Received: 30 October 2020 | Revised: 6 April 2021 | Accepted: 1 June 2021 | Published on: 17 June 2021

DOI: 10.1002/joc.7240

RESEARCH ARTICLE

International Journal of Climatology RMetS

A comprehensive study on the Bayesian modelling of extreme rainfall: A case study from Pakistan

Touqeer Ahmad[1] | **Ishfaq Ahmad**[2] | **Irshad Ahmad Arshad**[3] | **Nicolas Bianco**[1]

[1]Department of Statistical Sciences, University of Padova, Padova, Italy

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

[3]Department of Statistics, Faculty of Sciences, Allama Iqbal Open University, Islamabad, Pakistan

Correspondence
Touqeer Ahmad, PhD scholar in the Department of Statistical Sciences, University of Padova, Via Cesare Battisti, 241, 35121 Padova, PD, Italy.
Email: touqeer.ahmad@studenti.unipd.it

Funding information
Università degli Studi di Padova

Abstract

In this paper, the modelling of extreme rainfall is carried out in Pakistan by analysing annual daily maximum rainfall data via frequentist and Bayesian approaches. In frequentist settings, the parameters and return levels of the best fitted probabilistic model (i.e., generalized extreme value) are estimated using maximum likelihood and linear moments method. On the other side, under the Bayesian framework, the parameters and return levels are calculated both for noninformative and informative priors. This task is completed with the help of the Markov Chain Monte Carlo method using the Metropolis-Hasting algorithm. This study also highlights a procedure to build an informative prior through historical records of the underlying processes from other nearby weather stations. The findings attained from the Bayesian paradigm demonstrate that the posterior inference could be affected by the choice of past knowledge used for the construction of informative priors. Additionally, the best method for the modelling of extreme rainfall over the country is decided with the support of assessment measures. In general, the Bayesian paradigm linked with the informative priors offers an adequate estimations scheme in terms of accuracy as compared to frequentist methods, accounting for ambiguity in parameters and return levels. Hence, these findings are very helpful in adopting accurate flood protection measures and designing infrastructures over the country.

KEYWORDS

annual daily maximum rainfall series, Bayesian method, generalized extreme value, informative priors, maximum likelihood estimation, return levels

1 | INTRODUCTION

Pakistan is a country with exclusive rainfall patterns and appearances than other countries in the world (Arif *et al*., 2019). In particular, Pakistan faces two different rainfall seasons namely summer and winter. In summer, rainfall happens mainly during the monsoon season (early July to September). Therefore, July and August are considered the peak months for monsoon rainfalls. In winter, mostly rainfall events occur from (mid-December to March) (Ahmad *et al*., 2013). Unquestionably, extreme rainfall events are frequently connected with climate fluctuations, which may cause a series of natural disasters such as flash floods, heavy winds, and landslides.

Consequently, rapid fluctuations in the climate have frequently increased the number of heavy floods in the country.

During the 2020 monsoon rainfall spell, numerous rain-related losses were reported in Pakistan. According to Aljazeera News (2020), 31 casualties were reported in the southern Sindh province, whereas 23 people expired in Khyber Pakhtunkhwa province. Furthermore, 15 deaths were reported in southwestern Baluchistan province and 8 in Punjab province. Likewise, 13 more people passed away elsewhere in Pakistan's northern areas, including three in Pakistan administered Kashmir.

According to the Federal Flood Commission (FFC) report, floods have now become a regular feature in the country. Due to downpours, the government has faced an overall financial loss of more than US$ 38 billion during the past 70 years. The massive defeat in the economy, particularly in the agriculture sector, has dramatically influenced the country's progress.

Rainfall patterns have continuously been examined including the estimation of the rainfall distribution and the identification of damp or dry events on a specific day. However, information about the amount and happenings of extreme rainfall is inevitable for different purposes such as sustainable water resource management, government planning for water-related disasters and preparation of different hydraulic structures (Ahmad *et al*., 2016; Hussain *et al*., 2017). On the other side, it is unquestionable that the infrastructures and the region's economy might be affected by extreme events. For these reasons, statistical modelling and prediction of extreme environmental events are required for future planning. Moreover, the challenging task in modelling extreme events is to realize the happening probabilities linked with events that are extrapolated beyond the observed data.

Procedures to analyse extreme values comprise the frequency of happening of extreme events with the practice of probabilistic models for both similar or dissimilar processes for one or more climate variables (Noto and La Loggia, 2009; Lenderink and Fowler, 2017). Several research studies exist in literature in favour of selection and evaluation of different extreme value probabilistic models for extreme data, but due to the accessibility of the small length of observed data as compared to return periods of interest, continuously this work has been challenging and provocative (Fadhilah *et al*., 2007; El Adlouni and Ouarda, 2010; Olofintoye *et al*., 2009; Suhaila *et al*., 2011; Rahman *et al*., 2013; Khudri and Sadia, 2013; Marani and Ignaccolo, 2015; Ahmad *et al*., 2016; Ahmad *et al*., 2019).

Extreme value analysis might be helpful to assess both the probability of happening and the magnitude of extreme events. So, extreme value theory authorizes researchers to measure an event's random behaviour that originates in the upper or lower tails. Standard extreme value analysis is frequently performed based on the most straightforward inferential procedure; however, the data structure might be complex. Consequently, the statistical modelling of extreme weather events using extreme value theory agrees to accomplish the complex system that is natural in the extreme data to enhance inferential procedures.

For the modelling of extreme weather variables, the two essential methods are engaged in extreme value theory, namely block maxima and peak-over-threshold (POT) (see, e.g., Coles, 2001; Rivas *et al*., 2008; Bücher *et al*., 2019). In the block maxima method, we model a maximum value of each year acquired from the large sample via generalized extreme value (GEV) distribution (Ferreira and De Haan, 2015). On the other side, the POT method deals with those values that exceed the high threshold level in the observed data with the application of generalized Pareto distribution (Davison and Smith, 1990; Ferreira and De Haan, 2015). However, according to Madsen *et al*. (1997) and Eastoe and Tawn (2012)), the block maxima is the favourite for modelling extremes, whereas in the POT procedure, sometimes selecting an appropriate threshold is not an easy task.

From the advances in the statistical modelling point of view, extreme value analyses customarily have been carried out using frequentist approaches. For instance, Hosking (1990) introduced the linear moments (LM) method for the study of extreme data. Many studies are available in the literature concerned with applying the LM method (for instance, Elamir and Seheult, 2003; Hosking and Wallis, 2005; Khan *et al*., 2021). Furthermore, Ahmad *et al*. (2013) and Ahmad *et al*. (2016) used LM method to model the monsoon rainfalls patterns in Pakistan. They establish the best fit distribution among five extreme value distributions by classical modelling. On the other side, Coles and Dixon (1999), Coles (2001), and Ahmad *et al*. (2019) used likelihood-based inference methods for modelling extreme value models. Researchers are more interested in Bayesian modelling than a classical setup to obtain more valuable results about uncertainty extreme environmental events.

Meanwhile, extreme data are scarce by their nature. The statistical inference on extremes could be enhanced by the Bayesian paradigm's support that allows supplementary evidence about the processes via prior knowledge. For interested readers, many studies exist in the literature (see, e.g., Coles and Tawn, 1996, 2005; Coles and Powell, 1996; Beirlant *et al*., 2004; Chu and Zhao, 2011; Naghettini, 2017, chap. 11; Diriba *et al*., 2017; Ahmad *et al*., 2019; Diriba and Debusho, 2020). However, the Bayesian analysis of extreme events is not dependent on the critical assumptions that are obligatory

for the frequentist framework by the asymptotic theory (Smith, 1985; Coles, 2001; Smith, 2005). Generally, the prior knowledge on extreme value model parameters is rare to discover that fits the probabilistic models via Bayesian procedure. Therefore, impressive results on this strengthen us to increase the precision of the estimates. For instance, Diriba *et al*. (2017) have been examined the properties of priors on the parameter estimates of GEV distribution and the prior effect on return levels for the wind speed data of Cape Town, South Africa.

Furthermore, Ahmad *et al*. (2019)) have studied the effects of priors on the parameters as well as return levels for the rainfall data of Lahore station, Punjab, Pakistan. Although, they did not generalize their results over the country because their findings were limited only to one province. They did not study any other frequentist procedure except maximum likelihood estimation (MLE). These studies inspire the authors to develop extreme rainfall modelling over the country and explore an entire narrative of uncertainty in parameters and return levels (RLs) of GEV distribution. For this reason, some sites over the country are considered for experimental work. Consequently, the essential purpose of this modelling is to predict extreme rainfall events in the future over the country. The occurrence of uncertainty in the future forecasts of extremes makes the study of extreme events even more vital and critical (Coles *et al*., 2003; Zhu *et al*., 2013; First, 2019). Thus, it is mandatory to characterize their behaviour statistically.

In application point of view, the key objective of the study for extreme environmental events is to recognize the properties of the larger RLs of the variable of interest. For instance, the estimates of RLs for an annual maximum of the extreme event could be predicted as these observations provide an expected value of return level that exceeds once, on average, in a given return period (Coles, 2001; Diriba and Debusho, 2020). Hence, statistical findings from the systematic study of extreme climatic events suggest high analytical power. Also, this research study aims to examine extreme climate fluctuations and varying patterns of events which may help to know the behaviour of extreme weather events.

The rest of the paper is organized as follows. In Section 2, the materials and methods are presented for data analysis. The data description with their exploratory analysis and generalized extreme value model with block maxima are explained. Also, both frequentist (MLE and LM methods) and Bayesian Markov Chain Monte Carlo (MCMC) paradigm with noninformative (NIPs) plus Informative Priors (IPs) for parameters estimation and RLs are established in the same section. Also, the assessment measures are described in Section 2. In Section 3, the results are discussed. For instance, how NIPs and IPs affect parameter estimates of GEV and RLs in Bayesian settings, the results of all three methods are compared based on assessment measures. Conclusion and some recommendations are given in the final Section 4.

2 | MATERIALS AND METHODS

2.1 | Data description and exploratory analysis

Throughout this paper, the data comprise a daily aggregate of rainfall (in millimetres) of 11 weather stations all over Pakistan recorded by automatic weather stations. The data had been taken from Pakistan Metrological Center Karachi corresponds to 32 years from 1985 to 2016. Data have been selected on the following standard criteria: the length of the data, variability, quality, urbanization, and climate changes. Later, the ADMRS was extracted from the daily rainfall data using the block maxima method. AMDRS is a single maximum value for any specific year and station among all recorded daily rainfall values. The extracted data of 11 weather stations, namely Lahore, Drosh, Chitral, Jacobabad, Khuzdar, Rohri, Nawabshah, Lasbela, Hyderabad, Chhor, and Pasni, were utilized for analysis and prediction. For deriving of IPs, the rainfall characteristics of two new weather stations at various distances were engaged, namely Mohenjo-daro and Dera Ismael Khan (D.I. Khan). The length of the data for these stations was 30 years from 1987 to 2016.

The D.I. Khan station is considered for the construction of IPs and it is located in the centre of the country. Besides, the D.I. Khan district is situated between district Bhakkar of south Punjab, Mianwali of North Punjab, Zhob of Baluchistan, Indus river, and South Waziristan of Pakistan tribal belt. The Mohenjo-daro station was chosen for prior elicitation due to very short distances from other various sites of the Sindh and Baluchistan provinces. The plots under the Google map spatial linkage encompass selected areas studied in the present research are shown in Figure 1. Blue pinpoints highlight the observed weather stations, and green pinpoints indicate the stations used for IPs construction.

The descriptive analysis for the amount of ADMRS of different stations selected all over the country is briefed in Table 1. The mean of ADMRS fluctuates from 34.44 to 87.61 mm. The ADMRS of the Jacobabad station has comparatively large variation, that is, the observations are more spread (sample CV) than other data sets. One observation (e.g., 323 mm) in the Jacobabad ADMRS is big enough compared to others and may be a source for this large CV. It can be observed that the Drosh station gained less relative variation against other stations in the study.

Besides, most Sindh province stations have larger sample CVs compared to Punjab and Khyber Pakhtunkhwa

FIGURE 1 Spatial map plot of selected sites used in the study for the period of 1985–2016. Blue pinpoints are denoted the observed data and green indicates those sites used for generating the informative priors [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Summary statistics of the AMDRS recorded from different stations

	Summary statistics of selected weather stations						
Study locations	**Minimum (mm)**	**Maximum (mm)**	**Mean**	**Standard deviation**	**Coefficient of variation**	**Skewness**	**Kurtosis**
Jacobabad	3.00	323	58.83	70.87	1.20	2.25	4.94
Rohri	5.00	173.7	46.75	40.58	0.87	1.49	1.97
Nawabshah	0.00	143	42.49	34.52	0.81	1.07	0.49
Hyderabad	4.00	153	48.35	33.63	0.69	0.878	0.77
Chhor	2.30	214.6	71.27	44.26	0.62	1.14	1.36
Mohenjo-daro	5.00	119.6	34.44	25.44	0.79	1.28	1.63
Khuzdar	15.00	223	43.33	34.55	0.80	4.34	19.87
Lasbela	8.00	269.6	48.22	46.55	0.96	3.49	13.41
Pasni	0.00	131.8	35.35	26.64	0.75	1.56	3.13
D.I. Khan	21.1	150	59.09	27.01	0.45	1.49	2.47
Drosh	22	131	51.95	21.21	0.41	2.10	5.15
Chitral	23.40	161.20	51.43	28.62	0.55	2.13	4.81
Lahore	29.40	189.70	87.61	38.71	0.44	0.89	0.17

(KPK). The ADMRS of two Baluchistan sites (i.e., Lasbela and Khuzdar) and one station of KPK province as Chitral are pretty more skewed compared to other data sets.

Moreover, it is necessary to test the fundamental assumptions of any annual maximum series before conducting a final analysis in the field of statistical hydrology

because the final results could be doubtful without satisfying the basic assumptions. The hydrological series fundamental assumptions are independence, homogeneity, randomness, and stationarity (Naghettini, 2017, chap. 7). Thus, we ensured that the data fulfilled the basic assumptions and can be used to model extreme rainfall. The most suitable probability model for ADMRS is decided among various models (generalized extreme value, Pearson type three, generalized Pareto, generalized logistic) by using some nonparametric procedures. Hence, the GEV distribution remains most suitable for the observed ADMRS from various country sites. Moreover, the best-fitted probability model and estimation of its parameters via frequentist and Bayesian techniques are presented in the subsequent section.

2.2 | Block maxima and generalized extreme value distribution

To model the extreme observations using GEV, a data of N independent values $w_1, w_2, ..., w_N$ is first blocked into k block of size n, with n essentially large, and hence $N=kn$. For rainfall data, the block size is usually a month, season or a year. For instance, 1 year stands for $n \approx 365$ days. Then the maxima or extreme value $(M_i, i=1,...,k)$ is selected from each block. This produces a data of k annual maxima series named block maxima to which the GEV distribution family can be fitted. Suppose the yearly maxima $w_1, w_2, ..., w_n$ are independent and identically distributed (i.i.d) with distribution function of $G(w)$. Let $M_n = \max(w_1, w_2, ..., w_n), n \in \mathrm{N}$ and if there are sequences of normalizing constants $\{c_n > 0\}$ and $d_n \in$ such that

$$pr\left\{\frac{(M_n - d_n)}{c_n} \leq w\right\} \to G^n(c_n w + d_n) \to F(w) \qquad (1)$$

as $n \to \infty$, where F is a nondegenerate distribution function, the distribution function G is called to be in the domain of attraction of extreme value distribution F, {i.e., $G \in F(w)$}. Besides, the F follows the family of the probability distribution that has the form

$$F(w,\mu,\sigma,\kappa) = \begin{cases} \exp\left[-\left\{1+\kappa\left(\frac{w-\mu}{\sigma}\right)\right\}^{-\frac{1}{\kappa}}\right], \kappa \neq 0 \\ \exp\left[-\exp\left(-\frac{w-\mu}{\sigma}\right)\right], \kappa = 0 \end{cases} \qquad (2)$$

where $\{w : 1+\kappa(w-\mu)/\sigma\}$, $\mu, \sigma > 0$, and κ are location, scale, and shape parameters of GEV distribution (Beirlant *et al.*, 2004). The shape parameter affects the behaviour of the upper tail of the distribution. Moreover, GEV distribution is the mixture of three limiting extreme value distributions, that is, Gumbel distribution, Freshet distribution, and Weibull distribution. If $\kappa \to 0$ the GEV distribution in (2) relates to the Gumbel distribution. For $\kappa > 0$ and $\kappa < 0$, the expression given in (2) called Frechet and negative Weibull distributions, respectively (Coles, 2001).

2.3 | Parameter estimation of GEV distribution

Primarily, the maximum likelihood estimation method and linear moments method were applied to estimate GEV distribution parameters. In MLE, we differentiate the function given in (2) for w_i, for instance, when $\kappa \neq 0$ the density of GEV is given by

$$f(w_i,\kappa,\mu,\sigma) = \frac{1}{\sigma}\left[1+\kappa\frac{(w_i-\mu)}{\sigma}\right]^{-\left(1+\frac{1}{\kappa}\right)} \exp\left[-\left\{1+\kappa\frac{(w_i-\mu)}{\sigma}\right\}^{-\frac{1}{\kappa}}\right] \qquad (3)$$

The maximum likelihood estimates (MLEs) of the parameters μ, σ, and κ, say $\hat{\mu}, \hat{\sigma}$, and $\hat{\kappa}$ are calculated by maximizing the logarithm of the joint likelihood, that is, maximizing

$$l(\mu,\sigma,\kappa; w_1, w_2, .., w_n) = -n\log\sigma - \left(1+\frac{1}{\kappa}\right)\sum_{i=1}^{n}\log\left[1+\kappa\frac{(w_i-\mu)}{\sigma}\right] - \sum_{i=1}^{n}\left\{1+\kappa\frac{(w_i-\mu)}{\sigma}\right\}^{-\frac{1}{\kappa}} \qquad (4)$$

concerning unknown parameters, say μ, σ, and κ. Since the solution of log-likelihood is not an easy task, in particular, the maximization is solved by quasi-Newton procedure with numerical iteration (Diriba *et al.*, 2017; Diriba and Debusho, 2020).

In LM method computations, we use the linear combinations of order statistics values. This method was introduced by Hosking (1990). The LM provides simple and more efficient estimators of extremal data characteristics and the parameters of the distribution. Let $W_1, W_2, ..., W_r$ be the random sample of magnitude n, with cumulative distribution function $F(w)$ and quantile function $w(F)$. Suppose $W_{1:r} \leq W_{2:r} \leq ... \leq W_{r:r}$ be the order statistics of the selected random samples. For the random variable W, the rth population LM as explained by Ahmad *et al.* (2016):

$$\lambda_r = \frac{1}{r}\sum_{i=0}^{r-1}(-1)^k\binom{r-1}{i}E(X_{r-i:r}),\ r=1,2,... \qquad (5)$$

Usually, we require the first four LM for $r=1,2,3,4$. Additionally, LM can also be considered as the linear combinations of probability-weighted moments as given:

$$\lambda_{r+1}=\sum_{i=0}^{r}\beta_i(-1)^{r-i}\binom{r}{i}\binom{r-1}{i} \tag{6}$$

The first four LM are

$$\lambda_1=\mu+\frac{\sigma}{\kappa}\{1-\Gamma(1+\kappa)\}, \tag{7}$$

$$\lambda_2=\frac{\sigma}{\kappa}\{(1-2^{-\kappa})\Gamma(1+\kappa)\}, \tag{8}$$

$$\tau_3=\frac{2(1-3^{-\kappa})}{(1-2^{-\kappa})}-3, \tag{9}$$

$$\tau_4=\frac{\{5(1-4^{-\kappa})-10(1-3^{-\kappa})+6(1-2^{-\kappa})\}}{(1-2^{-\kappa})} \tag{10}$$

The LM ratios $\tau=\lambda_2/\lambda_1$, τ_3, and τ_4 denote the linear coefficient of variation, linear skewness, and linear kurtosis, respectively. The GEV parameters were estimated by using some approximations discussed by Hosking and Wallis (2005, p. 196). Moreover, the theoretical estimates of the parameters of GEV are given as follows

$$\hat{\mu}=\lambda_1+\frac{\hat{\sigma}[\Gamma(1+\hat{\kappa})-1]}{\hat{\kappa}} \tag{11}$$

$$\hat{\sigma}=\frac{\lambda_2\hat{\kappa}}{(1-2^{-\hat{\kappa}})\Gamma(1+\hat{\kappa})} \tag{12}$$

$$\hat{\kappa}=7.8590\hat{c}+2.95554\hat{c}^2 \tag{13}$$

where

$$\hat{c}=\frac{2}{3+\tau_3}-\frac{\log 2}{\log 3} \tag{14}$$

2.4 | Return level estimation for GEV model

The attention to extreme environmental events analysis sometimes does not generally rely on the estimates of extreme value distribution parameters; however, it applies the fitted model to calculate the other quantities. Return level estimates play a dynamic role in rainfall modelling for calculating the future hazard associated with return periods conforming to a fitted model. For instance, the estimates of extreme quantiles for ADMRS of an event could be calculated because these observations assess the return level of the event predicted on averagely exceeds once in a specific number of years.

The RLs for the GEV model corresponding to the return period $T=1/p$, denoted by w_p where $F(w_p=1-p)$ and $0<p<1$, is attained by using quantile function by the inverse of (2) given by (Coles, 2001) and also discussed by Ahmad *et al*. (2019) & Diriba and Debusho (2020).

$$w_p=\begin{cases}\mu-\frac{\sigma}{\kappa}[1-\{-\log(1-p)\}^{-\kappa}], & \kappa\neq 0,\kappa=0\\ \mu-\sigma\log\{-\log(1-p)\},\end{cases} \tag{15}$$

The return level w_p is determined by quantiles of GEV distribution associated with the upper tail probability p. For the GEV model, the MLEs of the return level w_p, indicated by $\hat{w}_p$ is gained by substituting the MLEs $\hat{\mu},\hat{\sigma}$, and $\hat{\kappa}$ (Rao, 1973).

2.5 | Bayesian analysis

As in the maximum likelihood procedure, suppose ADMRS $w=(w_1,w_2,...,w_n)$ are i.i.d and their distribution falls within the parametric GEV family. Moreover, now in the Bayesian setting, the GEV distribution parameters (μ,σ, and κ) are dealt as random variables for which we identify the prior distributions. The prior information helps us enhance the knowledge provided by the observed data. Let $\theta=(\mu,\sigma,\kappa)$ and suppose the prior for θ with no evidence to the actual data can be expressed by a probability density function $g_\theta(\theta)$. Then using Bayes theorem to combine the likelihood and prior knowledge and to get the posterior density for θ has the following form:

$$f(\theta/w)=\frac{L(\theta/w)g_\theta(\theta)}{\int_\Theta L(\theta/w)g_\theta(\theta)d\theta}\propto L(\theta/w)g_\theta(\theta) \tag{16}$$

Where $L(\theta/w)$ indicates the likelihood function of GEV distribution given in (4) and $f(\theta/w)$ is the posterior distribution for θ, and the integral is set over the parametric space Θ. In this research, both the NIPs and IPs were engaged. The NIPs were specified by considering there is least or no external information accessible about the parameters, separate from the data. To generate the NIPs for the GEV parameter designated $\theta=(\mu,\sigma,\kappa)$, the parametrization $\varphi=\log\sigma$ is done in the place of σ due to more manageable tasks in the specification of prior and to secure the positivity of scale parameter σ. Since for priors specification, the joint density for θ was supposed in the following form

$$g(\theta)=g(\mu,\varphi=\log\sigma,\kappa)=g_{\mu}(\mu)g_{\varphi}(\varphi)g_{\kappa}(\kappa) \tag{17}$$

The following marginal independent NIPs in different studies (Coles and Tawn, 2005; Fawcett and Walshaw, 2008; Eli *et al*., 2012; Diriba *et al*., 2017; Ahmad *et al*., 2019; Diriba and Debusho, 2020) were used.

$$g_{\mu}(\mu)\sim N(0,10{,}000), g_{\varphi}(\varphi)\sim N(0,10{,}000), g_{\kappa}(\kappa)\sim N(0,100) \tag{18}$$

These are known as independent Gaussian priors with mean 0 and large variances (e.g., $N(0,10{,}000)$ indicates Gaussian distribution with 0 mean and 10,000 variances). The higher variances of the distribution enforce enough to create flat marginal priors, which confirm the lack of external information. On the other hand, the IPs were built by using the procedure given by (Coles and Tawn, 1996), that is, prior knowledge was provoked in terms of extreme quantiles. The method engaged for GEV is briefly described in the following paragraph.

Remember that the return level $w_{p_i}, i=1,2,3$ in expression (15) with $p_1>p_2>p_3$, be the quantiles calculated corresponding to T return period from historical extreme rainfall data of two suitable sites over the country. For example, the quantiles w_{p_i} are estimated independently for both Mohenjo-daro and D.I. Khan stations by replacing MLEs of GEV parameters in Equation (2). Coles and Tawn (1996) discuss a joint prior distribution for GEV parameters generated from extreme quantile $(w_{p_1}, w_{p_2}, w_{p_3})$ by employing given probabilities $p_1>p_2>p_3$. One minor complication with these techniques is that the quantiles $w_{p_i}, i=1,2,3$ must be in natural order (e.g., $w_{p_1}<w_{p_2}<w_{p_3}$); hence, the fundamental assumption of independent priors $w_{p_i}, i=1,2,3$ would not be fulfilled. Subsequently, they recommended to use the quantile differences:

$$\tilde{w}_{p_i}=w_{p_i}-w_{p_{i-1}}, i=1,2,3 \tag{19}$$

where w_{p_0} denotes a physically lower endpoint of the process variable (e.g., rainfall) and supposed to be $w_{p_0}=0$. It can be noticed that the change in quantile endorses the ordering of quantiles. Since independent marginal priors based on the quantile differences are now supposed to be independent gamma distribution with parameters (ν_i,γ_i), $i=1,2,3$, and can be written in the following form:

$$\tilde{w}_{p_i}\sim G(\nu_i,\gamma_i)\quad \nu_i>0,\ \gamma_i>0;\ i=1,2,3. \tag{20}$$

From Equations (19) and (20), we can develop the joint prior for the (ν_i,γ_i), $i=1,2,3$, from the Gamma distribution in the following form

$$\tilde{w}_{p_1}\sim G(\nu_1,\gamma_1)\propto w_{p_1}^{\nu_1-1}\exp\left(-\gamma_1 w_{p_1}\right),$$

$$\tilde{w}_{p_2}\sim G(\nu_2,\gamma_2)\propto \left(w_{p_2}-w_{p_1}\right)^{\nu_2-1}\exp\left(-\gamma_2\left(w_{p_2}-w_{p_1}\right)\right),$$

and

$$\tilde{w}_{p_3}\sim G(\nu_3,\gamma_3)\propto \left(w_{p_3}-w_{p_2}\right)^{\nu_3-1}\exp\left(-\gamma_3\left(w_{p_3}-w_{p_2}\right)\right).$$

Then the joint prior for $(w_{p_1}, w_{p_2}, w_{p_3})$, by considering $w_{p_0}=0$, is stated as

$$g\left(w_{p_1},w_{p_2},w_{p_1}\right)\propto w_{p_1}^{\nu_1-1}\exp\left(-\gamma_1 w_{p_1}\right)\times\left(w_{p_2}-w_{p_1}\right)^{\nu_2-1}\exp\left\{-\gamma_2\left(w_{p_2}-w_{p_1}\right)\right\}\times\left(w_{p_3}-w_{p_2}\right)\exp\left\{-\gamma_3\left(w_{p_3}-w_{p_2}\right)\right\}$$

and has written in a short form

$$g\left(w_{p_1},w_{p_2},w_{p_1}\right)\propto \tilde{w}_{p_1}^{\nu_1-1}\exp\left(-\gamma_1\tilde{w}_{p_1}\right)\times\prod_{i=2}^{3}\tilde{w}_{p_i}^{\nu_i-1}\exp\left(-\gamma_i\tilde{w}_{p_i}\right) \tag{21}$$

with $w_{p_1}<w_{p_2}<w_{p_3}$ discussed by Diriba *et al*. (2017), Ahmad *et al*. (2019), and Diriba and Debusho (2020). Then incorporating expression (15) in Equation (21) and multiply by the Jacobian of the transformation from $\left(w_{p_1},w_{p_2},w_{p_3}\right)\rightarrow\theta=(\mu,\sigma,\kappa)$, it provides an expression for the prior in terms of the GEV parameter vector θ. Moreover, it has the following form

$$g_{\theta}(\theta)\propto\prod_{i=1}^{3}\tilde{w}_{p_i}^{\nu_i-1}\exp\left(-\gamma_i\tilde{w}_{p_i}\right)\times \mathrm{J} \tag{22}$$

for $w_{p_1}<w_{p_2}<w_{p_3}$. According to Diriba and Debusho (2020), the Jacobean used in (22) is found in such a way

$$\mathrm{J}=\begin{vmatrix}\frac{\partial w_{p1}}{\partial\mu} & \frac{\partial w_{p1}}{\partial\sigma} & \frac{\partial w_{p1}}{\partial\kappa}\\ \frac{\partial w_{p2}}{\partial\mu} & \frac{\partial w_{p2}}{\partial\sigma} & \frac{\partial w_{p2}}{\partial\kappa}\\ \frac{\partial w_{p3}}{\partial\mu} & \frac{\partial w_{p3}}{\partial\sigma} & \frac{\partial w_{p3}}{\partial\kappa}\end{vmatrix}$$

and their results expressed in Equation (23) as follows:

$$\mathrm{J}=\begin{cases}\frac{\sigma}{\kappa^2}\left|\sum_{i<j}(-1)^{i+j}(z_i\times z_j)\log\left(\frac{z_j}{z_i}\right)\right|, & i,j\in(1,2,3);\kappa\neq 0\\ \frac{\sigma}{2}\left|\sum_{i<j}(-1)^{i+j}\log z_i\times\log z_j\log\left(\frac{z_j}{z_i}\right)\right|, & i,j\in(1,2,3);\kappa=0\end{cases} \tag{23}$$

where $z_i=-\log(1-p_i)$, $i=1,2,3$.

2.6 | Assessment measures

Assessment measures were used to compare the performance between the classical or frequentist (i.e., MLE and LM approaches) and the Bayesian MCMC paradigm with NIPs and IPs in estimating GEV parameters and RLs for ADMRS recorded from different weather stations over the country. Moreover, these measures could distinguish the accuracy among results obtained through classical and Bayesian procedures. So, the proposed assessment measures are relative root mean square error (RRMSE), relative absolute error (RAE). Both measures encompass evaluating the difference between the observed and the estimated values corresponding to the assumed distribution (Ahmad *et al.*, 2019). The mathematical form of the tests are given in (24) and (25)

$$\mathrm{RRMSE}=\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left[\left|\frac{w_{j:n}-\hat{q}\left(F_j\right)}{w_{j:n}}\right|\right]^2} \tag{24}$$

$$\mathrm{RAS}=\frac{1}{n}\sum_{j=1}^{n}\left|\frac{w_{j:n}-\hat{q}\left(F_j\right)}{w_{j:n}}\right| \tag{25}$$

where $w_{j:n}$ means the observed sample values of jth order statistics of a random sample, while $\hat{q}\left(F_j\right)$ are estimated quantiles parallel to jth Weibull plotting position

TABLE 2 Estimated parameters with their SE through MLE and LM methods for the GEV model based on ADMRS from different stations over the country

	MLEs			L-moments estimates		
Study locations	$\hat{\mu}(\boldsymbol{SE})$	$\hat{\sigma}(\boldsymbol{SE})$	$\hat{\kappa}(\boldsymbol{SE})$	$\hat{\mu}(\boldsymbol{SE})$	$\hat{\sigma}(\boldsymbol{SE})$	$\hat{\kappa}(\boldsymbol{SE})$
Rohri	26.15 (4.73)	22.14 (4.02)	0.30 (0.20)	26.63 (4.72)	23.84 (4.08)	0.21 (0.21)
Nawabshah	25.26 (4.51)	21.95 (3.60)	0.18 (0.16)	24.91 (4.40)	22.61 (3.64)	0.16 (0.15)
Hyderabad	32.45 (5.29)	25.05 (4.09)	0.05 (0.18)	33.06 (5.27)	27.65 (4.09)	–0.03 (0.11)
Chhor	51.04 (6.28)	31.66 (4.64)	0.05 (0.13)	49.85 (6.29)	31.36 (4.60)	0.10 (0.12)
Khuzdar	32.14 (2.24)	11.62 (1.74)	0.21 (0.10)	31.15 (2.35)	9.735 (1.77)	0.41 (0.10)
Lasbela	30.89 (3.44)	17.61 (2.76)	0.25 (0.12)	29.50 (3.45)	14.64 (2.78)	0.42 (0.11)
Pasni	22.85 (3.38)	16.77 (2.62)	0.15 (0.15)	22.51 (3.44)	17.21 (2.58)	0.14 (0.14)
Jacobabad	23.58 (4.81)	22.47 (4.97)	0.63 (0.23)	25.17 (4.79)	25.54 (4.88)	0.43 (0.11)
Drosh	43.04 (2.51)	13.01 (1.83)	0.09 (0.10)	42.40 (2.48)	11.80 (1.92)	0.19 (0.12)
Chitral	37.99 (2.40)	12.08 (2.12)	0.37 (0.15)	37.57 (2.38)	11.78 (2.01)	0.38 (0.14)
Lahore	69.30 (5.67)	27.88 (4.29)	0.07 (0.16)	68.84 (5.64)	28.82 (4.21)	0.07 (0.15)

TABLE 3 Estimated return levels for ADMRS of different weather stations over the country by using MLE and L-moments methods

	MLE method					L-moment method				
Study locations	**10**	**25**	**50**	**100**	**500**	**10**	**25**	**50**	**100**	**500**
Rohri	97.40	145.22	190.62	246.27	429.89	95.56	136.13	172.02	213.41	336.37
Nawabshah	86.71	121.42	151.46	185.46	283.28	86.31	119.68	147.97	179.42	267.26
Hyderabad	92.39	119.90	141.24	163.24	217.34	93.57	118.07	135.88	153.23	192.21
Chhor	127.18	162.23	189.55	217.78	287.51	128.73	167.47	198.613	231.69	317.27
Khuzdar	65.82	85.69	103.29	123.62	184.13	67.15	95.53	124.99	163.96	310.77
Lasbela	84.28	117.64	148.06	184.14	296.44	84.35	128.25	174.17	235.35	468.66
Pasni	67.78	91.75	111.87	134.07	195.20	68.43	92.81	113.23	135.69	197.33
Jacobabad	136.43	259.55	413.21	651.77	1846.23	122.63	202.22	286.44	399.75	839.37
Drosh	75.67	91.63	104.43	117.10	152.99	75.627	94.536	110.96	129.62	183.62
Chitral	80.64	112.48	144.57	185.96	335.03	79.605	111.42	143.69	185.56	337.96
Lahore	137.57	169.91	195.40	222.05	289.15	139.08	172.16	198.16	225.26	293.18

$F=j/(n+1)$, where j shows the ranks of the data. The method that gained the lowest RRMSE and RAS would be considered an efficient method for modelling Pakistan's extreme rainfall data.

3 | RESULTS AND DISCUSSION

3.1 | Generalized extreme value distribution using frequentist methods

In this section, the GEV model of Equation (2) with a block maxima was fitted to ADMRS using MLE and LM methods. Moreover, the parameter estimates ($\hat{\mu}, \hat{\sigma}$, and $\hat{\kappa}$) with their standard errors (SE) of the GEV stationary model are given in Table 2.It can be noted from Table 2 that the shape parameter $\hat{\kappa}$ obtained through the LM method is less than zero for the Hyderabad station. This specifies a bounded upper tail to the distribution of ADMRS. The negative shape parameter suggests a heavier tail for GEV distribution, which offers smaller quantiles, particularly when the quantiles for ADMRS are estimated for longer return periods (Hosking, 1990).

The estimated RLs for different return periods are presented in Table 3, for both MLE and LM procedures. As a result, the RL estimates are smaller and consistent for the LM method, while RLs are larger for the MLE method. Moreover, the variations in RLs could be due to the skewness in the ADMRS. Consequently, to distinguish the summary of a series for skewed distributions, the median is more energetic to deal with outliers than the mean. The variation between the median and the mean can signify the magnitude of unusual values in the RLs (Diriba and Debusho, 2020). Hence, this task is further examined with the Bayesian paradigm application.

3.2 | Bayesian modelling of ADMRS using noninformative and informative priors

This section deals with inferences about ADMRS obtained through Bayesian analysis by NIPs and IPs support. The IPs were constructed independently for Mohenjo-daro and D.I. Khan. Therefore, two sites were used to elicit prior distributions. Both sites were lying at different distances from the observed sites. The distance's influence on the parameters and the RLs were evaluated in general.

The NIPs were built for the GEV parameters $\theta=(\mu,\sigma,\kappa)$ by assuming there is no reliable prior information about the process to express the prior distributions apart from the data. Thus, the priors joint density for θ given in expression (17) was assumed with the parametrization $\varphi=\log\sigma$. Hence, the noninformative independent priors given in (18) were incorporated. The scale parameter of GEV (i.e., σ) was reparametrized as $\varphi=\log\sigma$ to hold the positive property of variance. The Gaussian distributions having zero mean and higher variances in expression (18) enforce the flat independent marginal priors (also known as diffuse or vague priors), which exhibit the lack of external information (Eli *et al.*, 2012; Ahmad *et al.*, 2019).

The MCMC procedure with the Metropolis-Hasting (M-H) recipe was used for generating the samples from posterior distributions. By applying H-M, 50,000

TABLE 4 Posterior means, standard deviations and confidence intervals for GEV parameters to ADMRS data of different weather stations via noninformative priors

	Parameters estimates of GEV distribution		
Study locations	$\hat{\mu}(SE)$ **[CI]**	$\hat{\sigma}(SE)$ **[CI]**	$\hat{\kappa}(SE)$ **[CI]**
Rohri	26.07 (4.97) [17.13, 36.44]	23.80 (4.68) [15.97, 34.23]	0.35 (0.22) [0.01, 0.87]
Nawabshah	25.21 (4.83) [16.43, 35.29]	23.77 (4.19) [16.84, 33.50]	0.22 (0.18) [−0.09, 0.61]
Hyderabad	31.96 (5.42) [21.52, 42.95]	26.56 (4.51) [18.85, 36.27]	0.13 (0.19) [−0.17, 0.57]
Chhor	50.61 (6.65) [37.99, 64.14]	34.15 (5.51) [25.20, 46.74]	0.08 (0.13) [−0.15, 0.37]
Khuzdar	32.12 (2.45) [27.46, 37.11]	12.56 (2.08) [9.17, 17.33]	0.25 (0.11) [0.07, 0.51]
Lasbela	30.94 (3.71) [24.04, 38.67]	19.04 (3.30) [13.63, 26.51]	0.29 (0.13) [0.08, 0.60]
Pasni	22.95 (3.56) [16.17, 30.26]	18.19 (2.98) [13.24, 24.95]	0.18 (0.15) [−0.07, 0.51]
Jacobabad	24.23 (5.11) [15.43, 35.35]	24.92 (5.93) [15.57, 39.04]	0.66 (0.23) [0.27, 1.18]
Drosh	42.78 (2.69) [37.88, 48.36]	14.046 (2.160) [10.444, 18.92]	0.12 (0.12) [−0.05, 0.37]
Chitral	38.18 (2.58) [33.47, 43.60]	13.251 (2.53) [9.158, 19.17]	0.39 (0.16) [0.12, 0.74]
Lahore	69.18 (5.99) [57.54, 81.57]	30.205 (4.91) [21.846, 41.18]	0.09 (0.16) [−0.17, 0.42]

iterations were produced for all sites, of which the initial 10,000 were burn-in. For simulation, different starting points were considered to perceive the chain convergence. Hence, all the chains had mixed well with the original data. The posterior means (PMs), standard deviations (*SD*s), and 95% CIs via NIPs for the GEV parameters for different sites are given in Table 4. Hence, it can be observed that the PMs and SDs are close to MLEs and LM estimates of GEV parameters except for the shape parameter estimates for various sites obtained through L-moments (see Table 2). It is anticipated that for vague priors PMs would be close to L-moments and MLEs as they incorporate slight evidence to the likelihood.

On the other hand, the historical data of rainfall of two weather stations, namely Mohenjo-daro and D.I. Khan were used to formulate the IPs. Using the procedure given in Section 2.5, the prior distributions for the GEV model were built with the support of quantiles $p_i = 10^{-i}$, $i = 1, 2, 3$. From the historical data of Mohenjo-daro station, we found $\tilde{w}_{p_1} \sim G(8.760, 8.071)$, $\tilde{w}_{p_2} \sim G(0.423, 273.255)$, and $\tilde{w}_{p_3} \sim G(0.076, 3400.395)$. Similarly, from D.I. Khan station historical data, authors have also attained $\tilde{w}_{p_1} \sim G(39.612, 2.457), \tilde{w}_{p_2} \sim G(1.301, 50.941)$, and $\tilde{w}_{p_3} \sim G(0.294, 282.395)$. The PMs with their SDs and 95% CIs of GEV parameters from IPs are also given in Table 5. The findings show that PMs for the GEV model parameters from the settings of informative priors are very close to the results gained from the NIPs. Also, the IPs built from Mohenjo-daro and D.I. Khan stations abridged the posterior SDs of GEV parameters for various sites than the SDs obtained via NIPs and frequentist methods. The smaller SDs indicate a reduction in uncertainty. This is happened due to the use of historical information from two nearby weather stations over the country.

To understand how GEV parameters were affected by the IPs based on the historical data of two different weather stations. The posterior densities (PDs) of the parameters found through NIPs and IPs were compared. The estimated densities of GEV parameters (μ, σ, and κ) for three sites, namely Hyderabad, Khuzdar, and Chitral are plotted in Figure 2. The posterior densities of the parameters of the model for remaining are given in Figure S1 (see, e.g., supporting information file). Notice that the distributions of location parameters are symmetric for all three sites.

TABLE 5 Posterior means, standard deviations, and confidence intervals for GEV parameters to ADMRS data of different weather stations via informative priors

Study locations	Mohenjo-daro			D.I. Khan		
	$\hat{\mu}(SE)$[CI]	$\hat{\sigma}(SE)$[CI]	$\hat{\kappa}(SE)$[CI]	$\hat{\mu}(SE)$[CI]	$\hat{\sigma}(SE)$[CI]	$\hat{\kappa}(SE)$[CI]
Rohri	24.92 (4.59) [16.58, 34.53]	21.75 (3.89) [15.15, 30.19]	0.25 (0.14) [−0.01, 0.54]	27.40 (4.37) [19.33, 36.47]	23.94 (3.73) [17.69, 32.04]	0.17 (0.108) [−0.03, 0.38]
Nawabshah	24.40 (4.49) [15.86, 33.88]	22.34 (3.37) [16.09, 30.59]	0.16 (0.14) [−0.09, 0.45]	26.52 (4.38) [18.27, 35.49]	24.36 (3.70) [18.29, 32.70]	0.12 (0.12) [−0.11, 0.34]
Hyderabad	30.83 (5.18) [21.12, 41.45]	25.08 (4.07) [17.81, 34.11]	0.07 (0.14) [−0.17, 0.37]	32.32 (5.02) [22.91, 42.28]	26.41 (4.01) [19.56, 35.30]	0.04 (0.12) [−0.18, 0.29]
Chhor	47.92 (6.22) [36.15, 60.40]	30.86 (4.54) [23.54, 40.98]	0.04 (0.10) [−0.14, 0.25]	47.54 (5.81) [36.18, 59.15]	30.02 (4.06) [23.26, 39.11]	0.022 (0.08) [−0.14, 0.19]
Khuzdar	31.97 (2.12) [27.48, 36.82]	12.35 (1.65) [9.29, 16.58]	0.24 (0.09) [0.07, 0.45]	33.38 (2.12) [28.73, 38.20]	13.87 (1.66) [10.58, 18.09]	0.26 (0.09) [0.09, 0.44]
Lasbela	30.07 (3.45) [23.60, 37.13]	17.73 (2.64) [13.29, 23.61]	0.25 (0.10) [0.07, 0.46]	31.70 (3.45) [25.08, 38.67]	19.03 (2.57) [14.56, 24.64]	0.23 (0.08) [0.07, 0.40]
Pasni	22.54 (3.47) [15.98, 29.62]	17.67 (2.73) [13.12, 23.91]	0.16 (0.12) [−0.06, 0.43]	24.73 (3.55) [18.15, 32.03]	19.91 (2.88) [15.11, 26.42]	0.17 (0.11) [−0.06, 0.39]
Jacobabad	22.15 (4.26) [14.69, 31.20]	20.37 (3.93) [13.63, 29.13]	0.43 (0.12) [0.18, 0.68]	24.91 (4.16) [17.12, 33.55]	22.79 (3.72) [11.39, 19.32]	0.29 (0.09) [0.12, 0.47]
Drosh	42.66 (2.67) [37.51, 47.96]	12.45 (2.13) [8.90, 17.22]	0.34 (0.12) [0.11, 0.59]	39.11 (2.48) [34.47, 44.26]	13.68 (2.18) [9.98, 18.53]	0.31 (0.10) [0.09, 0.50]
Chitral	37.82 (2.46) [33.30, 42.95]	12.45 (2.13) [8.90, 17.22]	0.34 (0.12) [0.11, 0.59]	39.12 (2.48) [34.47, 44.26]	13.68 (2.18) [9.98, 18.53]	0.30 (0.10) [0.09, 0.50]
Lahore	66.87 (5.65) [56.04, 78.54]	27.22 (4.27) [20.11, 37.02]	0.05 (0.12) [−0.17, 0.29]	65.82 (5.22) [55.86, 76.33]	26.10 (3.84) [19.74, 34.81]	0.03 (0.10) [−0.16, 0.23]

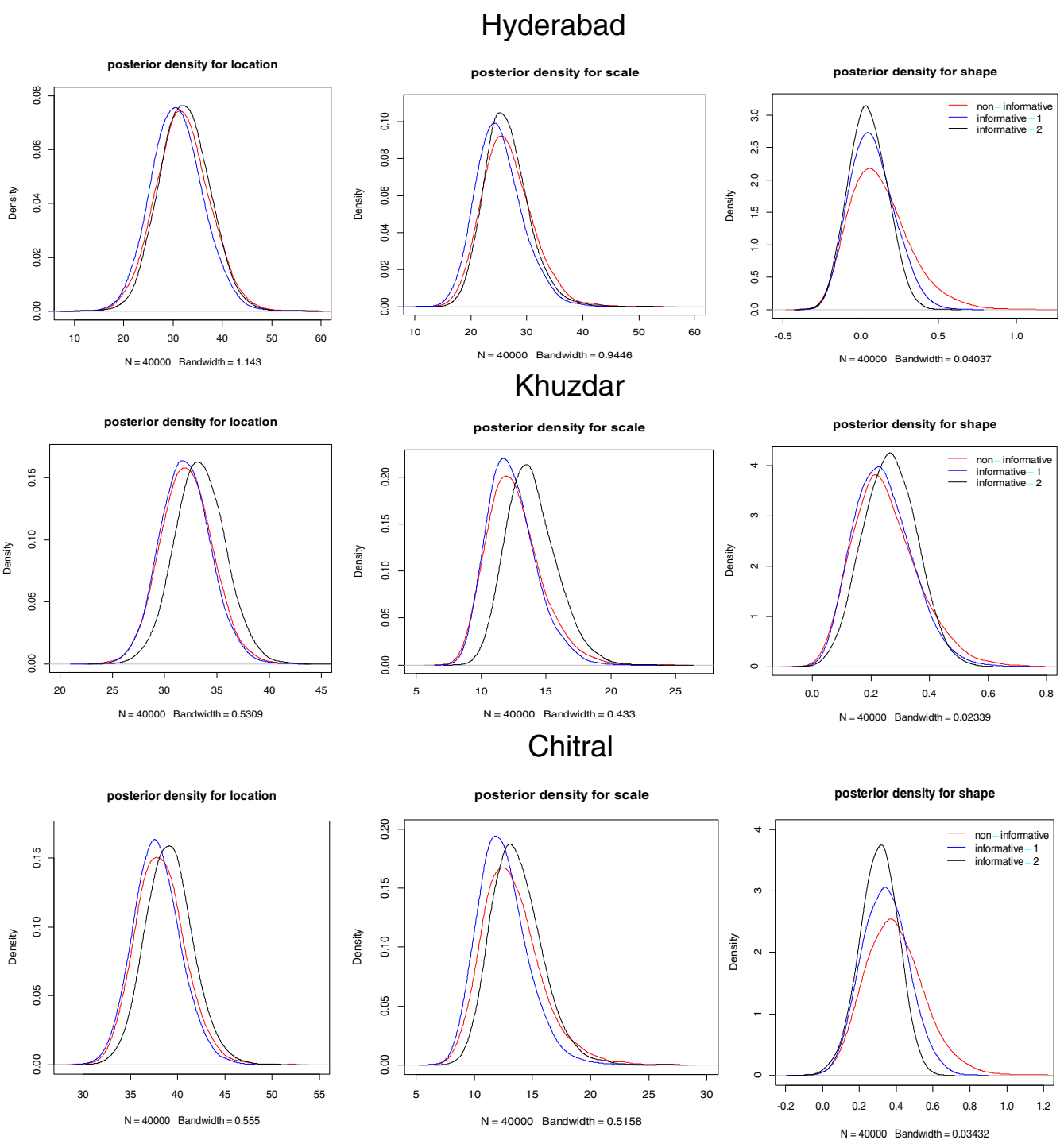

FIGURE 2 Posterior densities of GEV distribution parameters for Hyderabad, Khuzdar and Chitral weather stations via noninformative and informative priors [Colour figure can be viewed at wileyonlinelibrary.com]

On the other hand, the distributions of scale and shape parameters are positively skewed. Furthermore, the location, scale, and shape parameters for IPs are built using the information from Mohenjo-daro and D.I. Khan showing high peaks. The fluctuations in posterior densities indicate that the posterior distributions are sensitive to the IPs from which the prior knowledge was produced. As we discussed earlier, IPs were based on the knowledge that considered the functions of two stations over the country with the combination of mean and quantiles of the data. Weather conditions of those regions from which the data acquired are assumed to be relatively homogeneous. For instance, the Lahore station was selected by pretending that the environmental conditions of other

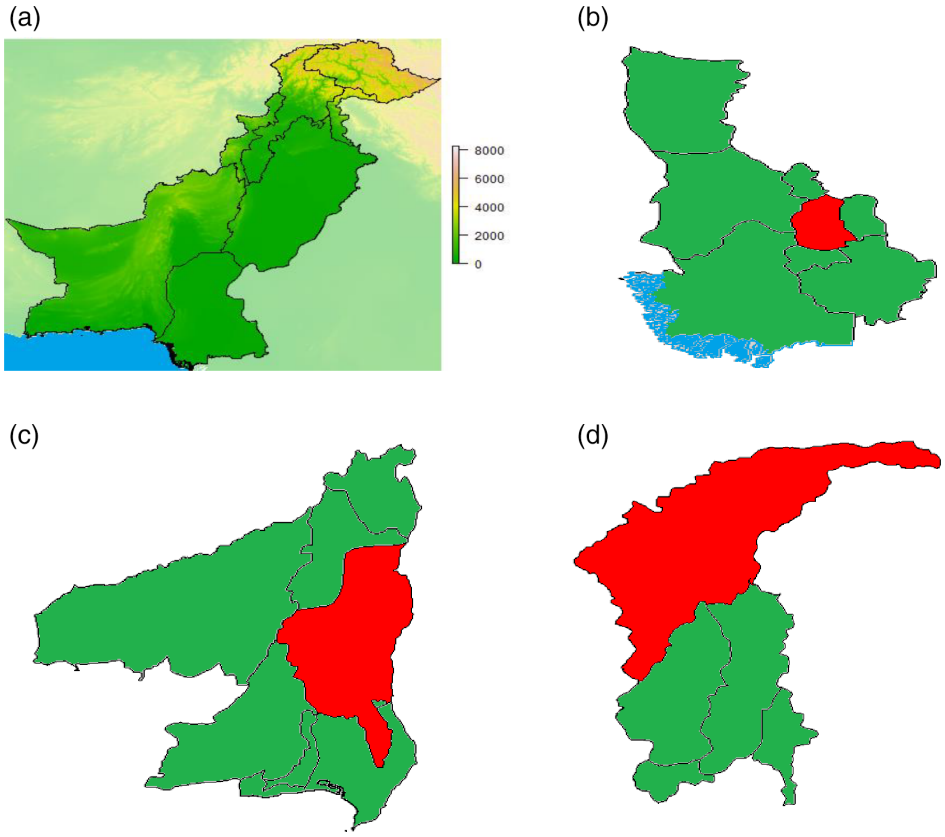

FIGURE 3 (a) The Pakistan spatial map with the elevation above the sea level (m), (b) the Hyderabad division of the province Sindh, the red colour indicates Hyderabad district, which is most affected due to heavy rains, (c) the Kalat division of the province Baluchistan, the red colour represents the affected site namely Khuzadar, and (d) described the Malakand division of the province Khyber Pakhtunkhwa, the red colour indicates the affected area Chitral with substation Darosh [Colour figure can be viewed at wileyonlinelibrary.com]

cities of the province Punjab similar to Lahore. According to our evaluation, the estimates of GEV parameters based on IPs are susceptible. Therefore, selecting appropriate weather stations for the formulations of IPs is also an important task. The validation and robustness of estimates are verified by observing the effect of other IPs on the model parameters. For this purpose, the IPs were elicited by exploiting the Jaisalmer station historical records of the neighbouring country India. Figure S3 (see, e.g., supporting information file) shows that the influence of IPs on parameters of the models for Rohri data was similar to IPs generated from Mohenjo-daro and D.I. Khan historical records. Based on this evidence, our current methodology could be extended easily to other ungagged sites of the country. Moreover, the same discussion can be followed to interpret the remaining stations namely Rohri, Nawabshah, Chhor, Lasbela, Pasni, Jacobabad, Drosh, and Lahore.

Generally, the above-discussed sites are very important from a geographical point of view over the country. A spatial map of the country and three sites (namely Hyderabad, Khuzdar, and Chitral) with their neighbouring areas in concern division are presented in Figure 3. In addition, Figure 3b covered the area of the Hyderabad division of the Sindh province and the red area indicates the Hyderabad city which was quite affected during the latest monsoon seasons. The left bottom map described the Kalat division of the Balouchistan province. The red colour indicates the Khuzdar district of the Kalat division. This division also received a lot of damages during monsoon seasons. Figure 3d presented the Malakand division of KPK province, and the red colour indicates the Chitral district. During the 2020 monsoon, the Malakand division got extreme rain events and faced land sliding problems, and flashed floods. Furthermore, the daily life of the people in these areas of the country is mostly affected by heavy rainfalls during every monsoon. The modelling presented in this paper could be very helpful in policymaking and the country's development.

3.3 | Influence of priors on return levels

To inspect the influence of the NIPs and IPs on RLs, the posterior density plots for $(0<p<1)$ were constructed by considering the vector of observations. Consequently, these are obtained from the marginal posterior distributions of GEV parameters. For instance, the posterior densities for 10, 25, 50, 100, and 500 years were obtained against the different posterior distributions. The RLs are also sensitive in the context of the choice of values p (e.g., $p=0.1, 0.04, 0.02, 0.01$, and 0.005).

The posterior densities plots of site Khuzdar for 10-, 25-, 50-, 100-, and 500 years RLs based on NIPs and IPs are

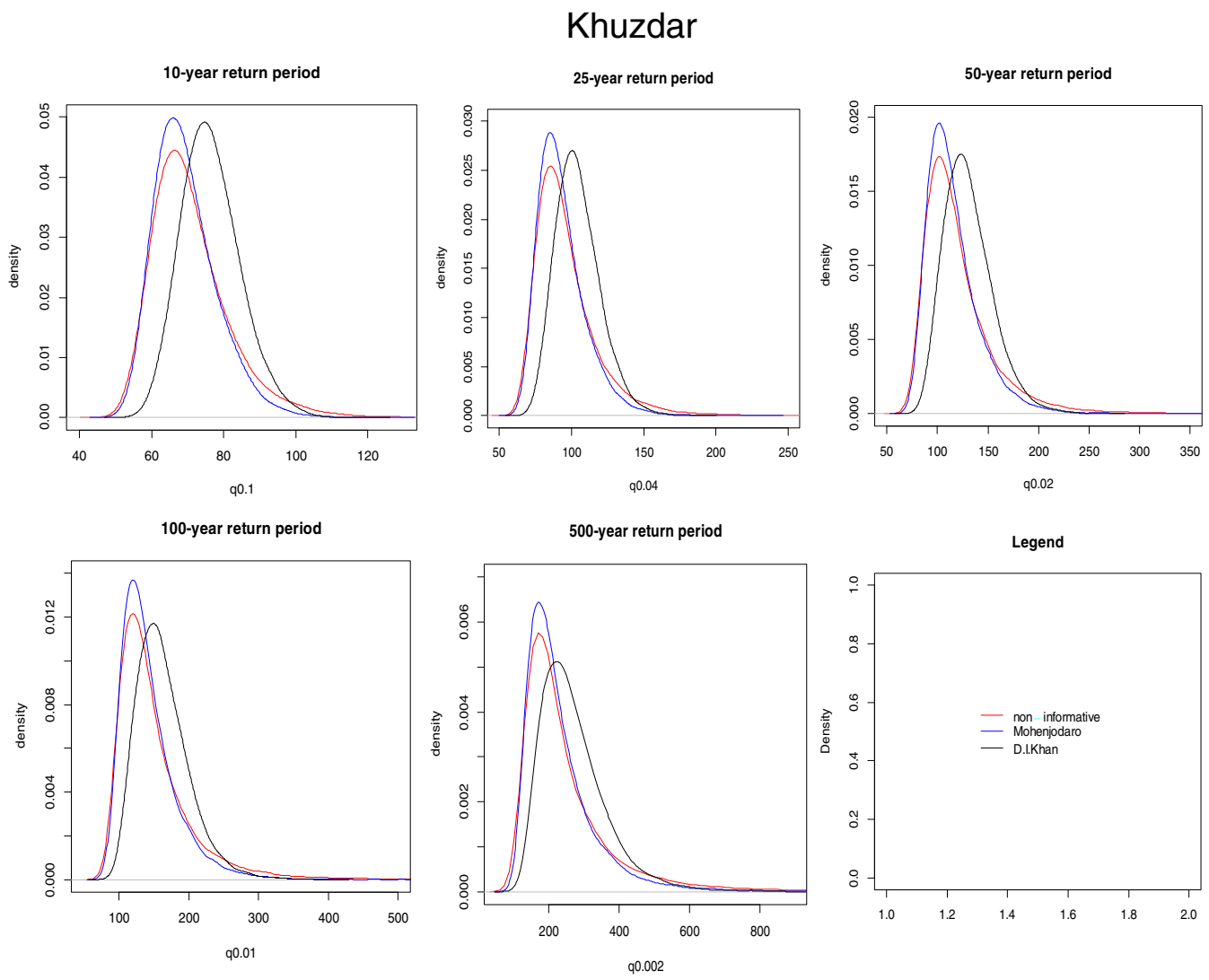

FIGURE 4 Posterior densities of 10-, 25-, 50-, 100-, and 500-year return levels via noninformative and informative priors [Colour figure can be viewed at wileyonlinelibrary.com]

presented in Figure 4. Also, the posterior densities plots for the RLs of the remaining sites namely Rohri, Nawabshah, Hyderabad, Chhor, Lasbela, Pasni, Jacobabad, Drosh, Chitral, and Lahore are presented in Figure S2 (see for example, supporting information file). From the plots given in Figure 4, it can be realized that the IPs also affected the RLs distribution. Furthermore, the PDs of IPs have appeared with high peaks as compared with PDs of NIPs. Also, to some extent, the distributions of all RLs are skewed to the right side. On the other hand, the posterior densities for higher RLs of stations (i.e., Rohri, Hyderabad, and Jacobabad) are not interpretable for the case of NIPs. So far, when comparing MLE with the LM method in a frequentist framework, we acquired different mean RLs, which could be due to the heavy tail of the GEV distribution or skewness detected in the data. The mean RLs obtained for Jacobaabad weather station through MLE are very high as compared to the LM method. Thus, the RLs estimates could be improved for all sites by choosing suitable summary measures for the inference. Generally, the posterior medians could be used in place of means in the Bayesian setting. Additionally, the skewed densities of return levels indicate the uncertainty inside the model for developing reasonable upper bounds of the return levels as compared to lower limits for higher return periods (Coles and Tawn, 2005) and (Ahmad *et al.*, 2019). Therefore, posterior medians were obtained as the best choice than the posterior mean for 10, 25, 50, 100, and 500 years RLs of ADMRS using NIPs and IPs. The RLs via NIPs are provided in Table 6 while the RLs based on the IPs are given in Table 7.

The posterior medians of RLs obtained via IPs are close to RLs based on MLEs except for the Jacobabad weather station. From the results of Table 7, it can be seen that the posterior RLs are quite reasonable also for

TABLE 6 Estimated return levels corresponding to different return periods of ADMRS for 11 weather stations across the country by using noninformative priors

	Noninformative prior				
Study locations	**10**	**25**	**50**	**100**	**500**
Rohri	107.87	167.21	226.00	300.74	563.38
Nawabshah	94.00	134.53	170.40	211.77	334.81
Hyderabad	101.06	136.44	165.57	197.14	281.54
Chhor	135.08	175.63	207.83	241.70	327.83
Khuzdar	70.23	94.03	115.76	141.53	221.74
Lasbela	91.48	131.50	169.26	215.09	364.60
Pasni	73.35	101.47	125.65	152.85	230.37
Jacobabad	145.86	278.57	443.95	700.33	1980.87
Drosh	79.38	98.00	113.289	129.823	173.91
Chitral	86.08	122.90	160.57	209.80	391.32
Lahore	144.79	181.67	211.21	242.50	323.07

TABLE 7 Estimated return levels corresponding to different return periods of ADMRS for 11 weather stations over the country by using Bayesian paradigm under informative priors

	Informative prior (Mohenjo-daro)					Informative prior (D.I. Khan)				
Study locations	**10**	**25**	**50**	**100**	**500**	**10**	**25**	**50**	**100**	**500**
Rohri	90.43	131.03	167.93	211.51	346.32	93.19	129.49	160.52	195.26	293.44
Nawabshah	85.17	118.27	146.36	77.64	265.14	89.78	122.19	148.82	177.66	254.69
Hyderabad	92.26	121.37	144.33	168.34	228.82	94.90	123.25	145.08	167.45	221.93
Chhor	120.89	153.82	179.15	205.06	268.02	116.87	147.16	170.06	193.15	247.93
Khuzdar	68.97	91.74	112.37	136.65	211.36	76.12	103.32	128.41	158.41	253.22
Lasbela	83.70	117.08	147.51	183.52	295.32	87.64	121.26	151.33	186.34	292.06
Pasni	70.71	97.01	119.36	144.27	214.09	79.55	109.88	135.82	164.91	247.24
Jacobabad	99.73	162.94	229.76	319.56	667.32	97.78	146.31	192.21	248.25	431.93
Drosh	77.88	95.75	110.35	126.09	167.82	82.90	103.57	120.80	139.68	191.19
Chitral	79.97	110.00	139.47	176.60	305.18	83.44	113.43	142.02	177.20	293.94
Lahore	131.76	161.39	184.32	207.89	265.67	126.54	153.35	173.73	194.37	243.72

Jacobabad station. On the other hand, RLs calculated through L-moments are smaller as compared to other methods. This might be happened due to the negativity of the shape parameter. The best estimation method for future modelling could be decided by assessment measures. For this purpose, we used two assessment measures. A useful discussion for those procedures was given in the next section.

3.4 | Model selection through assessment measures

The methods used for the analysis and/or for the modelling of extreme rainfall data were compared in this section. The comparison task was carried out by using the assessment measures namely RRMSE and RAE given in Equations (24) and (25), respectively. Also, Table 8 was created by using the "Fgmutils" R package.

From the findings of Table 8, it can be observed the Bayesian MCMC approach have smaller values for both RRMSE and RAS when compared with those of the frequentist methods. Besides, the Bayesian paradigm with the support of IPs also had smaller values for RRMSE and RAS than Bayesian analysis based on noninformative priors. Furthermore, the results of RRMSE and RAS for the Bayesian MCMC method linked with the IPs generated from the data of D.I. Khan station were smaller than the Bayesian MCMC method based on IPs built from the data of Mohenjo-daro station. Honestly speaking, overall results based on the Bayesian method when compared with classical approaches discovered the

TABLE 8 Assessment measures results for the frequentist methods and Bayesian method (both for noninformative and informative priors)

	Frequentist methods				Bayesian method					
	MLE		L-moments		Noninformative		Mohenjo-daro		D.I. Khan	
Study locations	RRMSE	RAE	RRMSE	RAE	RRMSE	RAE	RRMSE	RAE	RRMSE	RAE
Rohri	0.245	1.576	0.285	1.581	0.230	1.575	0.221	1.358	0.214	1.534
Nawabshah	0.254	1.872	0.293	1.961	0.249	1.824	0.246	1.784	0.231	1.813
Hyderabad	0.239	1.739	0.263	1.930	0.225	1.706	0.210	1.697	0.205	1.690
Chhor	0.206	2.044	0.224	2.262	0.202	2.010	0.198	1.672	0.191	1.622
Khuzdar	0.177	1.674	0.182	1.836	0.174	1.103	0.176	1.629	0.173	1.611
Lasbela	0.220	1.618	0.239	1.946	0.221	1.608	0.204	1.387	0.227	1.401
Pasni	0.217	1.776	0.24	2.077	0.215	1.743	0.203	1.661	0.200	1.594
Jacobabad	0.271	1.239	0.257	1.212	0.276	1.282	0.247	1.208	0.198	1.190
Drosh	0.114	1.945	0.117	2.023	0.112	1.923	0.111	1.901	0.109	1.894
Chitral	0.135	1.396	0.151	1.738	0.134	1.347	0.133	1.302	0.130	1.116
Lahore	0.145	1.997	0.154	2.218	0.142	1.991	0.131	1.989	0.114	1.985

reputation of Bayesian inference for the extreme rainfall data. This approach deals with the uncertainties linked with excesses of weather variables efficiently. Essentially, the IPs for the Bayesian method constructed from surrounding weather stations increase the accuracy of the parameter estimates than the frequentist approaches. This study partially supports the Bayesian paradigm results of (Ahmad *et al*., 2019). Even though it is arguable to say which of the methods offer accurate estimates, it can be contended that the supplement of uncertainties via IPs in the Bayesian framework significantly enhanced the findings of the estimates for ADMRS at different weather stations over the country. Conversely, the precision of the estimates could be more improved with a suitable choice of weather stations used for the elicitation of IPs.

4 | CONCLUSION

In this paper, our attention was to develop the modelling of extreme rainfall patterns all over the country by applying frequentist and Bayesian methods. Frankly speaking, we were unable to analyse the data of all stations across the country. Therefore, the ADMRS of suitable weather stations of the country were chosen by keeping in mind that the environmental conditions are homogenous across the stations in the provinces of the country. The data used for this study from different weather stations around the country were shown to follow the family of extreme value distributions (i.e., GEV distribution). In a frequentist setting, the parameters of GEV distribution were estimated through MLE and LM methods. Furthermore, RLs for (10, 25, 50, 100, and 500 years) were also calculated for MLE and LM methods. The RLs based on LM were showed consistency while examined by the birds-eye view. But, the results obtained from both methods provided evidence that there would be extreme rain events across the country in the future.

Modelling the behaviour of such extremes events within the Bayesian paradigm at different weather stations throughout the country offers more beauty to this paper. Bayesian MCMC is respected when climatic indications are unusual, and also the behaviour of extreme rainfall is similar over the region from which the data were acquired. Consequently, in these circumstances, the authors have preferred a Bayesian paradigm over the frequentist methods. This needs a genuine construction of IPs, thus it provides great estimation accuracy. Similar to the frequentist setting, in the Bayesian framework, the parameter of GEV distribution and RLs for (10, 25, 50, 100, and 500 years) were estimated via NIPs and IPs formulated from two suitable weather stations over the country. Moreover, the parameter estimates and RLs for GEV were sensitive to those sites used for the elicitation of IPs. Consequently, the present study also supports proper choices of the neighbouring stations, since the devising of IPs is significant as the estimates and the accuracies are profound to these priors.

Additionally, assessment measures were used to adopt a superior method for modelling ADMRS among frequentist and Bayesian approaches. The smaller values of assessment measures proved the precision of the Bayesian MCMC method associated with IPs. Thus, our

current methodology could be implemented easily to other ungagged sites of the country. Also, the information from the neighbouring countries (for instance, India, Afghanistan, and Iran) could be utilized as prior knowledge. On the other hand, the findings of the proposed method could be very helpful for policymakers and hydrologists. Hence, engineers can take help from this study in designing dams, bridges, culverts, and flood control devices in Pakistan. The study could be improved more by considering nonstationary rainfall series and by inspecting a linear time trend in the location parameter of GEV, and also with the exercise of spatial modelling.

ACKNOWLEDGEMENTS

The authors express their gratitude towards two anonymous referees and editor and associate editor for many useful comments that have helped to improve the earlier version of the manuscript. Also, the authors would like to thank the Department of Statistical Sciences, Università degli Studi di Padova, Italy for giving administrative and technical support. The authors are also grateful to Pakistan Meteorological Department, Karachi for providing the rainfall data. Open Access Funding provided by Universita degli Studi di Padova within the CRUI-CARE Agreement. [Correction added on 18 May 2022, after first online publication: CRUI-CARE funding statement has been added.]

AUTHOR CONTRIBUTIONS

Touqeer Ahmad: Data curation; investigation; methodology; software; writing-original draft. **ishfaq ahmad:** Conceptualization; project administration; supervision; validation. **Irshad Arshad:** Formal analysis; validation; writing-review & editing. **Nicolas Bianco:** Data curation; methodology; visualization; writing-original draft.

ORCID

Touqeer Ahmad https://orcid.org/0000-0002-2046-6268
Ishfaq Ahmad https://orcid.org/0000-0002-1124-7485

REFERENCES

Ahmad, I., Shah, S.F., Mahmood, I. and Ahmad, Z. (2013) Modeling of monsoon rainfall in Pakistan based on kappa distribution. *Science International (Lahore)*, 25(2), 333–336.

Ahmad, I., Abbas, A., Saghir, A. and 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), 925–993. https://doi.org/10.15244/pjoes/61715.

Ahmad, I., Ahmad, T. and Almanjahie, I.M. (2019) Modeling of extreme rainfall in Punjab, Pakistan using Bayesian and frequentist approach. *Applied Ecology and Environmental Research*, 17(6), 13729–13748. https://doi.org/10.15666/aeer/1706_1372913748.

Bücher, A., Volgushev, S. and Zou, N. (2019) On second-order conditions in the multivariate block maxima and peak over threshold method. *Journal of Multivariate Analysis*, 173, 604–619. https://doi.org/10.1016/j.jmva.2019.04.011.

Aljazeera News (2020) Flooded streets, submerged cars as rain causes havoc in Karachi. https://www.aljazeera.com/news/2020/08/flooded-streets-submerged-cars-rain-havoc-karachi-200825120348316.html.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2004) *Statistics of Extremes: Theory and Applications*. Wiley, Chichester.

Coles, S. and Tawn, J.A. (1996) A Bayesian analysis of extreme rainfall data. *Applied Statistics*, 6(10), 463–478. https://doi.org/10.2307/2986068.

Coles, S.G. and Dixon, M.J. (1999) Likelihood-based inference for extreme value models. *Extremes*, 2(1), 5–23. https://doi.org/10.1023/A:1009905222644.

Coles, S.G. and Powell, E.A. (1996) Bayesian methods in extreme value modeling: a review and new developments. *International Statistical Review*, 64(1), 119–136. https://doi.org/10.2307/1403426.

Coles, S. (2001) *An introduction to statistical modeling of extreme values*. London: Springer.

Coles, S., Pericchi, L.R. and Sisson, S. (2003) A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 273(1–4), 35–50. https://doi.org/10.1016/S0022-1694(02)00353-0.

Coles, S. and Tawn, J. (2005) Bayesian modeling 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. https://doi.org/10.1098/rsta.2005.1574.

Chu, P.S. and Zhao, X. (2011) Bayesian analysis for extreme climatic events: a review. *Atmospheric Research*, 102(3), 243–262. https://doi.org/10.1016/j.atmosres.2011.07.001.

Davison, A.C. and Smith, R.L. (1990) Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3), 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x.

Diriba, T.A., Debusho, L.K., Botai, J. and Hassen, A. (2017) Bayesian modeling of extreme wind speed at Cape Town, South Africa. *Environmental and Ecological Statistics*, 24(2), 243–267. https://doi.org/10.1007/s10651-017-0369-z.

Diriba, T.A. and Debusho, L.K. (2020) Modeling dependency effect to extreme value distributions with application to extreme wind speed at Port Elizabeth, South Africa: a frequentist and Bayesian approaches. *Computational Statistics*, 35, 1–31. https://doi.org/10.1007/s00180-019-00947-2.

Elamir, E.A. and Seheult, A.H. (2003) Trimmed L-moments. *Computational Statistics & Data Analysis*, 43(3), 299–314. https://doi.org/10.1016/S0167-9473(02)00250-5.

El Adlouni, S. and Ouarda, T.B.M.J. (2010) Frequency analysis of extreme rainfall events. In: Testik, F.Y., & Gebremichael, M. (Eds.). *Rainfall: State of the Science*, *Geophys. Monogr. Ser.* (Vol. 191, pp. 171–188). Washington, D.C.: AGU.

Eastoe, E.F. and Tawn, J.A. (2012) Modeling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1), 43–55. https://doi.org/10.1093/biomet/asr078.

Eli, A., Shaffie, M. and 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.

Ferreira, A. and De Haan, L. (2015) On the block maxima method in extreme value theory: PWM estimators. *The Annals of Statistics*, 43(1), 276–298.

Fadhilah, Y., Md, Z., Nguyen, V., Suhaila, S. and Zulkifli, Y. (2007) Fitting the best-fit distribution for the hourly rainfall amount in the Wilayah Persekutuan. *Jurnal Teknologi*, 46(1), 49–58.

Fawcett, L. and Walshaw, D. (2008) Modeling environmental extremes. In: *Short Course for the 19th Annual Conference of The International Environmetrics Society*. The University of British Columbia kanagan, Kelowna, Canada.

First, P.J. (2019) Global warming of 1.5 C. An IPCC special report on the impacts of global warming of 1.5 C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change. *Sustainable Development, and Efforts to Eradicate Poverty*. https://www.ipcc.ch/sr15/.

Hosking, J.R. (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1), 105–124.

Hosking, J.R.M. and Wallis, J.R. (2005) *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, New York.

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. https://doi.org/10.1007/s12205-016-0946-y.

Khudri, M.M. and Sadia, F. (2013) Determination of the best fit probability distribution for annual extreme precipitation in Bangladesh. *European Journal of Scientific Research*, 103(3), 391–404.

Khan, S., Hussain, Z. and Ahmad, I. (2021) Effects of L-moments, maximum likelihood and maximum product of spacing estimation methods in using Pearson Type-3 distribution for modeling extreme values. *Water Resources Management*, 35, 1415–1431. https://doi.org/10.1007/s11269-021-02767-w.

Lenderink, G. and Fowler, H.J. (2017) Understanding rainfall extremes. *Nature Climate Change*, 7(6), 391–393. https://doi.org/10.1038/nclimate3305.

Madsen, H., Pearson, C.P. and 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. https://doi.org/10.1029/96WR03849.

Marani, M. and Ignaccolo, M. (2015) A metastatistical approach to rainfall extremes. *Advances in Water Resources*, 79, 121–126. https://doi.org/10.1016/j.advwatres.2015.03.001.

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. https://doi.org/10.1007/s11269-008-9378-x.

Naghettini, M. (Ed.). (2017) *Fundamentals of Statistical Hydrology*. Cham: Springer.

Olofintoye, O.O., Sule, B.F. and Salami, A.W. (2009) Best-fit probability distribution model for peak daily rainfall of selected cities in Nigeria. *New York Science Journal*, 2(3), 1–12.

Rao, C.R. (1973) *Linear Statistical Inference and Its Applications*, Vol. 2. New York: Wiley Sons.

Rahman, A.S., Rahman, A., Zaman, M.A., Haddad, K., Ahsan, A. and Imteaz, M. (2013) A study on selection of probability distributions for at-site flood frequency analysis in Australia. *Natural Hazards*, 69(3), 1803–1813. https://doi.org/10.1007/s11069-013-0775-y.

Rivas, D., Caleyo, F., Valor, A. and Hallen, J.M. (2008) Extreme value analysis applied to pitting corrosion experiments in low carbon steel: comparison of block maxima and peak over threshold approaches. *Corrosion Science*, 50(11), 3193–3204. https://doi.org/10.1016/j.corsci.2008.08.002.

Sathar, Z. and Khan, K. (2019). (eds.). *Climate, population, and vulnerability in Pakistan: Exploring evidence of linkages for adaptation*. Islamabad: Population Council. https://knowledgecommons.popcouncil.org/departments_sbsr-pgy/733/.

Smith, R.L. (1985) Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, 72(1), 67–90. https://doi.org/10.1093/biomet/72.1.67.

Smith, E. (2005). *Bayesian Modelling of Extreme Rainfall Data*. Doctoral dissertation, University of Newcastle upon Tyne (Unpublished). Newcastle upon Tyne:

Suhaila, J., Ching-Yee, K., Fadhilah, Y. and Hui-Mean, F. (2011) Introducing the mixed distribution in fitting rainfall data. *Open Journal of Modern Hydrology*, 1(2), 11–22. https://doi.org/10.4236/ojmh.2011.12002.

Zhu, J., Forsee, W., Schumer, R. and Gautam, M. (2013) Future projections and uncertainty assessment of extreme rainfall intensity in the United States from an ensemble of climate models. *Climatic Change*, 118(2), 469–485. https://doi.org/10.1007/s10584-012-0639-6.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ahmad, T., Ahmad, I., Arshad, I. A., & Bianco, N. (2022). A comprehensive study on the Bayesian modelling of extreme rainfall: A case study from Pakistan. *International Journal of Climatology*, *42*(1), 208–224. https://doi.org/10.1002/joc.7240