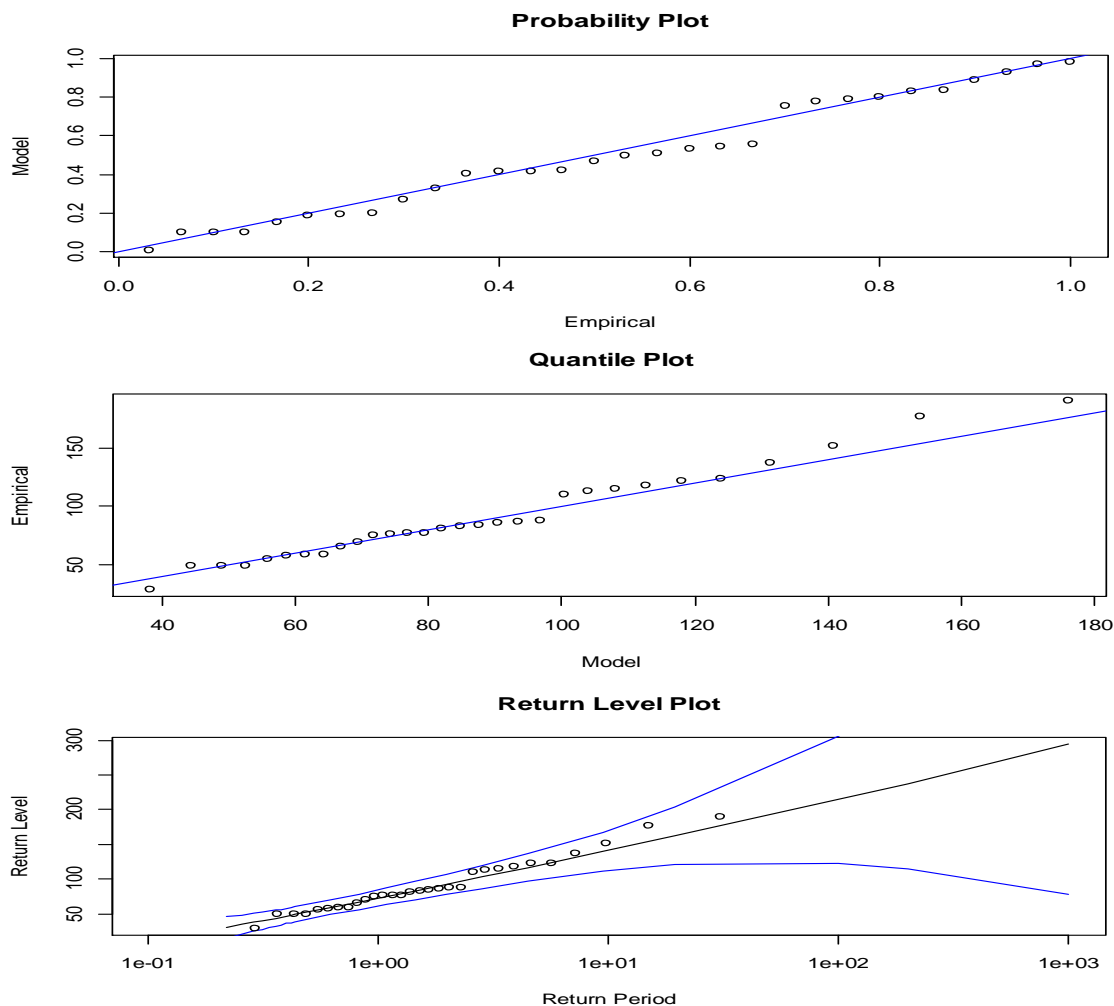
***Figure 2.*** *Diagnostic plots based on profile likelihood method to assess goodness-of-fit of the GEV distribution fitted to AMDR data. From top to bottom: probability, quintile and return level*

***Table 4.*** *Return level and 95% profile likelihood confidence intervals for 10, 25, 50, 75 and 100-year return levels of AMDR based on MLE*

| Return periods | Return levels | Profile likelihood confidence intervals |
|:---:|:---:|:---:|
| 10 | 139.6837 | (118.3549, 187.0867) |
| 25 | 169.2826 | (138.9474, 263.5811) |
| 50 | 191.8235 | (153.1579, 340.1084) |
| 75 | 205.1583 | (160.5263, 393.9222) |
| 100 | 214.7013 | (166.8421, 437.5112) |

## Bayesian analysis

In Bayesian analysis, we were employed the non-informative and informative priors to the extreme rainfall data of Lahore weather station. The non-informative priors used (Coles and Tawn, 2005) were constructed for GEV parameters by assuming there is no reliable prior information available about the parameters given in priors specification

subheading (*non-informative priors*). Non-informative independent priors for GEV parameters were assumed $g(\mu) \sim N(0,10000)$; $g(\varphi) \sim LN(0,10000)$ and $g(\xi) \sim N(0,100)$.

The scale parameter of GEV (i.e. $\sigma$) was reparametrized as $\varphi = \log(\sigma)$ to hold the positivity of this parameter. The Gaussian distribution having zero mean and higher variance impose the flat marginal priors (also known as diffuse or vague priors), which show the absence of external information. The MCMC technique with combination of Gibbs Sampler (GB) and Metropolis-Hasting (M-H) algorithm was applied to AMDR series. Different starting points were used in simulations to check that the chains had converged to the original series and all the chains mixed well. The posterior means, standard deviation (SD) and 95% CIs of the parameters are given in *Table 4*. Furthermore, the posterior means, standard deviations (SD) are close to ML parameter of GEV distribution (see *Table 3*). The CIs of posterior parameters are little bit wider as compared to MLEs. It is expected that for flat-priors that posteriors means of GEV parameters with their CIs would be close to MLEs because they incorporate little information to the likelihood.

The informative priors were articulated by using historical records of rainfall data from three nearby weather stations to Lahore station. Moreover, the prior distributions were elicited using the method given by Coles and Tawn (1996) and discussed in subheading (*informative priors*) with quantiles $p_i = 10^{-i}$; for $i = 1,2,3$. By using this procedure, we have found $\tilde{q}_{p_1} \sim gamma(25.9065,4.015)$, $\tilde{q}_{p_2} \sim gamma(0.5702,117.4019)$ and $\tilde{q}_{p_3} \sim gamma(0.0999,804.3856)$ from Faisalabad weather station data. Similarly, the informative prior priors for Sialkot and Jhelum weather station were elicited by applying the same procedure. The posterior means and SD of GEV parameters with CIs are also given in *Table 5*. It can be seen the posterior means and CIs for GEV parameters based informative prior are close posterior means and CIs of non-informative priors.

*Table 5. Posterior means, standard deviations and confidence intervals for GEV parameters to AMDR data of Lahore weather station*

| Study locations | Parameters estimate of GEV distribution | | |
|---|---|---|---|
| | Location ($\hat{\mu}$) (SE) [CI] | Scale ($\hat{\sigma}$) (SE) [CI] | Shape ($\hat{\xi}$) (SE) [CI] |
| Non-informative priors | 72.5232 (6.3956) [60.362, 85.5687] | 31.0642 (5.1108) [22.7240, 42.6316] | 0.0527 (0.1501) [-0.2130, 0.3696] |
| Informative priors | | | |
| Faisalabad | 70.4847 (5.8650) [59.2650, 82.3369] | 28.3486 (4.4383) [21.1682, 38.4110] | -0.0006 (0.1108) [-0.2169, 0.2179] |
| Sialkot | 73.40324 (6.2712) [61.3944, 86.1910] | 31.76948 (4.9098) [23.7582, 42.9742] | 0.07116 (0.1429) [-0.2006, 0.3549] |
| Jhelum | 72.76406 (6.3700) [60.6154, 85.6824] | 31.00531 (4.9640) [22.8808, 42.2235] | 0.05265 (0.1430) [-0.2077, 0.3497] |

The informative priors constructed from the surrounding weather stations reduced the posterior SD for all GEV parameters as compared to non-informative priors and MLEs. This abatement in the SD would possibly overcome in uncertainty due to using extra information from the surrounding weather stations. This impact became most important for the SD of location, scale and shape parameters.

The estimated posterior densities plots of GEV parameters are given in *Figure 3*. It can be noticed that the distributions of location and shape parameters are symmetrical for non-informative and informative priors, while the shape parameter is negatively skewed for the prior elicited from the Faisalabad weather station that is situated at 118 km distance from Lahore. But the densities of location, scale and shape parameters have high peaks at center for the informative priors which were constructed using the Faisalabad and Sialkot data.

The densities for scale parameter were slightly positively skewed for non-informative and for all informative priors. It can be observed from densities plots, the densities of location, scale and shape parameters are shifted lift for the priors elicited from Sialkot and Jhelum data. These variations indicate that the posterior results might be changed due to the choice of priors. Furthermore, the distance between a weather station was used to articulate priors and Lahore weather station.
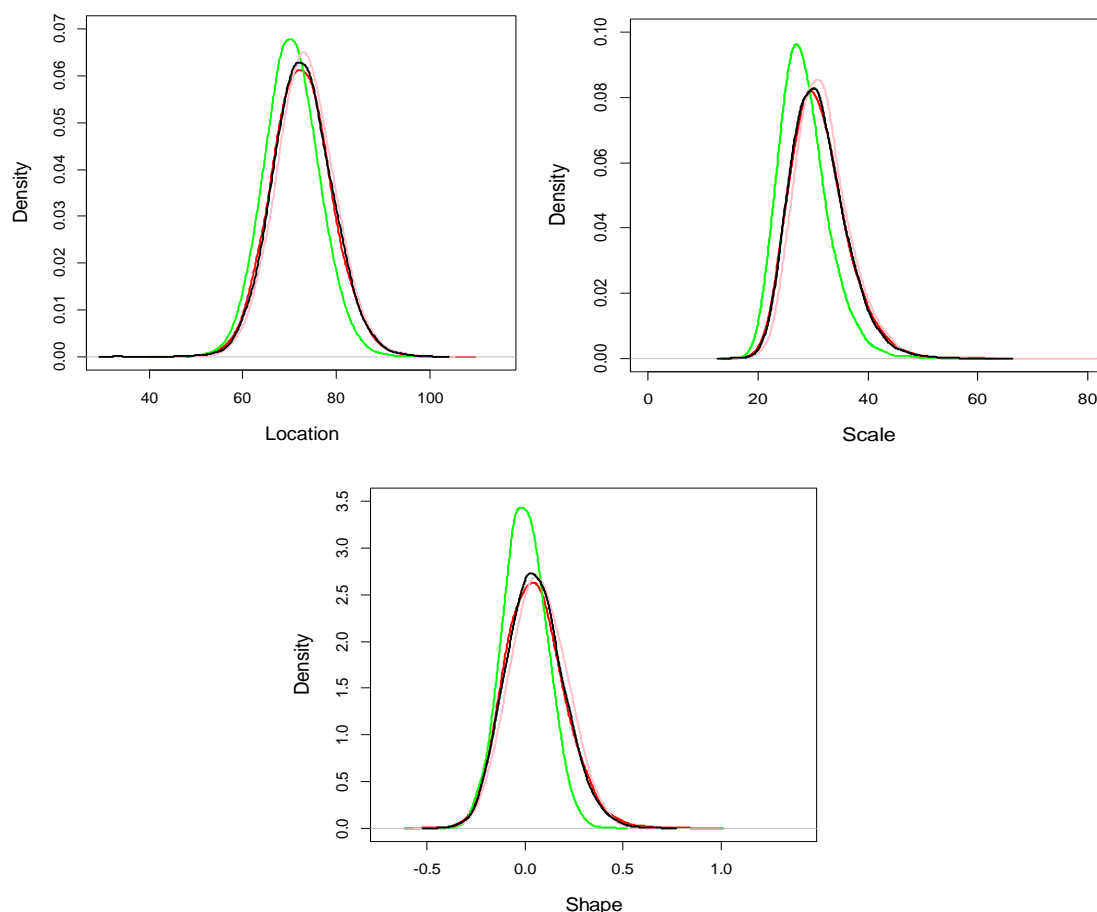


*Figure 3. Posterior densities plots of the GEV parameters to AMDR data using non-informative (red line) and informative priors (green line for Faisalabad, pink line for Sialkot and black line for Jhelum)*

*Priors effect on return level*

To look at how the non-informative and informative priors have affected the return levels, the authors obtained the posterior densities of the return levels. Additionally, the posterior density plots were constructed by employing the observations vectors obtained from the marginal posterior distributions of GEV parameters in *Equation 3,* for $0 < p < 1$. This process was carried out to attain the posterior densities of return levels for 10, 25, 50, 75, and 100 years return periods. Therefore, the posterior densities of these return levels for both non-informative and informative priors are given in *Figure 4*. The return level densities were symmetric and peaks high for the prior formulated from the Faisalabad weather station data, while the all other densities of the return levels were positively skewed and mesokurtic for both non-informative and the informative priors obtained from the Sialkot and Jhelum data. From this skewness, it can be cleared that the uncertainty might reproduce in the model for founding upper limits of the return levels as compare to lower limits for higher return periods (Coles and Tawn, 2005; Diriba et al., 2015). Furthermore, the posterior median of AMDR data for Lahore weather station was calculated. The results for different return periods are given in *Table 6*. The posterior medians of return levels for both non-informative and informative prior are not much different to their parallel MLEs. Moreover the results of *Tables 6* and *4* showed that the posterior medians of return levels for non-informative priors and informative priors constructed from nearby stations except Faisalabad are higher than MLEs across all considered return periods. The variations in higher medians of return levels (i.e. 75, 100) are smaller as compared to lower return levels. Hence the medians of returns levels vary due to distance of these stations (Faisalabad, Sialkot and Jhelum) and Lahore weather station, which was used to informative priors construction. However, the results of return level against different return periods are higher as compared to previous study which was conduct by Ahmad at el. (2016). Therefore, it is undeniable that these results are significant for policy implications in the region. They are also beneficial for meteorologists and hydrologists in water resource management.

**Table 6.** *Posterior medians for 10, 25, 50, 75 and 100-year return levels of AMDR using both non-informative and informative priors*
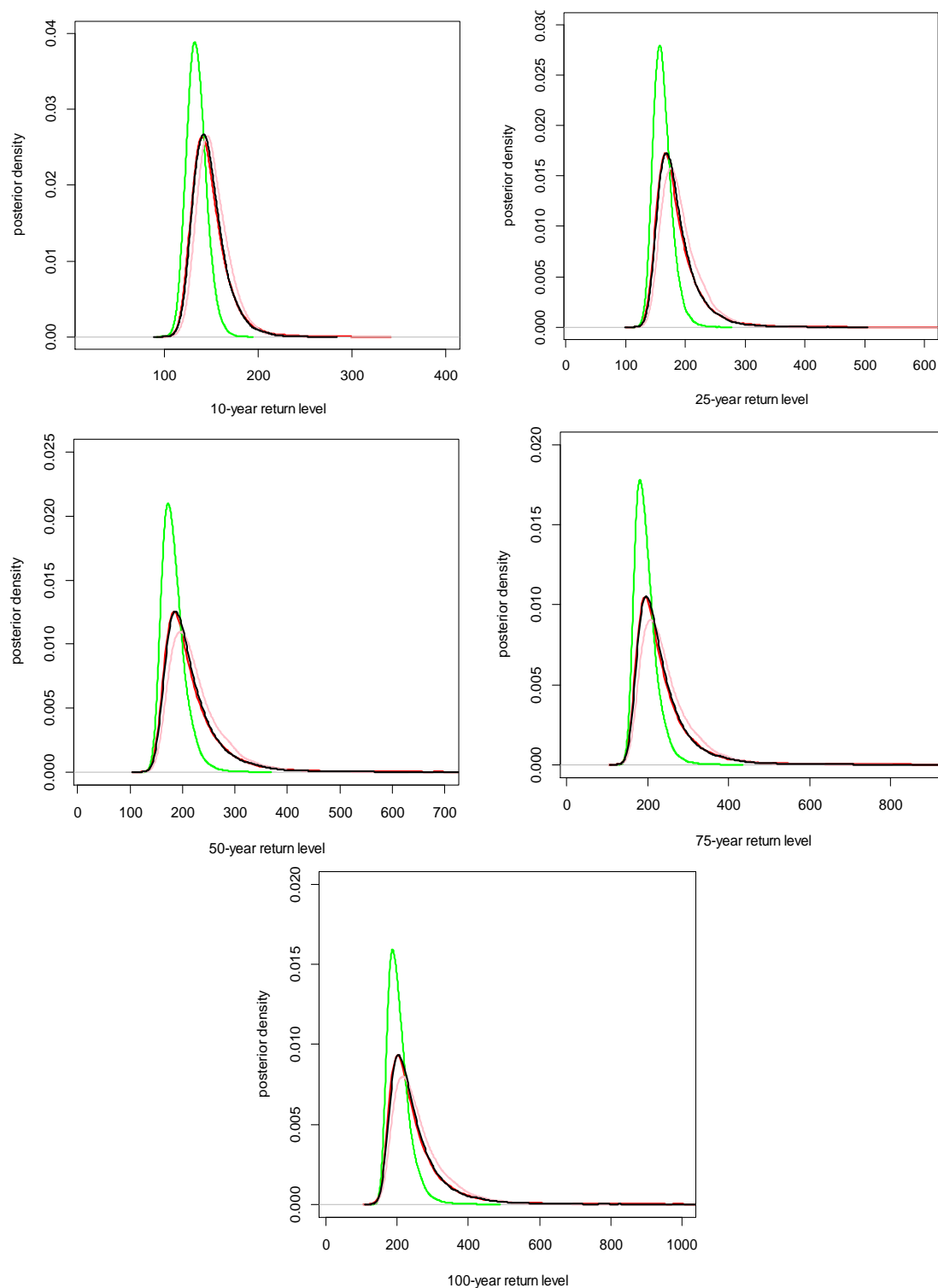
| Priors | Return levels | | | | |
|---|---|---|---|---|---|
| | **10** | **25** | **50** | **75** | **100** |
| Non-informative | 146.7440 | 180.7501 | 207.0989 | 222.8688 | 234.2373 |
| Informative | | | | | |
| Faisalabad | 134.2359 | 161.0712 | 180.9693 | 192.5310 | 200.712 |
| Sialkot | 150.9387 | 187.5129 | 216.2855 | 233.6831 | 246.3059 |
| Jhelum | 146.8390 | 180.7752 | 207.0684 | 222.8045 | 234.1485 |

*Robustness measures*

On the basis of result given in *Table 7,* it can be decided the Bayesian MCMC paradigm for both non-informative and informative priors is slightly performed better than MLE. Because, the smaller values of RRMSE and RASE show little difference between the observed and estimated values. Furthermore, the smaller value of RRMSE and RASE corresponding to informative priors also indicate that Bayesian MCMC

method provide more precise results, when the expert prior were used in estimation. With the support of the results of PPCC test, it can be declared that Bayesian MCMC method was superior in modeling of extreme rainfall in the Punjab, Pakistan.



***Figure 4.*** *Posterior median densities plots for the 10, 25, 50, 75 and 100 year return levels using AMDR data using non-informative (red line) and informative priors (green line for Faisalabad, pink line for Sialkot and black line for Jhelum)*

*Table 7. Comparison of performance between MLE and Bayesian method (both for non-informative and informative priors)*

| Method of estimation | Test of robustness | | |
|---|---|---|---|
| | RRMSE | RASE | PPCC |
| MLE | 0.68443 | 0.66774 | 0.8858 |
| Bayesian | | | |
| Non-informative priors | 0.67973 | 0.66050 | 0.9123 |
| Informative priors | | | |
| Faisalabad | 0.64670 | 0.63706 | 0.97211 |
| Sialkot | 0.65682 | 0.63957 | 0.96455 |
| Jhelum | 0.66143 | 0.63781 | 0.96054 |

## Conclusion

Our objective was to conduct the extreme value analysis for the modelling of AMDR data of Lahore weather station, Punjab, Pakistan through frequentist and Bayesian approaches. The AMDR data was shown to follow GEV distribution which has a positive shape parameter, frequently appealing as it has a finite upper limit. In frequentist method, the return levels for (10, 25, 50, 75 and 100 years) were estimated by using parameter estimates obtained via MLE. These results indicated that there would be an extreme rain event occurring in future. In addition, the authors used Bayesian MCMC approach, which provide a more efficient analysis and predictions. Therefore, the results found for both non-informative and informative priors using MCMC techniques and were compared with MLEs and each other. As the posterior means of best fit distribution parameter estimates attained using non-informative priors were close to MLEs of parameters. In addition, the influence of the informative priors used in the statistical analysis on posterior mean and SE depend on the distance among the weather stations used for construction of informative priors and the weather station where GEV model was best fitted. Furthermore, the SE of posterior means of the parameters were smaller due to shorter distance. The smaller values of robustness measures were directed precision of Bayesian MCMC method (both for non-informative and informative priors). Although, the Bayesian analysis offers a more suitable and straight way to handling and expressing uncertainties suggested (Coles et al., 2003).

The Bayesian paradigm presented in this study to model the behavior extreme event at Lahore weather station is suitable when climatological information is scare. The limitation of the method is that it considers the behavior of extreme rainfall is same over the weather stations in the province from which the data were acquired. However, the extreme rainfall behavior at different parts of the province Punjab can be affected by the topography of the area and other information. To construct the prior distributions for weather stations in addition to historical records, the authors suggest that the Bayesian analysis should utilize this kind of information. Moreover, this study shows that the effects of informative priors on posterior means and accuracy of parameter estimates of GEV depend on distance. In general, the results of this work are very helpful to future plan in the designing of dam, bridges, culverts, and flood control devices in the province Punjab, Pakistan. These findings could be enhanced further with the practice of spatial modelling and this is the theme of future research.

# REFERENCES

[1]     Adnan, S., Khan, A. H. (2009): Effective rainfall for irrigated agriculture plains of Pakistan. – Pakistan Journal of Meteorology 6(11): 61-72.

[2]     Ahmad, I., Mahmood, I., Malik, I. R., Arshad, I. A., Iqbal, Z. (2014): Probability analysis of monthly rainfall on seasonal monsoon in Pakistan. – International Journal of Climatology 34(3): 827-834.

[3]     Ahmad, I., Abbas, A., Saghir, A., Fawad, M. (2016): Finding probability distributions for annual daily maximum rainfall in Pakistan using linear moments and variants. – Polish Journal of Environmental Studies 25(3).

[4]     An, Y., Pandey, M. D. (2005): A comparison of methods of extreme wind speed estimation. – Journal of Wind Engineering and Industrial Aerodynamics 93(7): 535-545.

[5]     Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. L. (2006): Statistics of Extremes: Theory and Applications. – John Wiley & Sons, New York.

[6]     Chikobvu, D., Sigauke, C. (2013): Modelling influence of temperature on daily peak electricity demand in South Africa. – Journal of Energy in Southern Africa 24(4): 63-70.

[7]     Coles, S., Tawn, J. A. (1996): A Bayesian analysis of extreme rainfall data. – Applied Statistics 6(10): 463-478.

[8]     Coles, S., Tawn, J. (2005): Bayesian modelling of extreme surges on the UK east coast. – Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 363(1831): 1387-1406.

[9]     Coles, S., Bawa, J., Trenner, L., Dorazio, P. (2001): An Introduction to Statistical Modeling of Extreme Values. Vol. 208. – Springer, London.

[10]    Coles, S., Pericchi, L. R., Sisson, S. (2003): A fully probabilistic approach to extreme rainfall modeling. – Journal of Hydrology 273(1-4): 35-50.

[11]    Davison, A. C., Smith, R. L. (1990): Models for exceedances over high thresholds. – Journal of the Royal Statistical Society. Series B (Methodological): 393-442.

[12]    Diriba, T. A., Debusho, L. K., Botai, J., Hassen, A. (2014): Analysis of extreme rainfall at East London, South Africa. – Annual Proceedings of the South African Statistical Association Conference 2014(1): 25-32.

[13]    Diriba, T. A., Debusho, L. K., Botai, J. (2015): Modeling extreme daily temperature using generalized Pareto distribution at Port Elizabeth, South Africa. – Annual Proceedings of the South African Statistical Association Conference 2015(1): 41-48).

[14]    Diriba, T. A., Debusho, L. K., Botai, J., Hassen, A. (2017): Bayesian modelling of extreme wind speed at Cape Town, South Africa. – Environmental and Ecological Statistics 24(2): 243-267.

[15]    Eastoe, E. F., Tawn, J. A. (2012): Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. – Biometrika 99(1): 43-55.

[16]    Eli, A., Shaffie, M., Zin, W. Z. W. (2012): Preliminary study on Bayesian extreme rainfall analysis: a case study of Alor Setar, Kedah, Malaysia. – Sains Malaysiana 41(11): 1403-1410.

[17]    Faisal, N., Sadiq, N. (2009): Climatic zonation of Pakistan through precipitation effectiveness index. – Pak. J. Meteorol 6(16): 51-60.

[18]    Fawcett, L., Walshaw, D. (2008, June). Modelling environmental extremes. – Short Course for the 19th Annual Conference of the International Environmentrics Society, June 8, Kelowna, Canada.

[19]    Furrer, E. M., Katz, R. W. (2008): Improving the simulation of extreme precipitation events by stochastic weather generators. – Water Resources Research 44(12).

[20] Gingras, D., Adamowski, K. (1992): Coupling of nonparametric frequency and l-moment analyses for mixed distribution identification 1. – JAWRA Journal of the American Water Resources Association 28(2): 263-272.

[21] Gupta, S. C. (1982): Fundamentals of Mathematical Statistics: A Modern Approach. – Sultan Chand and Sons, New Delhi.

[22] Hamed, K., Rao, A. R. (2000): Flood Frequency Analysis. – CRC Press, Boca Raton, FL.

[23] Haroon, M. A., Rasul, G. (2009): Principal component analysis of summer rainfall and outgoing long-wave radiation over Pakistan. – Pakistan Journal of Meteorology 5(10).

[24] Jackman, S. (2009): Bayesian Analysis for the Social Sciences. Vol. 846. – John Wiley & Sons, Chicherster.

[25] Jenkinson, A. F. (1955): The frequency distribution of the annual maximum (or minimum) values of meteorological elements. – Quarterly Journal of the Royal Meteorological Society 81(348): 158-171.

[26] Jonathan, P., Ewans, K. (2013): Statistical modelling of extreme ocean environments for marine design: a review. – Ocean Engineering 62: 91-109.

[27] Karim, M. A., Chowdhury, J. U. (1995): A comparison of four distributions used in flood frequency analysis in Bangladesh. – Hydrological Sciences Journal 40(1): 55-66.

[28] Khamkong, M. (2012): Statistical modeling of annual monthly maximum rainfall in upper northern region of Thailand. – The 6th International Days of Statistics and Economics, 13-15 September, Prague.

[29] Madsen, H., Pearson, C. P., Rosbjerg, D. (1997): Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling. – Water Resources Research 33(4): 759-769.

[30] Naghettini, M. (ed.) (2017): Fundamentals of Statistical Hydrology. – Springer, Cham.

[31] Pickands III, J. (1975): Statistical inference using extreme order statistics. – The Annals of Statistics 3(1): 119-131.

[32] Rao, C. R. (1973): Linear Statistical Inference and Its Applications. Vol. 2. – Wiley, New York, pp. 263-270.

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

[34] Rasul, G., Chaudhry, Q. Z., Mahmood, A. (2009): Numerical simulation of heavy rainfall case in South Asia. – Pakistan Journal of Meteorology 6(11).

[35] Shabri, A. B., Daud, Z. M., Ariff, N. M. (2011): Regional analysis of annual maximum rainfall using TL-moments method. – Theoretical and Applied Climatology 104(3-4): 561-570.

[36] Shamshad, K. M. (1988): The Meteorology of Pakistan. First Ed. – Royal Book Company Publishers, Karachi.

[37] Smith, E. (2005): Bayesian modelling of extreme rainfall data. – Doctoral Dissertation, University of Newcastle upon Tyne (unpublished).

[38] Smith, E. P. (2002): An Introduction to Statistical Modeling of Extreme Values. – Springer, London.

[39] Smith, R. L. (1985): Maximum likelihood estimation in a class of non-regular cases. – Biometrika 72(1): 67-90.

[40] Smith, R. L. (1989): Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. – Statistical Science 4(4): 367-377.

[41] Tang, W. H., Ang, A. (2007): Probability Concepts in Engineering: Emphasis on Applications to Civil & Environmental Engineering. – Wiley, New York.

[42] Zawiah, W. Z. W., Jemain, A. A., Ibrahim, K., Suhaila, J., Sayang, M. D. (2009): A comparative study of extreme rainfall in Peninsular Malaysia: with reference to partial duration and annual extreme series. – Sains Malaysiana 38(5): 751-760.

[43] Zin, W. Z. W., Jamaludin, S., Deni, S. M., Jemain, A. A. (2010): Recent changes in extreme rainfall events in Peninsular Malaysia: 1971-2005. – Theoretical and Applied Climatology 99(3-4): 303.