Energy Reports 9 (2023) 2980–2992

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

An efficient Bayesian modelling of extreme winds in the favour of energy generation in Pakistan

Touqeer Ahmad [a,*], Ishfaq Ahmad [b], Irshad Ahmad Arshad [c], Ibrahim Mufrah Almanjahie [d,e]

[a] Department of Statistical Sciences, University of Padova, Via Cesare Battisti, 241, 35121, Padova, PD, Italy
[b] Department of Mathematics and Statistics, Faculty of Basic and Applied Sciences, International Islamic University, 44000 Islamabad, Pakistan
[c] Department Statistics, Faculty of Sciences, Allama Iqbal Open University, 44000 Islamabad, Pakistan
[d] Department of Mathematics, King Khalid University, Abha 62529, Kingdom of Saudi Arabia
[e] Statistical Research and Study Support Unit, King Khalid University, 62529, Abha, Kingdom of Saudi Arabia

ARTICLE INFO

Article history:
Received 18 January 2022
Received in revised form 17 March 2022
Accepted 19 January 2023
Available online 10 February 2023

Keywords:
Bayesian
Extreme winds
EVDs
Generalized Pareto distribution
Informative priors

ABSTRACT

Daily and annual maximum wind speed quantiles can be estimated using extreme value theory for any metrological site of interest. These estimates are of vast importance for modelling and predicting maximum wind speed. This paper develops an efficient modelling paradigm of extreme winds by analysing daily and annual maximum wind speed data via frequentist and Bayesian methodologies. For this purpose, the generalized extreme value (GEV) model is used for yearly maxima, and the generalized Pareto distribution (GPD) is used for daily exceedance over a high threshold. In frequentist inference, the parameters of both models are estimated using the maximum likelihood and linear moments method. In contrast, the Bayesian Markov Chain Monte Carlo procedure with the Metropolis–Hasting algorithm is used. In addition, the informative priors for both models are constructed empirically using historical records of wind speed data from five other weather stations of Pakistan and one belonging to India. The results show that the Bayesian modelling provides apparent benefits in terms of improved accuracy in the estimation of the parameters as well as return levels of both distributions. Furthermore, the Bayesian analysis expresses that posterior inference might be affected by the choice of priors used to formulate the informative priors. Overall, based on assessment measures, the GPD fitted through Bayesian informative priors provides an efficient estimation strategy in terms of precision than other frameworks when uncertainty in parameters and return levels are taken into account. Our methodology can be implemented easily to other regions by considering the prior information from the border area stations of other countries (e.g., China, Afghanistan, India, and Iran). Moreover, the return level estimates of the GPD based on informative Bayesian priors are very beneficial in policymaking and wind energy generation engineering for the Thatta region of the country.

1. Introduction

In a developing world, energy plays a crucial role in the progress of any country. Energy is considered a primary ingredient of the industrial economy (Saulat et al., 2020). People's daily life activities are genuinely linked with energy supply in the modern era, such as cooking, heating, lighting, health, production, storage, and education. Therefore, a reliable, continuous, and cheap energy supply is a prerequisite for reducing poverty, boosting investment, and enhancing economic growth. Currently, energy systems have been revamped. It is assumed that poverty will not be diminished in emerging states without the larger utilization of recent modes of energy. Due to the rising world population, energy demand and consumption are increasing day by day, but the resources are decreasing (e.g., water storage in dams, expensive fuel, chronic natural gas etc.). The growing trend of energy consumption and demand indicates that energy will be the most crucial issue of the world (Rafique and Rehman, 2017; Hulio et al., 2017). Unfortunately, Pakistan is facing severe energy crises for the last few decades. Even though extensive access

Abbreviations: GEV, Generalized extreme value distribution; GPD, Generalized Pareto distribution; EVD, Extreme value distribution; RL, Return level; MCMC, Markov Chain Monte Carlo; NIP, Noninformative prior; IP, Informative prior; POT, Peak over threshold; MLE, Maximum likelihood method; LMM, Linear moments method; SD, standard deviation

* Corresponding author.
E-mail address: touqeer.ahmad@studenti.unipd.it (T. Ahmad).

https://doi.org/10.1016/j.egyr.2023.01.093

to electricity (99% of the population had access to electricity in 2016, compared to 59% of the population in 1990), the country suffers enormous blackouts (load shedding of 6–8 h a day for households and 1–2 h a day for the industry). The energy generation resources are reduced day by day due to the exponential growth in demand. These resources are usually divided into two major mechanisms: (i) renewable energy resources (i.e., energy resources that can be revived in a short period of time are known as renewable energy resources such as wind energy and solar energy) and (ii) non-renewable energy resources (i.e. those resources that cannot be restored are known as non-renewable energy resources like coal and oil) (Raza et al., 2015; Saulat et al., 2020). The most popular and richly obtained renewable energy forms are solar, wind, hydrothermal energy. Pakistan has been blessed with many renewable energy resources such as solar, wind and biomass energy; however, Pakistan's energy resources have been used inefficiently for decades. As a result, the nation confronts a severe energy crisis. According to a survey by the World Bank, 66.7% of the businesses in Pakistan mention electricity shortages as a more significant barrier to business than other factors. Under the umbrella of these issues, there are urgent actions needed to modernize the country's energy sector. Nowadays, the country is moving its attention to the wind energy sector. Many areas of the Sindh and the Baluchistan provinces receive extreme windstorms for the whole year. An appropriate statistical modelling of extreme winds makes it possible to install wind turbines in that areas. For instance, in the case of extreme winds, statistical models of extremes can be helpful to measure the involved uncertainty and risk. The extreme wind is somehow connected with extreme environmental events.

Analysis of extreme environmental events of many natural disasters is a hot topic throughout the different areas of scientific research. Derivation of extremal features of observed phenomena and quantification of the stochastic behaviour of such events is one of the primary concerns in the statistical modelling of extremes (Coles, 2001). However, the fundamental issue is a lack of data, essentially modelling with few observations is not advantageous. Most statistical approaches focus on the central part of the distributions, whereas the tails are frequently disregarded. Procedures used to analyse the observations exist at the tail of the distribution partaking excellent statistical properties of a similar or dissimilar process for one or more variables. For example, such analysis could aid in estimating the frequency and magnitude of extreme events. This enables preventative steps to be implemented in order to avoid catastrophic events, plan for their influence, and lessen their consequences. Extreme value theory (EVT) enables us to measure the stochastic behaviour of an event found in extremes (upper or lower tails). Generally, extreme data are scarce by their nature. The development of inferential methods that exploit the use of available data has been a dominating research issue for the last few decades.

In particular, the EVT modelling framework has been split into two categories: block maxima and peak over a threshold (POT). Block maxim procedure models the maximum observations (yearly, monthly, weekly etc.) of data collected from the large sample using generalized extreme value (GEV) distribution (Coles, 2001; Ferreira and De Haan, 2015). On the other side, the second approach models the observations (hourly, daily etc.) that exceed a high threshold via generalized Pareto distribution (GPD) (Davison and Smith, 1990; Coles, 2001; Beirlant et al., 2004; Ferreira and De Haan, 2015; Coles and Powell, 1996). Numerous studies favouring block maxima and POT have been done in different parts of the world. Coles (2001) has extensively discussed applications of both schemes in extreme paradigm applying the maximum likelihood inference method. Ferreira and De Haan (2015) have done a wide survey of literature to compare block maxima and POT. They use the probability-weighted moments procedure to estimate the parameters of both schemes. Several works, including (Martins and Stedinger, 2001; Madsen et al., 1997; Wang, 1991) have examined the relative merits of POT and block maxima. Based on these studies, Ferreira and De Haan (2015) also identified two significant properties of block maxima and POT, with mixed views on performance. POT estimates are more efficient than block maxima in many studies, provided the number of exceedances is greater than the number of blocks on average. de Oliveira et al. (2011) found that the GPD fitted more efficiently than GEV during modelling extreme wind speed in grid points over the South Atlantic region from north of Argentina to southeast Brazil. Furthermore, extreme value distributions (EVDs) were used to describe the extreme behaviour of the wind speed variable at Port Elizabeth station in South Africa (Diriba and Debusho, 2020). Ahmad et al. (2019, 2021) used GEV distribution to model extreme precipitation data at different weather stations in Pakistan.

In addition, several studies have been conducted on wind data at the national level. Kamal and Jafri (1997) used an autoregressive moving average time series model for stochastic simulation and modelling hourly averaged wind speed in Quetta, Pakistan. Fawad et al. (2018) developed the regional frequency modelling of wind speed data using generalized normal, generalized logistic, Pearson Type-3, GPD, Weibull, log Pearson Type-3 distributions. They used linear moments method for parameter and quantiles estimations. Khan et al. (2019) developed the modelling for predictive assessment of wind energy at 10 m height as a power generation source at seven locations of Pakistan. They used different parametric probability distributions with classical inference to complete this task. Shahzad et al. (2020) studied new Hybrid autoregressive autoregressive and neural network models in the context of modelling and forecasting wind speed over the three regions of Pakistan. Haq et al. (2020) investigated the fitting of five parametric probability distributions on wind speed data using classical inference procedure. Sumair et al. (2021) applied three probability distributions (i.e., Weibull, Rayleigh, and lognormal) to develop the wind modelling at six sites along the country's coastal belt. They practiced only classical techniques such as maximum likelihood estimation method. The present study develops the Bayesian framework compared with classical for modelling and forecasting extreme winds. In general, no study was found in the literature relevant to our proposed work over the region.

The main motive behind this work is to find the best model with an efficient estimation procedure. Extreme value models have a long history in this area; however, traditional methods are inefficient in maximum data utilization. Given the importance of evidence in extreme value modelling, it is only reasonable to investigate incorporating additional sources of information into the analysis. For instance, the time period for which data was collected may not be fully representative. There could be historical evidence, but it will not be in the shape of data and its behaviour substantially extremes. Therefore, there are various reasons to believe that an expert who understands the physical processes involved may have information on extreme behaviour independent of the available data. That is the reason the Bayesian Markov Chain Monte Carlo (MCMC) inferential procedure is a natural choice for conducting extreme value analysis of extreme wind data over the region. So, the historical records of different weather stations are considered to elicit prior information.

The rest of the paper is organized as follows. Section 2 is dealt with information about the study area and some exploratory analysis of the data. Comprehensive details regarding models and their inference are given in Section 3. For instance, classical and Bayesian MCMC techniques for block maxima and POT have been briefly discussed. The results and fruitful discussions exist in Section 4. Lastly, the conclusions and some recommendations for future work are described in Section 5.

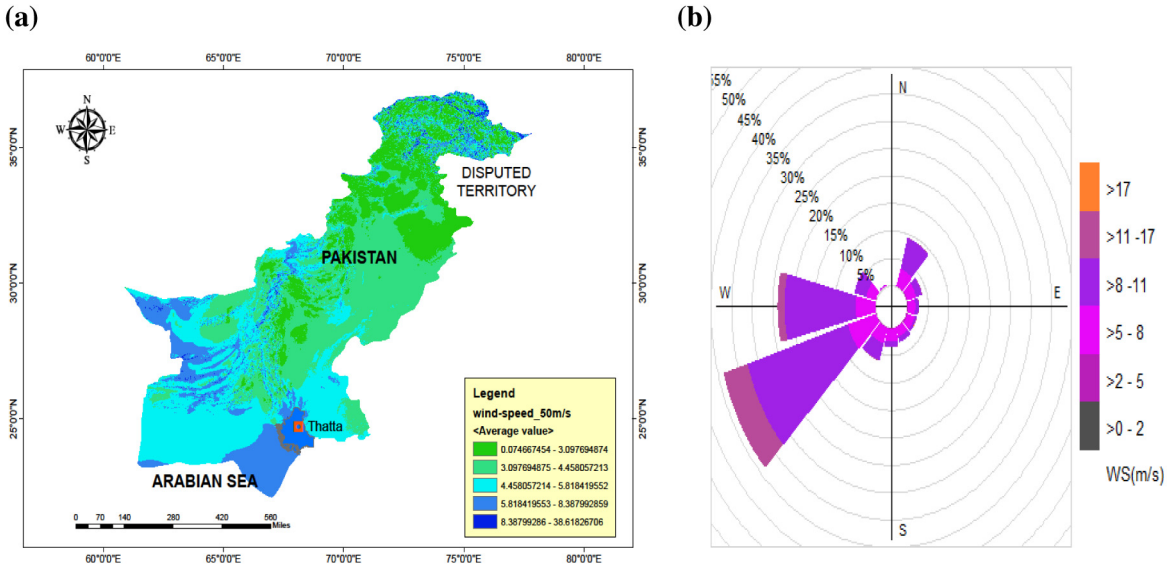

Fig. 1. (a) Geographical location of the Thatta station and blue colour indicates maximum wind speed (see legends) (b) presents the direction of wind speed with count frequency at Thatta station. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2. Leading areas information and exploratory analysis

The intended research study is the first submission in respect of frequency analysis of wind speed using both frequentist and Bayesian paradigms in Pakistan. The data considered in this study consist of an aggregate of daily maximum wind speed at 50 meters per second at the Thatta district (Latitude 24.5457°N and longitude 67.9524°E) of the province Sindh, Pakistan. The data have been taken from the website https://power.larc.nasa.gov/data-access-viewer/. The data used correspond to 34-years, from 1981 to 2014 series of daily maximum wind speed records. The windy areas and geographical position with the wind direction on the observed site are illustrated in Fig. 1. Thus, Fig. 1(a) pointed out those areas of the country that gained maximum wind at 50 m height. The purple colour rectangle in the map indicates the observable stations Thatta. Fig. 1(b) shows the direction of the wind speed in the Thatta region. It is clear in Fig. 1(b) that the wind speed direction at the selected station is to the southwest.

For the construction of informative priors for the parameters of EVDs in the Bayesian setting, the wind characteristics of six sites (namely, Badin, Hyderabad, Jamshoro, Nawabshah, Naushahro Feroze and Lakhpat) at various distances (93 km, 85 km, 85.37 km, 173 km, 233.67 km and 135 km). Thus, due to the very short distance from the Thatta station, the Lakhpat station is selected from the Indian region Gujrat. For the elicitation of priors, all stations are from the same province except the Lakhpat station, so it is sufficient to justify the precision in the construction of informative priors that the primary criterion is geographical closeness or homogeneous environmental characteristics.

The summary statistics for the Thatta station are given in Table 1. From the findings, it can be noticed that the distribution of the series is positively skewed with a high peak. This suggested the use of EVDs for analysis. For a complete analysis, both annual and daily data were used. The best suitable distribution of the wind speed is performed using extreme value theory by applying to GEV and GP distributions. The fitting and estimation of the parameters of these models through frequentist and Bayesian settings are presented next.

Table 1
Summary statistics of daily maximum wind speed at Thatta station.

Station	Mean	SD	Min	Max	Skewness	Kurtosis	CV
Thatta	8.22	1.73	2.68	18.46	0.50	4.10	0.21

3. Methodology

3.1. Modelling of block maxima

In extreme value theory, GEV and GP distributions are widely used for the modelling and characterizing of extremes. To model variables with extreme nature using GEV, a data of N independent values $v_1, v_2, \ldots, v_N$ are first blocked into h blocks of size n, with n quite large, and hence $N = hn$. Generally, a month, season, or year could be considered a block size for the data of wind speed. For example, if n is the number of values in a year, then M_n indicates the maximum observation corresponds to a year (Coles, 2001). However, the maxima or extreme observation ($M_j, j = 1, \ldots, h$), is taken from each block. This generates the data of h annual maxima series so-called block maxima to which the family of GEV models can be fitted. The annual maxima $v_1, v_2, \ldots, v_n$ are independent and identically distributed (i.i.d) with the distribution function of $H(v)$. Let $M_n = \max(v_1, v_2, \ldots, v_n)$, $n \in \mathrm{N}$ and if there are sequences of normalizing constants $\{b_n > 0\}$ and $c_n \in \Re$ such that

$$pr\left\{\frac{(M_n - c_n)}{b_n} \leq v\right\} \to H^n(b_n v + c_n) \to F(v) \tag{1}$$

as $n \to \infty$, where F is a non-degenerate distribution function, the distribution function H is called to be in the domain of attraction of the extreme value distribution $F\{i.e.\ H \in F(v)\}$. Also, the F belongs to GEV family

$$F(v, \alpha, \beta, \xi) = \begin{cases} \exp\left[-\left\{1 + \xi\left(\frac{v-\alpha}{\beta}\right)\right\}^{-\frac{1}{\xi}}\right], & \xi \neq 0 \\ \exp\left[-\exp\left(-\frac{v-\alpha}{\beta}\right)\right], & \xi = 0 \end{cases} \tag{2}$$

defined on $\{v : 1 + \xi(v - \alpha)/\beta\}$, where $-\infty < \alpha < \infty, \beta > 0$ and $-\infty < \xi < \infty$ are location, scale, and shape parameters

of GEV distribution. However, the shape parameter defines the behaviour of the upper tail of the distribution. The GEV distribution has three different limiting forms, i.e., Gumbel distribution, Freshet distribution and Weibull distribution. If $\xi \to 0$ the GEV distribution in (2) leads to the Gumbel distribution. For $\xi > 0$ and $\xi < 0$, the (2) correspond to Frechet and Weibull distribution families, respectively, (Coles, 2001; Beirlant et al., 2004).

3.2. Parameter estimation of GEV model through frequentist setting

Initially, the maximum likelihood estimation method (MLEM) and linear moments method (LMM) were employed to estimate the parameters of GEV distribution in the frequentist setting. In MLEM, we differentiate the function given in (2) for v_i, for example, when $\xi \neq 0$ the density of GEV is given

$$f(v_i,\alpha,\beta,\xi) = \frac{1}{\beta}\left[1+\xi\frac{(v_i-\alpha)}{\beta}\right]^{-\left(1+\frac{1}{\xi}\right)} \times \exp\left[-\left\{1+\xi\frac{(v_i-\alpha)}{\beta}\right\}^{-\frac{1}{\xi}}\right] \tag{3}$$

The Maximum Likelihood Estimates (MLEs) of the parameter vector (α,ξ), say $\hat{\alpha},\hat{\beta}$ and $\hat{\xi}$ are obtained by maximizing the log-likelihood with respect to unknown parameters. The log-likelihood is given in the following form

$$l(\alpha,\beta,\xi;v_1,v_2,\ldots,v_n) = -n\log\beta - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\log\left[1+\xi\frac{(v_i-\alpha)}{\beta}\right] - \sum_{i=1}^{n}\left\{1+\xi\frac{(v_i-\alpha)}{\beta}\right\}^{-\frac{1}{\xi}} \tag{4}$$

due to difficulty in the solution of expression (4), the maximization is obtained by quasi-Newton procedure with numerical iteration (Ahmad et al., 2019; Diriba et al., 2017; Diriba and Debusho, 2020).

For the estimation of linear moments via LMM, we apply the linear combinations of order statistics values (Hosking and Wallis, 2005). LLM offers simple and more efficient estimators of the characteristics of extremal data and the parameters of the distribution. Let a random sample $V_1, V_2, \ldots, V_r$ of size n, with cumulative distribution function $F(v)$ and quantile function $v(F)$. Suppose $V_{1:r} \leq V_{2:r} \leq \cdots \leq V_{r:r}$ be the order statistics of the sample. For the random variable V, the rth population linear moments explained by (Ahmad et al., 2019):

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

Generally, we require the first four linear moments for $r = 1,2,3,4$. Moreover, linear moments can also be considered as linear combinations of probability-weighted moments as in the form

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

A brief description of the first four linear moments and other quantities for the GEV model are given in (supplementary materials file **A.1**)

3.3. Peak-over-threshold and generalized Pareto distribution

In EVT, the method defined as Peak-Over-Threshold (POT) is commonly used. Let $v_1, v_2, \ldots$ is a sequence of independent and identically distributed (i.i.d) random variables having a continuous distribution function F which fulfils the condition in (1). Suppose v refer to an arbitrary term in the v_i sequence, it describes the stochastic behaviour of extreme events. Then, the conditional distribution function of v a suitably high threshold u is given in the following form

$$F_u(v) = \Pr(v-u|v>u) = \frac{F(v+u)-F(u)}{1-F(u)} \tag{7}$$

The expression (7) implies the probability that the observations of v exceeding the threshold u by $v-u$. According to Coles (2001), for a large sufficient threshold u, the distribution function of $(v-u)$, conditional on $v>u$, is approximately

$$K(v) = 1-\left(1+\frac{\xi v}{\widetilde{\beta}}\right)^{-1/\xi} \tag{8}$$

defined on $\{v: v>0 \text{ and } (1+\xi v/\widetilde{\beta})>0\}$, where ξ is a shape parameter and $\widetilde{\beta} = \beta+\xi(u-\alpha)$ is the modified scale parameter of the distribution. Note that the parameters α, β and ξ are location, scale and shape parameters as defined prior in (2), respectively. The distribution specified by expression (8) is called GPD. This means that, if the approximating distribution of block maxima is $F(v)$, then the approximate distribution corresponding to the observations that exceed the sufficient threshold lie within the generalized Pareto family with the shape ξ same as GEV and modified scale parameter $\widetilde{\beta} = \beta+\xi(u-\alpha)$ for any given threshold α. The distribution given in (16) is unbounded if $\xi = 0$ and it can be interpreted by taking the limit as $\xi \to 0$ in (16), which goes to

$$\lim_{\xi\to 0}K(v) = 1-\exp\left(-\frac{v}{\widetilde{\beta}}\right), \qquad v>0 \tag{9}$$

the exponential distribution with parameter $(1/\widetilde{\beta})$. Generally, the GPD is developed as a two-parameter in the following form

$$K(v,\beta,\xi) = \begin{cases} 1-\left\{1+\xi\dfrac{v}{\widetilde{\beta}}\right\}^{-1/\xi}, & \xi\neq 0 \\ 1-\exp\left(-\dfrac{v}{\widetilde{\beta}}\right), & \xi = 0 \end{cases} \tag{10}$$

where $v\in[0,\infty)$ for $\xi\leq 0$ and $v\in[0,\widetilde{\beta}/\xi)$ for $\xi>0$ (Jocković, 2012; Diriba and Debusho, 2020). As previously, the shape parameter establishes the form of distribution. For instance, the $\xi>0$ will produce heavy-tailed Pareto distribution, $\xi<0$ will create a Beta distribution with upper bound, and $\xi\to 0$ will provide an exponential distribution. Consequently, the asymptotic results depend on the selection of the threshold.

The selection of appropriate threshold is important; a very low threshold would lead to extreme bias while a very high threshold to larger variance (Ranjbar et al., 2020). For practical uses, an appropriate threshold u is selected using data-analytic instruments such as the mean residual life plot [for details see Coles (2001). This is established by the fact that if the excesses over the threshold u can be explained by a GPD with parameters $\xi<1$ and β, then it is easy to show that for any higher threshold $u_0\geq u$

$$E(v-u_0|v>u_0) = \frac{\beta_{u_0}+\xi(u_0-u)}{1-\xi} \tag{11}$$

the mean access function is linear in u_0 above u. In experimental applications, the mean residual life plot (which demonstrates the empirical mean excess opposite to the rising threshold values) is an advantageous path to decide the threshold and to prove the competence of the GPD as an approximation of the excess distribution.

3.4. Parameter estimation of GPD model through frequentist setting

Similar to the GEV model, both MLEM and LMM procedures were occupied to get the parameters estimates for GPD model in a frequentist setting. The density function of GPD is obtained by differentiating (10) with respect to v and has the following form

$$k(v,\widetilde{\beta},\xi)=\begin{cases}\frac{1}{\widetilde{\beta}}\left\{1+\xi\frac{v}{\widetilde{\beta}}\right\}^{-\frac{1}{\xi}-1}, & \xi\neq 0\\ \frac{1}{\widetilde{\beta}}\exp\left(-\frac{v}{\widetilde{\beta}}\right), & \xi=0\end{cases} \tag{12}$$

Suppose $v_1, v_2, \ldots, v_n$ are n exceedances over the threshold u, then the log-likelihood associated with (12) is given

$$l(v_1, v_2, \ldots, v_n, \widetilde{\beta}, \xi) = \begin{cases} -n\log\widetilde{\beta}-\left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\log\left(1+\frac{v_i}{\widetilde{\beta}}\right), & \xi\neq 0\\ -n\log\widetilde{\beta}-\frac{1}{\beta}\sum_{i=1}^{n}v_i & \xi=0\end{cases} \tag{13}$$

The MLEs (say $\widehat{\widetilde{\beta}}$ and $\hat{\xi}$) of $\widetilde{\beta}$ and ξ when $\xi\neq 0$ can be obtained by solving the equations given in (14)

$$\begin{cases}\frac{\partial l(v_1, v_2, \ldots, v_n, \widetilde{\beta}, \xi)}{\widetilde{\beta}}=0\\ \frac{\partial l(v_1, v_2, \ldots, v_n, \widetilde{\beta}, \xi)}{\xi}=0\end{cases} \tag{14}$$

simultaneously. Like GEV model, we could not solve the log-likelihood functions analytically, that is why the maximization is done by using a quasi-Newton numerical method. The standard errors of MLEs can be obtained asymptotically by inverting the information matrix.

In LMM paradigm, the parameters of GPD model were found similar to GEV model. The approximations for the parameters of GPD were also introduced by Hosking and Wallis (2005, p. 194). Linear moment defined for $\xi > -1$are briefly described in (supplementary materials file **A.2**)

3.5. Estimation of return levels for extreme value distributions

The estimates of the parameters of EVDs are not enough to study extreme environmental events. From the practical point of view, some other quantities (i.e., return levels) are estimated with the help of the fitted extreme value models. For instance, return level (RL) estimates play a dynamic role in the modelling of extreme winds for assessing hazards (for example destruction of infrastructures such as buildings and the spread of wildfires) linked with future periods adapting to EVDs models. On the other side, the prediction of future extreme wind speed is important to make a proper plan for the construction of wind energy power stations, and RLs are used for such prediction

The RLs for the GEV model opposite to the return period $T = 1/p$, represented by v_p where $F(v_p = 1-p)$ and $0 < p < 1$, is achieved by applying the quantile function with the inverse of (2) given by Coles (2001) and also discussed by (Ahmad et al., 2019; Diriba and Debusho, 2020).

$$v_p=\begin{cases}\alpha-\frac{\beta}{\xi}\left[1-\{-\log(1-p)\}^{-\xi}\right], & \xi\neq 0\\ \alpha-\beta\log\{-\log(1-p)\}, & \xi=0\end{cases} \tag{15}$$

In the GPD case, let v_p be the return level that is exceeded on the average once every p value. Thus, the v_p is acquired as a solution to $1-H(v_p,\widetilde{\beta},\xi)=1/p$ is sufficiently large to ensure that $v_p > u$, then v_p has the following form

$$v_p=\begin{cases}u-\frac{\widetilde{\beta}}{\xi}\left[1-\left\{p^{-\xi}-1\right\}\right], & \xi\neq 0\\ u-\widetilde{\beta}\log(p), & \xi=0\end{cases} \tag{16}$$

As a result, the return level v_p given in (15) and (16) is a quantile of GEV and GPD connected with the upper tail probability p. MLEs of the return level v_p, indicated by $\hat{v}_p$ is achieved by substituting the MLEs of the parameter vectors of GEV ($\hat{\alpha}$,$\hat{\beta}$,$\hat{\xi}$) and GPD ($\hat{\beta}$,$\hat{\xi}$), respectively.

3.6. Bayesian paradigm

As the likelihood approach, let $\{v_i, i = 1, 2, \ldots, n\}$ is i.i.d random variable which follows an EVDs (i.e. GEV or GPD) family. In the Bayesian modelling framework, the parameters of EVDs are treated as random variables for which we need to establish the prior distributions. From a practical point of view, prior knowledge assists the researchers to enhance the information offered by the observed data. Suppose $\boldsymbol{\psi_1} = (\alpha, \beta, \xi)$ and $\boldsymbol{\psi_2} = (\beta, \xi)$ are the vectors of parameters of EVDs and let $g_\psi(\boldsymbol{\psi}_i)$, $i = 1, 2$ express probability density function of the prior distribution for $\boldsymbol{\psi}_i$ $i = 1, 2$ with no evidence to the genuine data. Then applying Bayes theorem argument to discover the posterior distribution for $\boldsymbol{\psi}_i$ $i = 1, 2$ in the following form

$$f(\boldsymbol{\psi}_i|v)=\frac{L(\boldsymbol{\psi}_i|v)g_\psi(\boldsymbol{\psi}_i)}{\int_\Theta L(\boldsymbol{\psi}_i|v)g_\psi(\boldsymbol{\psi}_i)d\boldsymbol{\psi}}\propto L(\boldsymbol{\psi}_i|v)g_\psi(\boldsymbol{\psi}_i), \tag{17}$$

where $L(\boldsymbol{\psi}_i|v)$, $i = 1, 2$ show the likelihood function EVDs and Θ is the parametric space. For computational point of view, NIPs and IPs were considered. The NIPs were incorporated by assuming there is no extra knowledge available regarding parameters, apart from the data. On the other side, IPs were formed through the technique purposed by Coles and Tawn (2005) also used by Diriba et al. (2017), Diriba and Debusho (2020), Ahmad et al. (2019). According to this approach, extreme quantiles of the EVDs were taken into account to generate the prior information. A brief discussion about the approach proposed by Coles and Tawn (2005) is presented in the subsequent paragraphs.

Let the quantile function $[v_{p_i}, i = 1, 2, 3$ with $p_1 > p_2 > p_3]$ of GEV distribution in (15), be used to compute the quantiles from historical data nearby weather stations. The quantiles are assessed independently for each station by using the MLEs of GEV parameters. Hence, the joint prior distribution for GEV parameters could be produced via extreme quantiles $(v_{p_1}, v_{p_2}, v_{p_3})$ corresponding to $p_1 > p_2 > p_3$ probabilities. Due to natural ordering quantiles $v_{p_1} < v_{p_2} < v_{p_3}$, the assumption of independent priors on $v_{p_i}, i = 1, 2, 3$ is violated. That is why Coles and Tawn (2005) endorsed the quantile differences. In the present setting, the authors also used the quantile differences

$$\widetilde{v}_{p_i}=\widetilde{v}_{p_i}-\widetilde{v}_{p_{i-1}}, \qquad i=1,2,3 \tag{18}$$

where v_{p_0} is explained to a lower endpoint of the process variable (e.g.wind speed) and is usually considered $v_{p_0} = 0$. After the differencing of the quantiles, the appearance of quantiles in the ordering form validates the independence assumption. Hence, the priors based on the quantile differences are considered to be independent Gamma distributions with parameters (λ_i, γ_i), $i = 1, 2, 3$. That is,

$$\widetilde{v}_{p_i}\sim\Gamma(\lambda_i,\gamma_i), \qquad \lambda_i>0, \qquad \gamma_i>0; \qquad i=1,2,3. \tag{19}$$

The joint prior for the (λ_i, γ_i), $i = 1, 2, 3$, could be developed from the Gamma distribution in the following way

$$\widetilde{v}_{p_1}\sim\Gamma(\lambda_1,\gamma_1)\propto v_{p_1}^{\lambda_1-1}\exp(-\gamma_1 v_{p_1}),$$

$$\widetilde{v}_{p_2} \sim \Gamma(\lambda_2, \gamma_2) \propto (v_{p_2} - v_{p_1})^{\lambda_2 - 1} \exp(-\gamma_2(v_{p_2} - v_{p_1})),$$

and

$$\widetilde{v}_{p_3} \sim \Gamma(\lambda_3, \gamma_3) \propto (v_{p_3} - v_{p_2})^{\lambda_3 - 1} \exp(-\gamma_3(v_{p_3} - v_{p_2})).$$

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

$$\begin{aligned} g(v_{p_1}, v_{p_2}, v_{p_3}) &\propto v_{p_1}^{\lambda_1 - 1} \exp(-\gamma_1 v_{p_1}) \\ &\times (v_{p_2} - v_{p_1})^{\lambda_2 - 1} \exp\{-\gamma_2(v_{p_2} - v_{p_1})\} \\ &\times (v_{p_3} - v_{p_2})^{\lambda_3 - 1} \exp\{-\gamma_3(v_{p_3} - v_{p_2})\} \end{aligned}$$

and has composed in a short form

$$g(v_{p_1}, v_{p_2}, v_{p_1}) \propto \widetilde{v}_{p_1}^{\lambda_1 - 1} \exp(-\gamma_1 \widetilde{v}_{p_3}) \times \prod_{i=2}^{3} \widetilde{v}_{p_i}^{\lambda_i - 1} \exp(-\gamma_i \widetilde{v}_{p_i}) \quad (20)$$

with $v_{p_1} < v_{p_2} < v_{p_3}$ given that (Ahmad et al., 2019). By mixing (15) with (20) and multiplying by the Jacobian of the transformation from $(v_{p_1}, v_{p_2}, v_{p_3}) \rightarrow \boldsymbol{\psi} = (\alpha, \beta, \xi)$ precedes to a prior in terms of the GEV parameter vector ψ. That is, the form has given

$$g_{\psi}(\boldsymbol{\psi}) \propto \prod_{i=1}^{3} \widetilde{v}_{p_i}^{\lambda_i - 1} \exp(-\gamma_i \widetilde{v}_{p_i}) \times \mathrm{J} \quad (21)$$

with $v_{p_1} < v_{p_2} < v_{p_3}$, where the Jacobian has the following form

$$\mathrm{J} = \begin{vmatrix} \frac{\partial v_{p1}}{\partial \alpha} & \frac{\partial v_{p1}}{\partial \beta} & \frac{\partial v_{p1}}{\partial \xi} \\ \frac{\partial v_{p2}}{\partial \alpha} & \frac{\partial v_{p2}}{\partial \beta} & \frac{\partial v_{p2}}{\partial \xi} \\ \frac{\partial v_{p3}}{\partial \alpha} & \frac{\partial v_{p3}}{\partial \beta} & \frac{\partial v_{p3}}{\partial \xi} \end{vmatrix}$$

and the results is by

$$\mathrm{J} = \begin{cases} \frac{\beta}{\xi^2} \left| \sum_{i<j} (-1)^{i+j} (w_i \times w_j) \log\left(\frac{w_j}{w_i}\right) \right|, & i, j \in (1, 2, 3); \xi \neq 0 \\ \frac{\beta}{2} \left| \sum_{i<j} (-1)^{i+j} \log w_i \times \log w_j \log\left(\frac{w_j}{w_i}\right) \right|, & i, j \in (1, 2, 3); \xi = 0 \end{cases} \quad (22)$$

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

Similarly, the informative priors for the GPD were also obtained by using the historical data of different surrounding stations. More precisely, these informative priors for GPD were established in terms of extreme quantiles. The interested readers could find additional reading in, for instance (Diriba and Debusho, 2020; Ahmad et al., 2019; Beirlant et al., 2004; Coles and Tawn, 2005). Similar to GEV modelling, the assumption of independence priors would not be effective due to the natural ordering of the quantile in (16) i.e. $v_{p_1} < v_{p_2}$. Thus, we again considered the quantile difference $\widetilde{v}_{p_1} = v_{p_i} - v_{p_{i-1}}, i = 1,2$ of expression (16), where $v_{p_0} = 0$ is indicates the physical lower bound of the variable. For the GPD model, only two quantiles are required for scale and shape parameters. Frankly speaking, the authors are not considering the location parameter to GPD. Like the GEV model setting, again quantile differences are respected. These differences tend to be independent Gamma distributions with parameters (λ_i, γ_i), $i = 1, 2$, and have the following form

$$\widetilde{v}_{p_i} \sim \Gamma(\lambda_i, \gamma_i), \qquad \lambda_i > 0, \qquad \gamma_i > 0; \qquad i = 1, 2. \quad (23)$$

Additionally, using gamma distribution we can also consider

$$\widetilde{v}_{p_1} \sim \Gamma(\lambda_1, \gamma_1) \propto v_{p_1}^{\lambda_1 - 1} \exp(-\gamma_1 v_{p_1})$$

and

$$\widetilde{v}_{p_2} \sim \Gamma(\lambda_2, \gamma_2) \propto (v_{p_2} - v_{p_1})^{\lambda_2 - 1} \exp(-\gamma_2(v_{p_2} - v_{p_1})).$$

Hence, the joint prior for (v_{p_1}, v_{p_2}), by considering $v_{p_0} = 0$ has the following form

$$\begin{aligned} g(v_{p_1}, v_{p_2}) &\propto v_{p_1}^{\lambda_1 - 1} \exp(-\gamma_1 v_{p_1}) \\ &\times (v_{p_2} - v_{p_1})^{\lambda_2 - 1} \exp\{-\gamma_2(v_{p_2} - v_{p_1})\} \end{aligned} \quad (24)$$

given that $v_{p_1} < v_{p_2}$. Combining (23) with (24) and multiplying by the Jacobian of the transformation from $(v_{p_1}, v_{p_2}) \rightarrow \psi = (\beta, \xi)$, this precedes a prior in terms of the GPD parameter vector ψ and has been written in the following way

$$\begin{aligned} g_{\psi}(\boldsymbol{\psi}) \propto \left(\alpha + \frac{\beta}{\xi}(p_1^{-\xi} - 1)\right)^{\lambda_1 - 1} \exp\left(-\gamma_1 \left\{\alpha + \frac{\beta}{\xi}(p_1^{-\xi} - 1)\right\}\right) \times \\ \left(\frac{\beta}{\xi}(p_2^{-\xi} - p_1^{-\xi})\right)^{\lambda_2 - 1} \exp\left(-\gamma_2 \frac{\beta}{\xi}(p_2^{-\xi} - p_1^{-\xi})\right) \times |J| \end{aligned} \quad (25)$$

where Jacobian J is found as

$$\begin{aligned} \mathrm{J} &= \begin{vmatrix} \frac{\partial v_{p1}}{\partial \beta} & \frac{\partial v_{p1}}{\partial \xi} \\ \frac{\partial v_{p2}}{\partial \beta} & \frac{\partial v_{p2}}{\partial \xi} \end{vmatrix} \\ &= -\frac{\beta}{\xi}[(p_1 p_2)^{-\xi}(\log p_2 - \log p_1) - p_2^{-\xi} \log p_2 + p_1^{-\xi} \log p_1] \end{aligned}$$

The direct computation of the posterior densities $f(\boldsymbol{\psi}_i|v)$, $i = 1, 2$ of EVDs models is a challenging task. Therefore, the characteristics of the posterior distributions were estimated by engaging the MCMC procedure with the Metropolis–Hasting algorithm (Ahmad et al., 2019; Hastings, 1970).

3.7. Posterior predictive distributions

As discussed earlier, the core objective of extreme value analysis is often prediction. Generally, Bayesian analysis of extremes is played a vital role in this task through the posterior predictive distribution. For instance, if Y represents a future extreme of wind speed data with the density function $f(y|\psi)$, where $\psi \in \Psi$, and Ψ is a parametric space. As a result, the predictive distribution for our extremes based on EVDs is given that

$$\Pr(Y \leq y|v) = \int_{\psi} \Pr(Y \leq y|\boldsymbol{\psi}_i) f(\boldsymbol{\psi}_i|v) d\boldsymbol{\psi}, \quad (26)$$

where v shows the past observations of the process, $\boldsymbol{\psi}_i$, $i = 1, 2$ be generic parameters vectors of both extreme value models, and $f(\boldsymbol{\psi}_i|v)$ are posterior densities for $\boldsymbol{\psi}_i$, $i = 1, 2$. In addition, the $\Pr(Y \leq y|\boldsymbol{\psi}_i)$ $i = 1,2$ are the EVDs evaluated at Y.

4. Results and discussion

4.1. Testing the basic assumptions of extreme wind speed data

Before applying extreme value modelling to a series of excessive wind speeds, we tested the assumptions of the data. For instance, a trend in the series may affect the modelling. Therefore, we assessed assumptions (i.e., randomness, independence, homogeneity, and stationarity) using the following tests (i.e., NERC, Wald–Wolfowitz, Mann–Whitney, and Spearman test). The null hypothesis for all assumptions is that the respective assumption related to the data is not valid. For more details about the procedures [see, e.g., Naghettini (2017) and references therein]. Further, trend analysis was done by Kendall's tau test and Sen's slope estimators (Sen, 1968). The results of Table 2 clearly show that the data fulfil the above assumptions and can be used for

Table 2
Results of concerning assumptions of wind speed series using different tests. In brackets (p-value) is reported.

Station	NERC	WW	Mann–Whitney	Spearman	Kendall's tau	Sen's
Thatta	1.532 (0.063)	0.00 (0.497)	−1.119 (0.131)	−1.540 (0.062)	−0.175 (0.1503)	−0.186 (0.120)

Table 3
Estimated parameters with (standard errors in parentheses) and [confidence intervals in square brackets] for both GEV and GPD models via frequentist methods.

Models	MLEs			L-Moments Estimates		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\xi}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\xi}$
GEV	14.16 (0.22) [13.74, 14.59]	1.16 (0.15) [0.87, 1.45]	−0.15 (0.09) [−0.33, 0.03]	14.17 (0.22) [13.72, 14.59]	1.17 (0.16) [0.85, 1.47]	−0.18 (0.13) [−0.45, 0.06]
GPD	–	1.00 (0.21) [0.59, 1.421	−0.05 (0.15) [−0.34, 0.25]	–	0.95(0.14) [0.67, 1.45]	0.009 (0.17) [−0.34, 0.29]

further analysis. The probability value of Sen's slop test suggests no significant trend exists in wind speed data of Thatta station. Increased data length might change this argument.

4.2. Estimation of parameters using classical procedures

The Generalized extreme value and the generalized Pareto models were fitted to block maxima and daily maximum threshold exceedances, respectively, using MLE and LMM. For fitting GPD on daily maximum wind data, a suitable threshold $u = 14$ has been selected via the mean excess plot procedure. A stability plot for the shape and modified scale parameters can be used to demonstrate the effectiveness of the chosen threshold (Coles, 2001). Essentially, suppose the GPD is a sensible model for the exceedances above the threshold u. In that case, estimates of shape and modified scale parameters should be almost constant to all thresholds greater than u. Hence, the shape and scale parameters estimates are constant for thresholds higher than 14 for daily maximum wind speed.

The estimates based on MLE and LMM for the parameters of GEV (α, β and ξ) and/or GPD (β and ξ) with their associated standard errors (SE) and confidence intervals (CI) are given in Table 3. The SE and CI for the LMM estimates were estimated through the bootstrapping procedure. From Table 3, it can be seen that the estimated shape parameter is, therefore, less than zero for the GEV model, which specifies an upper tail of the appropriate extreme value models fitted to extreme wind speed. This suggests a thicker tail for GEV distribution, which delivers high quantiles. This is particularly accurate while the quantiles are estimated for higher return periods for maximum wind speed (Diriba and Debusho, 2020). Prior to calculating the return levels, the goodness-of-fit of both EVDs was checked. Thus, the probability, quantile, and return level plots for fitted GEV and GPD models to the extreme wind data are shown in Fig. 2 (a, b), respectively. The plots given in Fig. 2 recommends that the fitting requirements of EVDs have been contented.

The analysis was further elongated by using the classical parameter estimation methods to estimate the RLs of annual maxima and exceedances over the threshold for 2, 10, 25, 50 75, and 100 years of average recurrence interval. In Table 4, you can find the estimated RLs of EVD's based on MLE and LMM parameter estimation methods for annual maxima and POT. No significant differences are observed when comparing RLs of annual maximum over different return periods. On the other hand, the RLs for exceedances over the threshold slightly higher than the annual maxima. These results highlight the importance of using the full range of data available. The variation in RLs could be due to the skewness in daily yearly maximum wind speed data. Therefore, the median is a robust measure to define data summary for a skewed distribution, while the mean is not.

The change between the median and the mean can therefore signify the magnitude of irregular values in the return levels. This is further examined in the Bayesian approach.

4.3. Bayesian modelling of extreme wind speed data

This section deals with the findings of Bayesian analysis to the annual maxima and daily maximum wind speed data. Non-informative and informative priors were used. We constructed informative priors based on historical records independently for each of the surrounding weather stations (namely, Badin, Hyderabad, Jamshoro, Nawabshah, Naushahro Feroze). One more station (named Lakhpat) from the Indian side was also considered to ensure the accuracy of the study. The above-considered weather stations are situated at various distances from the Thatta station. The effects of various distances on parameters and return level estimates were assessed.

4.3.1. Priors effect on parameters of extreme value distributions

The fundamentals of the Bayesian paradigm require the prior distributions for GEV and GPD parameters denoted by $\boldsymbol{\psi_1} = (\alpha, \beta, \xi)$ and $\boldsymbol{\psi_2} = (\beta, \xi)$. Essentially, the non-informative for GEV and GPD parameters were fabricated by considering there is was no extremal information about the process apart from the data. By following the procedure of Coles and Tawn (2005), Ahmad et al. (2021), the joint density of $\boldsymbol{\psi}_i, i = 1, 2$was supposed to be

$$f_{\psi}(\boldsymbol{\psi}_1) = f(\alpha, \eta = \log(\beta), \xi) = f_{\alpha}(\alpha) f_{\eta}(\eta) f_{\xi}(\xi) \quad \text{and}$$
$$f_{\psi}(\boldsymbol{\psi}_2) = f(\eta = \log(\beta), \xi) = f_{\eta}(\eta) f_{\xi}(\xi)$$

The following independent NIPs were used by (Coles and Tawn, 2005; Ahmad et al., 2021) as

$$f_{\alpha}(\alpha) \sim N(0, 10000), \quad f_{\eta}(\eta) \sim N(0, 10000), \quad f_{\xi}(\xi) \sim N(0, 100)$$

where $N(0, 10000)$ represents a Gaussian distribution with mean 0 and variance 10,000. The higher variance guarantees the absence of external information.

The IPs were constructed by using the historical information of wind speed from nearby stations. To achieve more precision about the study, the historical records of one station (namely Lakhpat) from the Indian territory were also used to develop IPs for EVDs parameters. The procedure illustrated in Section 3.6 was practiced to build the IPs for GEV and GPD. For instance, $\widetilde{v}_{p_1} \sim \Gamma(133.462, 0.143)$, $\widetilde{v}_{p_2} \sim \Gamma(0.849, 8.948)$ and $\widetilde{v}_{p_3} \sim \Gamma(0.184, 69.367)$ quantiles were obtained by the authors for GEV model using historical records of the Lakhpat station. Same procedure was repeated to elicit IPs for other stations. Table 5, illustrates the posteriors means, SEs and confidence intervals of GEV and GPD parameters from NIPs and IPs. The posterior means

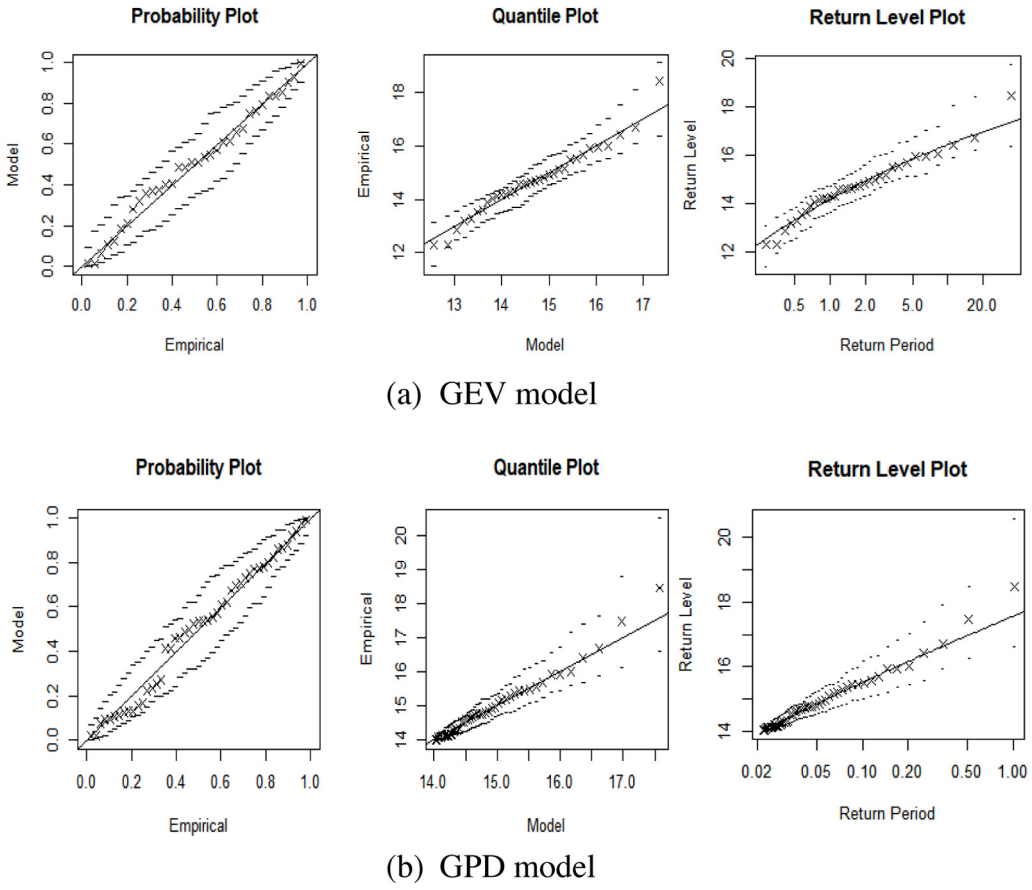

Fig. 2. The probability, quantile, and return level plots: (a) fitted GEV and (b) fitted GPD model.

Table 4
Estimated return levels and 95% confidence interval for GEV and GPD through frequentist methods.

Models	Maximum likelihood			L-Moments Estimates	
	Return period (years)	Estimates	[Confidence intervals]	Estimates	[Confidence intervals]
GEV	2	14.58	[14.13, 15.02]	14.59	[14.11, 15.02]
	5	15.72	[15.16, 16.27]	15.71	[15.11, 16.21]
	10	16.37	[15.71, 17.04]	16.34	[15.62, 16.91]
	25	17.10	[16.24, 17.97]	17.01	[16.05, 17.85]
	50	17.58	[16.51, 18.65]	17.44	[16.34, 18.74]
	75	17.84	[16.63, 19.05]	17.67	[16.44, 19.23]
	100	18.01	[16.70, 19.32]	17.82	[16.54, 19.59]
GPD	2	15.00	[14.56, 15.44]	14.97	[14.65, 15.30]
	5	15.86	[15.26, 16.46]	15.86	[15.30, 16.36]
	10	16.49	[15.72, 17.27]	16.54	[15.74, 17.22]
	25	17.29	[16.13, 18.46]	17.44	[16.19, 18.65]
	50	17.88	[16.28, 19.47]	18.12	[16.43, 20.11]
	75	18.21	[16.31, 20.11]	18.52	[16.55, 21.15]
	100	18.44	[16.30, 20.58]	18.81	[16.64, 21.94]

of the parameters EVDs parameters based on IP's are closed to the posterior means of NIPs. Table 5, show that the IPs produced from all stations reduced the posterior SD of EVDs except Lakhpat station. These decay in SD might reflect the diminution in uncertainty due to the incorporation of supplementary information from the neighbouring stations. When compared with standard errors of MLE and LMM, all the SDs of the EVDs parameters for the IPs and NIPs were smaller except Lakhpat station and scale parameter obtained through LMM.

To investigate how the information based on the historical records affected the EVDs parameters, the estimated posterior densities plots of GEV parameters for block maxima and GPD parameters for POT are given in Fig. 3. The distributions of location parameter of GEV and scale parameters of both models are look like symmetric. In addition, the densities of scales parameters of EVDs for informative priors built from the data of Naushero Firoz station had high peaks at the centre. On other hand, the densities of shape parameters of EVDs also had high peaks at centre for informative priors constructed from the information of Hyderabad station. This validates the findings of Table 5 that the IPs elicited have had an influence on EVDs parameters.

Table 5
Estimated parameters with (standard errors in parentheses) and [confidence intervals in square brackets] for both GEV and GPD models through Bayesian Paradigm.

	Parameter estimates of GEV and GP distributions				
	GEV			GPD	
Priors	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\xi}$	$\hat{\beta}$	$\hat{\xi}$
Non-informative					
	14.12 (0.23) [13.68, 14.58]	1.23 (0.17) [0.96, 1.63]	−0.12 (0.10) [−0.29, 0.12]	0.98 (0.21) [0.61, 1.44]	0.07 (0.18) [−0.22, 0.49]
Informative					
Hyderabad	14.23 (0.19) [13.83, 14.59]	1.23 (0.14) [0.99, 1.54]	−0.22 (0.05) [−0.32, −0.13]	1.28 (0.18) [0.96, 1.66]	−0.28 (0.06) [−0.33, −0.09]
Nawabshah	14.10 (0.20) [13.69, 14.49]	1.16 (0.14) [0.93, 1.47]	−0.17 (0.06) [−0.28, −0.04]	1.06 (0.17) [0.78, 1.43]	−0.11 (0.08) [−0.26, 0.06]
Badin	14.14 (0.21) [13.72, 14.54]	1.20 (0.15) [0.96, 1.54]	−0.16 (0.07) [−0.30, −0.017]	1.09 (0.18) [0.77, 1.50]	−0.10 (0.09) [−0.26, 0.09]
Naushahro-Firoz	14.03 (0.19) [13.63 ,14.39]	1.11 (0.13) [0.89, 1.40]	−0.16 (0.06) [-0.27 −0.05]	1.00 (0.15) [0.72, 1.32]	−0.11 (0.07) [−0.24, 0.04]
Jamshoro	14.18 (0.21) [13.75, 14.60]	1.24 (0.15) [0.98, 1.59]	−0.16 (0.07) [−0.30, −0.01]	1.16 (0.20) [0.81, 1.61]	−0.10 (0.09) [−0.28, 0.08]
Lakhpat	14.16 (0.24) [13.706, 14.622]	1.28 (0.17) [0.99, 1.67]	−0.06 (0.12) [−0.27, 0.21]	0.99 (0.29) [0.67, 1.47]	0.13 (0.19) [-0.20 0.53]

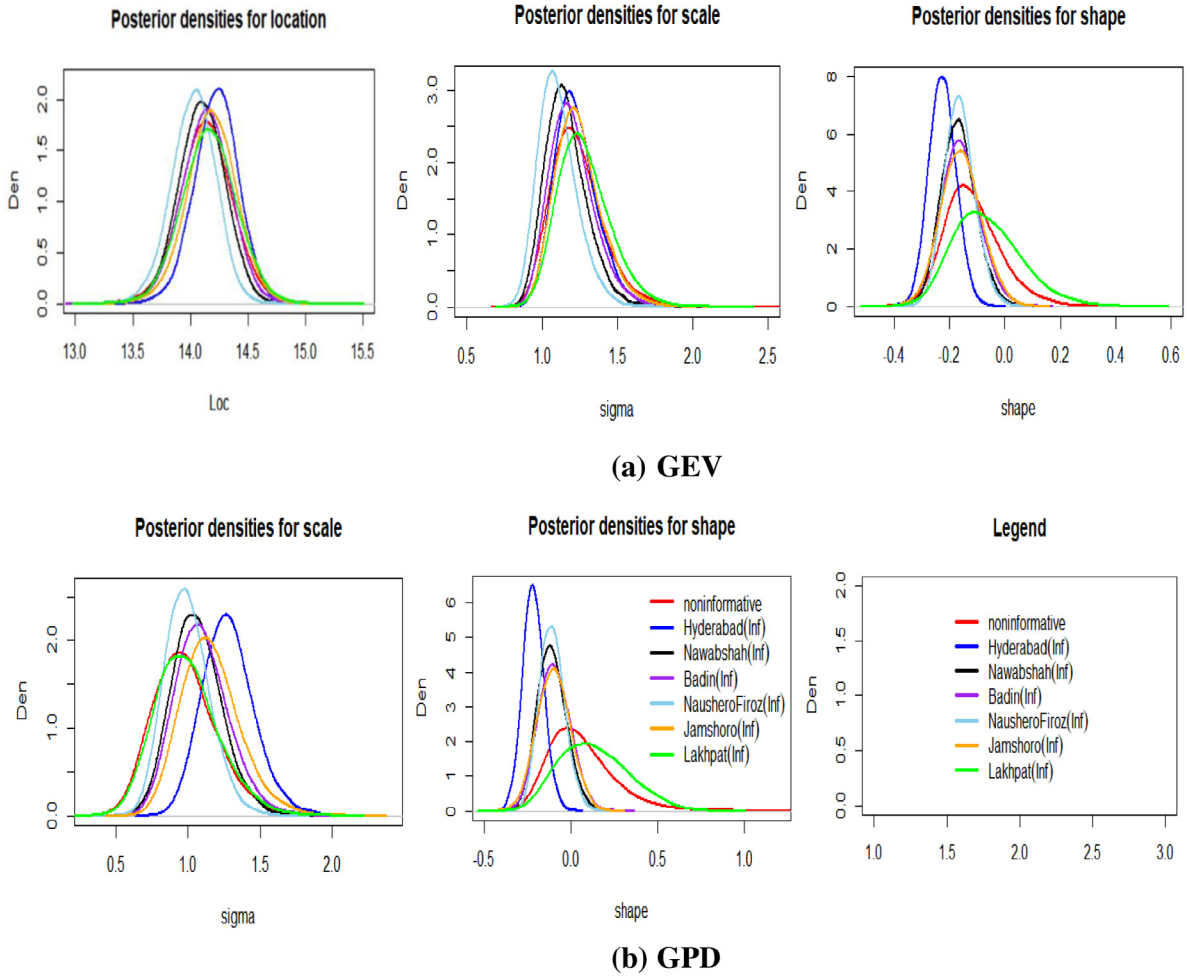

Fig. 3. Estimated Posterior densities of the generalized extreme value and generalized Pareto distribution parameters to annual maximum and daily maximum wind speed data using non-informative and informative priors.

4.3.2. Influence of priors on return levels

To explore the influence of the NIPs and IPs on the return levels of EVDs, the posterior density plots were constructed by putting the vector of observations from marginal posterior distribution of EVDs parameters into respective quantile functions define in (15) and (16), respectively, for 0 < p < 1. Similar to

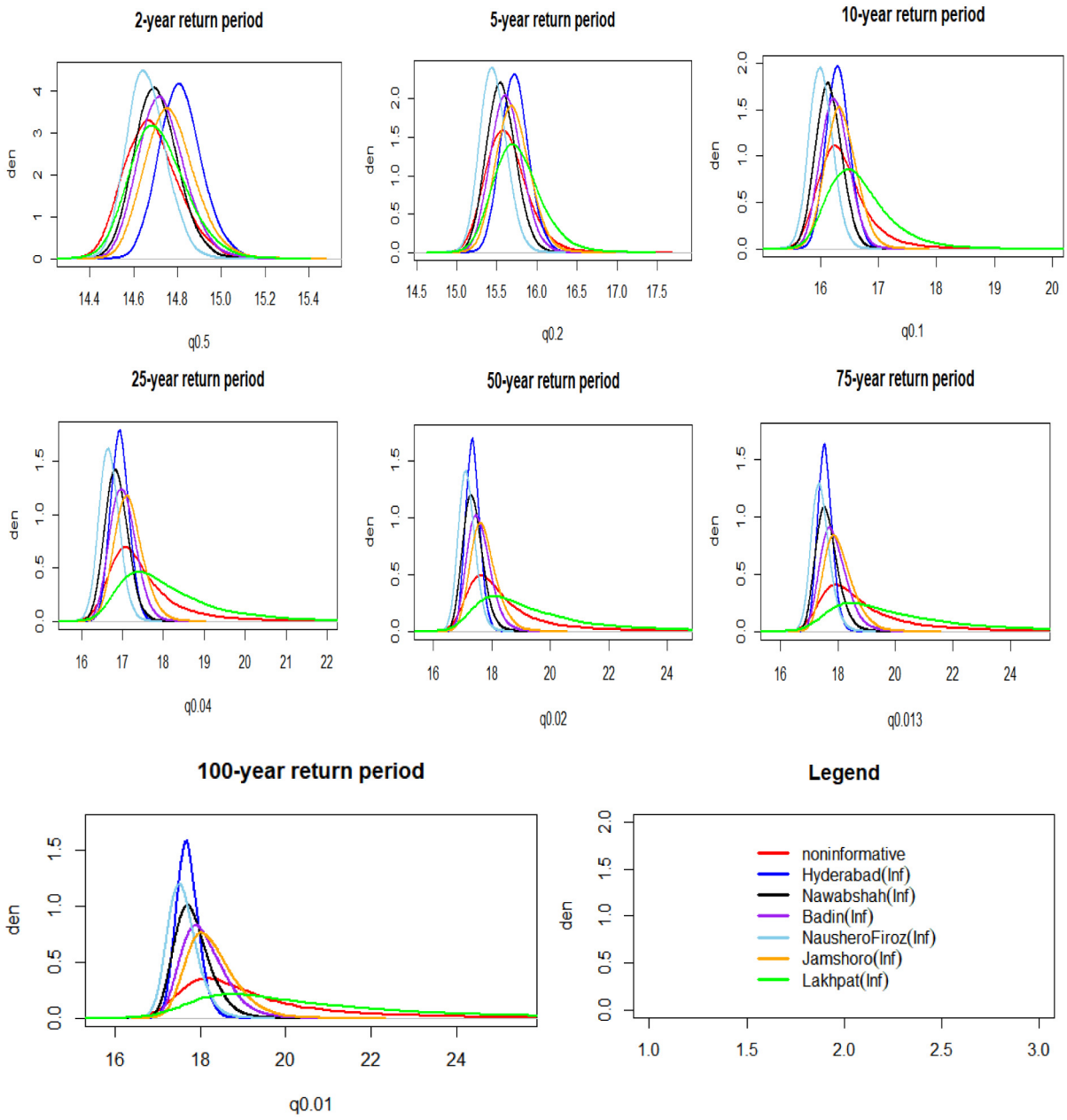

Fig. 4. Posterior densities for 2-, 5-, 10-, 25-, 50-, 75-, and 100-year return level.

frequentist methods, the posterior distribution of 2-, 5-,10-, 25-, 50-, 75-, and 100- years RLs were obtained corresponding the different value of $p = 0.5, 0.5, 0.1, 0.04, 0.02, 0.013$, and 0.01. The graphical evaluation of the densities of return levels based on the Bayesian parameters estimation method with NIPs and IPs for block maxima and daily wind speed data are also shown in this study; subsequently, the plotted densities of the RLs of all exceedance of wind speed data are shown in Fig. 4. In addition, the RLs density of block maxima are given in Figure S1 "see Supplementary materials file".

It is evidenced from Fig. 4 that the priors had an influence on the distribution of RLs. The posterior densities based on IPs have more elevated compared with the posterior densities of NIPs except Lakhpat station. This decay in RLs densities for Lakhpat station might be due to longer distance from the Thatta station. It can be observed that the distributions of RLs above 2-years are also slightly skewed to right tail. Further, the posterior distributions of the RLs are also sensitive to IPs from which the prior knowledge was produced.

In frequentist settings, the different mean RL values were observed for both models, which could be due to skewness perceived in the data. Hence, optimal selection of summary measures used for the RLs could be improved the results found from the posterior distribution. Additionally, the skewed RLs densities might imitate the uncertainty within the model for forming upper limits of the RLs comparative to lower limits for longer return periods (Coles and Tawn, 2005). The posterior median of RLs for EVDs based on IPs and NIPs are presented in Table 6. When comparing the results of Table 6 with Table 4, the posterior medians of RLs of the GEV model are close to the mean RLs of MLE and LMM, except for minor variation detected for some return level values. On the other hand, the median of RLs GPD

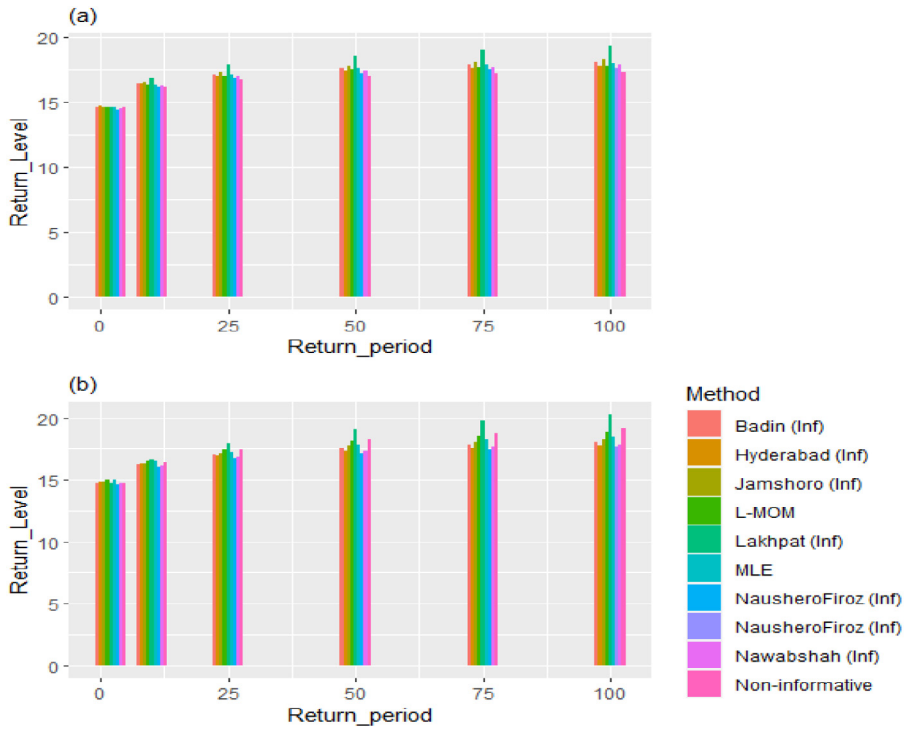

Fig. 5. Comparison of the return levels (a) block maxima of wind speed data using different GEV parameters estimation techniques (b) exceedances over threshold of daily wind speed data using different GPD parameters estimation techniques.

Table 6
Estimated return levels corresponding 2-, 5-, 10-, 25-, 50-, 75-, and 100- years of GEV and GPD based on NIPs and IPs methods.

Models	Return levels						
	2	5	10	25	50	75	100
GEV							
Non-informative	14.57	15.64	16.18	16.71	17.02	17.18	17.2
Informative							
Hyderabad	14.66	15.78	16.39	17.02	17.41	17.615	17.73
Nawabshah	14.51	15.64	16.28	16.98	17.43	17.67	17.83
Badin	14.56	15.77	16.40	17.14	17.62	17.87	18.04
NausheroFiroz	14.42	15.50	16.11	16.78	17.22	17.45	17.60
Jamshoro	14.63	15.84	16.53	17.30	17.80	18.06	18.24
Lakhpat	14.63	16.00	16.85	17.87	18.59	18.99	19.27
GPD							
Non-informative	14.69	15.64	16.40	17.45	18.27	18.77	19.13
Informative							
Hyderabad	14.82	15.74	16.32	16.97	17.37	17.59	17.73
Nawabshah	14.71	15.57	16.16	16.88	17.38	17.65	17.84
Badin	14.74	15.64	16.27	17.04	17.58	17.88	18.09
NausheroFiroz	14.66	15.47	16.03	16.71	17.17	17.43	17.61
Jamshoro	14.77	15.72	16.38	17.18	17.74	18.05	18.26
Lakhpat	14.72	15.78	16.67	17.97	19.06	19.75	20.26

establishes with IPs looking higher than mean return levels of MLEs and LMM. In addition, the RLs linked with Lakhpat station IPs are quite larger for the higher return period (for example, 100- years). The visual assessment of the estimated RLs of block maxima and POT corresponding to 2-, 5-, 10-, 25-, 50-, 75-, and 100- years return periods based on adopted different parameters estimation methods are shown in Fig. 5 (a, b), respectively. It is verified from Fig. 5 that the Bayesian paradigm with Lakhpat IPs for all exceedances show a substantial variability in RL of extreme wind speed data while examining for higher return period; in contrast, RLs for lower return periods displays slight difference for all methods. For verification point of view, posterior predictive distributions of EVDs with NIPs and IPs (with historical records of Hyderabad station) were constructed corresponding to 1, 2 and 5-years return periods. Figure S2 (a, b) correctly verified the RLs' findings based on Bayesian settings.

Generally, the Bayesian approach results showed a significant improvement in the inference for the extreme wind speed data compared with frequentist methods. Frankly speaking, the IPs for the Bayesian paradigm constructed from neighbouring weather stations improve the precision of the parameter estimates compared with the frequentist techniques. Hence, it can be claimed that the uncertainties added in the Bayesian framework in term of IPs has significantly enhanced the estimates for wind speed data of Thatta station. The precision could be increased by using the historical information of more suitable stations for prior construction (Ahmad et al., 2021).

4.4. Models assessment

Boxplots in Fig. 5 evaluate the efficiency of EVDs and the estimation procedures as well. Moreover, the boxplots are constructed corresponding to Root Mean Square Error (RMSE), Relative Root Mean Square Error (RRMSE). It can be acknowledged from Fig. 5 that the GPD based on Bayesian MCMC with IPs except the Lakhpat station is more appropriate to the wind speed modelling at Thatta station. In addition, the boxplots of GEV corresponding RMSE are also showing accurate fitting for all methods. In contrast, GEV boxplots based on RRMSE display the best fit for frequentist methods and Bayesian MCMC with Hyderabad, Badin and Lakhpat IPs. On the other hand, the RMSE and RRMSE boxplots construction for GPD are mainly proposing a good fit with smaller values of RMSE and RRMSE except for Bayesian MCMC with NIF and Lakhpat IPs. Hence, due to the lowest values of assessment measures, the GPD with Bayesian MCMC technique based on IPs is considered the most suitable and efficient choice for modelling wind speed data at Thatta station (see Fig. 6).

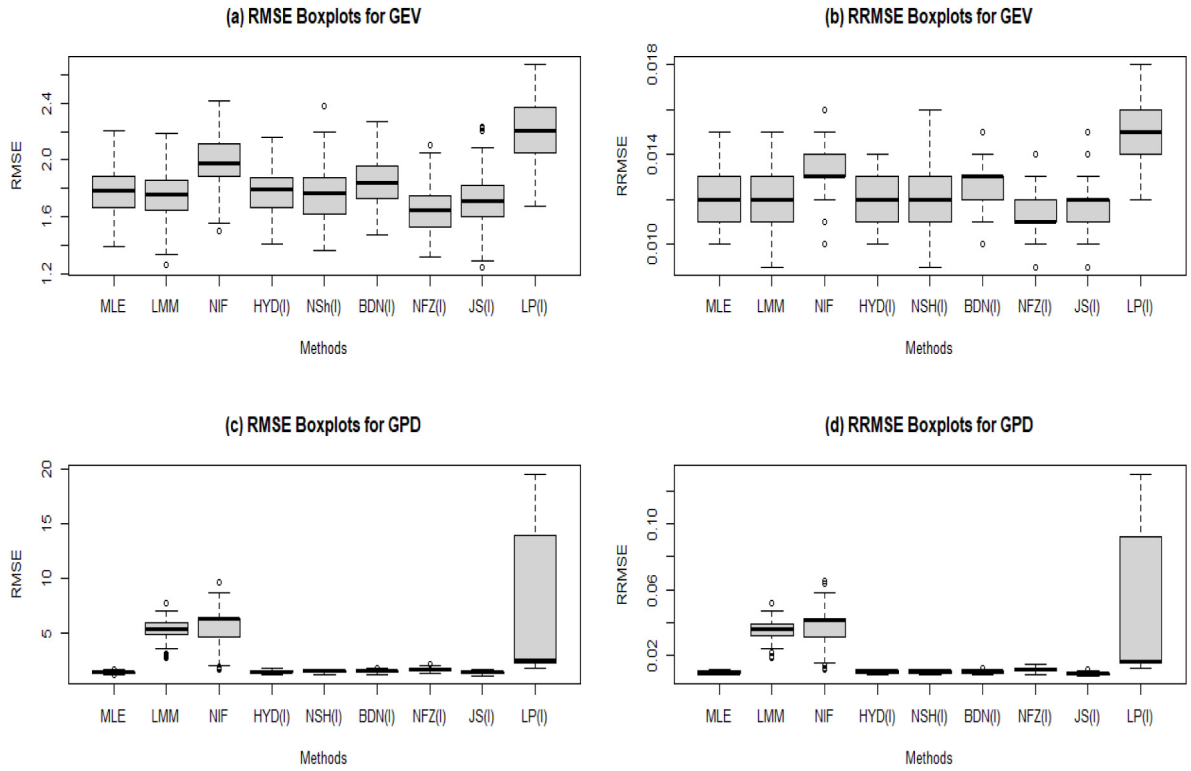

Fig. 6. Comparison of estimation methods and models (a) and (b) the root mean square error and relative root mean square error for GEV model corresponding to different estimation methods; (c) and (d) the root mean square error and relative root mean square error for GP model corresponding to different estimation methods.

5. Conclusions and recommendations

To conclude, this paper dealt with modelling of extreme wind speed at the Thatta location using the EVDs. Classical (MLE and LMM) and the Bayesian MCMC (with NIPs and IPs) techniques were employed to meet the inferential requirements of the EVDs models. Primarily, the inherent trend and the basic assumptions of the data were verified using different statistical tests, and no specific movement was found. On the other hand, this study offers artificial aids in terms of greater accuracy in estimating the considered models parameters. In addition, the parameters of EVDs, inferences for return levels, have also been enriched in precision. The present study's findings differed from the literature regarding applications on the topic.

Modelling with the Bayesian paradigm advances this paper to model the extreme behaviour of an extreme event at a given weather station. This approach is convenient when climatic evidence is scarce and assumes that extreme wind speed behaviour is homogeneous over the data's regions. Consequently, the authors would like to put a ball in the court of the Bayesian technique over the frequentists methods. The Bayesian framework needs a genuine construction of IPs, which delivers an augmented estimation accuracy. The parameter and RLs estimates for the extreme value models were sensitive to those weather stations used to develop IPs. Therefore, the present study also discusses the proper selection of the surrounding weather stations. Because the elicitation of IPs is essential as the estimates and their precision are linked to these priors.

In practice, the standard approach and the best estimation method were decided through assessment measures (see Section 4.4). Frankly speaking, wind speed modelling link with a peak over threshold approach show an enhancement over block maxima in that standard errors for GP model parameters are smaller. Overall, the RMSEs and RRMSEs of the GP model with IP-based Bayesian framework were lower than all others. The stability of the GP distribution in both Bayesian and frequentist settings authorizes as the predominant distribution to model extreme wind speed data at the Thatta station. Interestingly, GP distribution with Bayesian IPs produces RLs more accurately than others. The RLs values would be considered a future prediction of extreme wind. Based on the findings of this study, the authors recommend using GP distribution for modelling daily extreme wind speed data. More importantly, the future prediction of extreme winds given in this study could be helpful for engineers during the installation of wind energy turbines in the region. Furthermore, wind risk management, wind prediction, might be useful to install wind turbines at ungauged stations and those stations which are more likely to receive heavy wind and for better policy implications at vulnerable areas to wind disasters. Also, these predictions are essential to assist the government departments in protection preparedness to upcoming windstorms due to climate change. This study allows practitioners to consider variability and seasonality under climatic fluctuations. To some extent, it would be helpful in the production of wind energy at ungauged stations and also those areas where wind disasters have been seen but no turbines are installed there. We can opt for optimum structure engineering in these places plus can install some turbines to produce energy. Furthermore, this study can be improved by considering spatial settings or multivariate extreme value modelling or using extended versions of GP distribution.

CRediT authorship contribution statement

Touqeer Ahmad: Data curation, Investigation, Methodology, Software, Writing – original draft. **Ishfaq Ahmad:** Conceptualization, Project administration, Supervision, Validation. **Irshad Ahmad Arshad:** Formal analysis, Validation, Writing – review &

editing. **Ibrahim Mufrah Almanjahie:** Data curation, Methodology, Visualization, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Authors are very grateful to the CARIPARO foundation, Italy, for funding to this research. The corresponding author also avails of the PhD scholarship from the same foundation. Also, the authors would like to the express their gratitude to the University of Padova, Italy for giving administrative and technical support. In addition, the authors would thank to four referees and the editor for their useful comments and suggestions, which helped to improve the paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.egyr.2023.01.093.

References

Ahmad, I., Ahmad, T., Almanjahie, I.M., 2019. Modelling of extreme rainfall in punjab: pakistan using bayesian and frequentist approach. Appl. Ecol. Environ. Res. 17 (6), 13729–13748. http://dx.doi.org/10.15666/aeer/1706_1372913748.

Ahmad, T., Ahmad, I., Arshad, I.A., Bianco, N., 2021. A comprehensive study on the Bayesian modelling of extreme rainfall: a case study from Pakistan. Int. J. Climatol. http://dx.doi.org/10.1002/joc.7240.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.L., 2004. Statistics of Extremes: Theory and Applications. John Wiley & Sons.

Coles, 2001. An Introduction to Statistical Modeling of Extreme Values. Vol. 208, Springer, London.

Coles, S.G., Powell, E.A., 1996. Bayesian methods in extreme value modeling: a review and new developments. Internat. Statist. Rev. 64 (1), 119–136. http://dx.doi.org/10.2307/1403426.

Coles, S., Tawn, J., 2005. Bayesian modelling of extreme surges on the UK east coast. Phil. Trans. R. Soc. A 363 (1831), 1387–1406. http://dx.doi.org/10.1098/rsta.2005.1574.

Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds. J. R. Stat. Soc. Ser. B Stat. Methodol. 52 (3), 393–425, https://www.jstor.org/stable/2345667.

de Oliveira, M.M.F., Ebecken, N.F.F., de Oliveira, J.L.F., Gilleland, E., 2011. Generalized extreme wind speed distributions in South America over the Atlantic Ocean region. Theor. Appl. Climatol. 104 (3), 377–385. http://dx.doi.org/10.1007/s00704-010-0350-3.

Diriba, T.A., Debusho, L.K., 2020. Modelling dependency effect to extreme value distributions with application to extreme wind speed at Port Elizabeth, South Africa: a frequentist and Bayesian approaches. Comput. Statist. 1–31. http://dx.doi.org/10.1007/s00180-019-00947-2.

Diriba, T.A., Debusho, L.K., Botai, J., Hassen, A., 2017. Bayesian modeling of extreme wind speed at Cape Town, South Africa. Environ. Ecol. Stat. 24 (2), 243–267. http://dx.doi.org/10.1007/s10651-017-0369-z.

Fawad, M., Ahmad, I., Nadeem, F.A., Yan, T., Abbas, A., 2018. Estimation of wind speed using regional frequency analysis based on linear-moments. Int. J. Climatol. 38 (12), 4431–4444. http://dx.doi.org/10.1002/joc.5678.

Ferreira, A., De Haan, L., 2015. On the block maxima method in extreme value theory: PWM estimators. Ann. Statist. 43 (1), 276–298. http://dx.doi.org/10.1214/14-AOS1280.

Haq, M.A., Chand, S., Sajjad, M.Z., Usman, R.M., 2020. Evaluating the suitability of two parametric wind speed distributions: a case study from Pakistan. Model. Earth Syst. Environ. 1–9. http://dx.doi.org/10.1007/s40808-020-00899-3.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109. http://dx.doi.org/10.1093/biomet/57.1.97.

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

Hulio, Z.H., Jiang, W., Rehman, S., 2017. Technical and economic assessment of wind power potential of Nooriabad, Pakistan. Energy Sustain. Soc. 7 (1), 1–14. http://dx.doi.org/10.1186/s13705-017-0137-9.

Jocković, J., 2012. Quantile estimation for the generalized Pareto distribution with application to finance. Yugosl. J. Oper. Res. 22 (2), http://dx.doi.org/10.2298/YJOR110308013J.

Kamal, L., Jafri, Y.Z., 1997. Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. Sol. Energy 61 (1), 23–32. http://dx.doi.org/10.1016/S0038-092X(97)00037-6.

Khan, M.A., Çamur, H., Kassem, Y., 2019. Modeling predictive assessment of wind energy potential as a power generation sources at some selected locations in Pakistan. Model. Earth Syst. Environ. 5 (2), 555–569. http://dx.doi.org/10.1007/s40808-018-0546-6.

Madsen, H., Rasmussen, P.F., Rosbjerg, D., 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. Water Resour. Res. 33 (4), 747–757. http://dx.doi.org/10.1029/96WR03848.

Martins, E.S., Stedinger, J.R., 2001. Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series. Water Resour. Res. 37 (10), 2559–2567. http://dx.doi.org/10.1029/2000WR000009.

Naghettini, M. (Ed.), 2017. Fundamentals of Statistical Hydrology. Springer International Publishing, Switzerland.

Rafique, M.M., Rehman, S., 2017. National energy scenario of Pakistan–current status, future alternatives, and institutional infrastructure: An overview. Renew. Sustain. Energy Rev. 69, 156–167. http://dx.doi.org/10.1016/j.rser.2016.11.057.

Ranjbar, S., Cantoni, E., Chavez-Demoulin, V., Marra, G., Radice, R., Jaton-Ogay, K., 2020. Modelling the extremes of seasonal viruses and hospital congestion: The example of flu in a swiss hospital. arXiv preprint arXiv:2005.05808.

Raza, W., Hammad, S., Shams, U., Maryam, A., Mahmood, S., Nadeem, R., 2015. Renewable energy resources current status and barriers in their adaptation for Pakistan. J. Bioprocess. Chem. Eng. 3 (3), 1–9.

Saulat, H., Khan, M.M., Aslam, M., Chawla, M., Rafiq, S., Zafar, F., Khan, M.M., Bokhari, A., Jamil, F., Bhutto, A.W., Bazmi, A.A., 2020. Wind speed pattern data and wind energy potential in Pakistan: current status, challenging platforms and innovative prospects. Environ. Sci. Pollut. Res. 1–23. http://dx.doi.org/10.1007/s11356-020-10869-y.

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. J. Amer. Statist. Assoc. 63 (324), 1379–1389. http://dx.doi.org/10.1080/01621459.1968.10480934.

Shahzad, M.N., Kanwal, S., Hussanan, A., 2020. A new hybrid ARAR and neural network model for multi-step ahead wind speed forecasting in three regions of Pakistan. IEEE Access 8, 199382-199392. http://dx.doi.org/10.1109/ACCESS.2020.3035121.

Sumair, M., Aized, T., Gardezi, S.A.R., Bhutta, M.M.A., Rehman, S.M.S., Ur Rehman, S.U., 2021. Comparison of three probability distributions and techno-economic analysis of wind energy production along the coastal belt of Pakistan. Energy Explor. Exploitation 39 (6), 2191–2213. http://dx.doi.org/10.1177/0144598720931587.

Wang, Q.J., 1991. The POT model described by the generalized Pareto distribution with Poisson arrival rate. J. Hydrol. 129 (1–4), 263–280. http://dx.doi.org/10.1016/0022-1694(91)90054-L.