

Modeling of Reversed Fréchet Distribution under Linear Normalizing

H. A. Alaswed^{1*}

¹Assistant Professor at Statistics Department, Faculty of Science, Sebha University-Libya

DOI: <https://doi.org/10.36347/sjpm.2024.v11i11.002>

| Received: 18.09.2024 | Accepted: 27.10.2024 | Published: 06.11.2024

*Corresponding author: H. A. Alaswed

Assistant Professor at Statistics Department, Faculty of Science, Sebha University-Libya

Abstract

Review Article

Extreme value theory (EVT) is becoming increasingly important as a suitable model of extreme events to represent phenomena in various fields of applications where extreme values may appear and have detrimental effects. In EVT, the Reversed Fréchet (RF) distribution is a special case of the Reversed generalized extreme value distribution (RGEVD) for modeling extreme data under linear normalizing. The purpose of this paper is divided into four objectives: Firstly, we discuss the practical aspects of the two methods that belong to the domain attraction of RF distribution. The second objective is to use RF distribution to model the behavior of two extreme air pollutants. The third objective is based on a diagnostics plot and hypothesis testing used to evaluate an appropriate shape of the tail distribution. Lastly, for the fourth objective, we attempt to obtain the return level period (RLP) that is expected to be exceeded. Maximum likelihood estimation (MLE) is derived and used to estimate the parameters of RF distribution. The results show that applications are significant for modeling by RF distribution in all cases of two air pollutants. Software that was used in all computations and graphics is the R statistical program with packages fExtreme and ismev.

Keywords: Main Limiting Results in EVT, Reversed Fréchet Distribution, Determining the Domain Attraction of RF distribution, Parameter Estimation, Diagnostic plots, Hypothesis Testing.

Copyright © 2024 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

1 INTRODUCTION

Extreme value theory (EVT) is used to develop a statistical tool capable to describe and evaluate rare or extreme events, mainly which are considerably larger or smaller compared to usual happenings. Normally, the extreme values (EVs) refer to rare event, that means the analysts often need to assess the probability of occurrence of unusual extreme event than what has been observed in available set of data. The dare of EVT is to extrapolate beyond the available data and predict the unusual. Authors in [1], pioneered the contributions to this area, where they recognized a potential limit distribution for the maximum values in a dataset. Following this, [2], established that the EVT can only belong to one of three fundamental types of distribution namely; Gumbel, Fréchet, and Weibull. This revolutionary comprehension established the foundation for further developments, including the work by [3], who introduced basic but essential conditions for the weak convergence of maxima to these three types of EVT distributions. The authors in [4], built on this by founding a rigorous theoretical groundwork for EVT, providing necessary and sufficient conditions for the weak convergence issue in extremes. The EVT were further generalized by authors in [4], and studied in depth by [5], who expanded the applicability of EVT. Further improvement has been introduced by [6], who refined the understanding of asymptotic distributions. Subsequently, many authors have later expanded on these results and offer even more nuanced insights. In [7], the focus shifted to discussing various asymptotic results for extreme order statistics, adding further complexity to the field. The authors of [8], explored the practical applications of theory in statistical frameworks, while [9], introduced a latest, essential, and adequate condition that has advanced the theoretical understanding of EVT. Later [10], provided an innovative alternative formulation for EVT called von Mises conditions, offering another level of clarity. EVT plays a fundamental role across many fields, as specified by studies like [11, 12]. For the selection and validation of EVT models, the primary tools used are Probability Paper Plots (PPP), Quantile-Quantile Plots (QQP), and Return Level Plots (RLP). These tools are described extensively and demonstrated with several examples in [8-19]. To simplify evaluation process, packages of computer programs designed for EVT are available for download online, as referenced in [14-16]. For a more exhaustive theoretical elucidation and definition of the concepts involved in EVT, readers are directed to consult [17].

This paper is organized in the follows manner: In Section 2, light shad on a summary of the key results related to extreme value theory (EVT). Section 3 offers a general overview of some important findings regarding the RF distribution. While, section 4, discusses two approaches to determine whether a given distribution belongs to the domain of attraction (DA) condition of the RF distribution under linear normalization. Section 5 presents parameter estimation methods. In Section 6, focuses the diagnostic plots to assist in the selection of a suitable model. Section 7 is devoted to demonstrating hypothesis testing for the RF model. Section 8 reveals the application of the RF distribution to actual dataset. The final section provides conclusion and future work for possibly new research directions.

2 Main Limiting Results in EVT

The groundwork in EVT traced back to the work of authors in [1-3]. In [4], EVT summaries the possible limiting behavior of categorizations of maximum or minimum values under linear normalization, with an incomplete description that was fully achieved in [12], concerning the domains of attraction. One of the primary limiting results in EVT is summarized in the following theorem. It can be achieved by a well-established iterative procedure, like the Newton-Raphson method. This numerical method provides an effective solution, typically when analytical solution is not available. For further details on the application and theoretical background of these numerical techniques, reader is referred to [18].

2.1 Theorem of Limiting Results of Extreme Distributions

Consider $(X_i, i \in N)$ as a set of independent random variables with an identical distribution specified using the distribution function (d.f.) F . Suppose that the initial n random variables are sequenced in increasing order of magnitude and specified as $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$; in this case, the *ith* ordinary order statistic is denoted as $X_{i:n}$. Specifically, $X_{1:n} = \min_{i=1, \dots, n} X_i = -\max_{i=1, \dots, n} (-X_i)$ designate the lower values among X_1, X_2, \dots, X_n , hence, $P(X_{1:n} < x) = 1 - (1 - F(x))^n$ and let \mathfrak{R} specify a real line and F d.f. is considered among the min DA based on the linear normalisation of a non-degenerate d.f. G . It is typically denoted as $F \in D_{\min}(G_\beta)$ when normalising constants $b_n > 0$ and $a_n \in \mathfrak{R}$, $n \in N$ exist and follow the specified expression for all x .

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{1:n} - a_n}{b_n} \leq x\right) = \lim_{n \rightarrow \infty} (1 - (1 - F(b_n x + a_n))^n) = G_B(x) \quad 1$$

or equivalently

$$b_n^{-1}(X_{1:n} - a_n) \xrightarrow[n \rightarrow \infty]{d} G_\beta(x) \Leftrightarrow F \in D(RGEVD) \quad 2$$

for all x where $G_B(x)$ here is the Reversed generalized extreme value distribution (RGEVD) or the von Mises-Jenkinson form and it can be written as:

$$G_B(x) = (RGEVD) = \begin{cases} \frac{1}{\sigma} [1 + \beta(\frac{x-\mu}{\sigma})]^{-\frac{1}{\beta}-1} \exp\{-[1 + \beta(\frac{x-\mu}{\sigma})]^{-\frac{1}{\beta}}\} & \beta \neq 0 \\ \frac{1}{\sigma} \exp(\frac{x-\mu}{\sigma}) \exp\{-\exp[\frac{x-\mu}{\sigma}]\} & \beta = 0 \end{cases} \quad 3$$

and the cumulative density function (cdf) is given by:

$$F(x) = \begin{cases} 1 - \exp\{-[1 + \beta(\frac{x-\mu}{\sigma})]^{-\frac{1}{\beta}}\} & \beta \neq 0 \\ 1 - \exp\{\exp[-\frac{x-\mu}{\sigma}]\} & \beta = 0 \end{cases} \quad 4$$

If the conditions in Eq. 1 or 2 are met, we may indicate that CDF F resides in the min DA of $G_B(x)$ in (3) and is typically denoted as $F \in D_{\min}$. In extreme value analysis (EVA), the real parameter β is the chief parameter and is designated as the extreme value index (EVI). This EVI (β) ascertains how the right tail of F is shaped. The EVD, as specified in (3), is typically categorized into three variants: the first comprises CDFs with a small tail comprising a finite right threshold, such as the beta CDFs, that are specified as the Weibull ($\beta < 0$), e.g., beta and uniform distributions. The second case comprises the Gumbel ($\beta = 0$) situation that may be used for several applied science scenarios. It comprises numerous CDFs that have an exponential tail, such as exponential, gamma, and normal. Finally, Fréchet DA ($\beta > 0$) is the third scenario. It comprises heavy-tailed CDFs, such as Student-t, Cauchy, loggamma, and the Pareto Burr distributions. The problem of finding the extreme values of the distributions has been solved by [2], completed by [4] and later by [6]. They

demonstrate that, if Eq. 1 or 2 hold, the limiting distribution $G_B(x)$ must be one of just three types, i.e., Weibull, Gumbel, or Fréchet.

3 Reversed Fréchet distribution (RFD)

For the last few decades, the EVT domain has relied on the reversed Fréchet distribution (RFD) because it allows the modeling of several wide-ranging scenarios. It has been used for technological work and natural extremes like rainfall, floods, temperature, air pollution, and wind speed. The RFD is a unique scenario of RGEVD, where the shape (β) variation $\beta > 0$. Such distributions are also called Weibull or type II distributions. The Fréchet type is sometimes referred to as the Cauchy-Fréchet type. The three parameters of RFD were considered, namely, β , location (μ), and scale (σ), as well as the RF's ($D_{\min}(RF)$) probability density function (*pdf*), which is expressed as:

$$f(x) = \frac{\beta\sigma}{(\lambda - x)^2} \exp\left[-\left(\frac{\sigma}{\lambda - x}\right)^\beta\right] \left(\frac{\sigma}{\lambda - x}\right)^{\beta-1}, x < \lambda \quad 5$$

and has the cumulative distribution function (*cdf*):

$$F(x) = 1 - \exp\left[-\left(\frac{\sigma}{\lambda - x}\right)^\beta\right], x < \lambda \quad 6$$

Where (β, μ, σ) are shape, location and scale parameter respectively. For some properties and applications, see [20-22].

4 Determining the Domain Attraction under Linear Normalizing

This section discusses two methods and the necessary and sufficient conditions for belonging to the domain of attraction of the reversed Fréchet distribution (DARFD) under linear normalizing. However, first, some of the principles for the DA to belong to the RFD are discussed.

4.1 Domain Attraction of Reversed Fréchet distribution (DARFD)

An interesting problem, from the point of view of extremes, is knowing the DA of a given *cdf* F . To identify the DA of a given d.f. F and the associated location (a_n) and scale (b_n) sequences, [6, 9] have proposed the following two theorems.

Theorem 1 (Sufficient conditions of Castillo and Hadi for $F \in D_{\min}(RF)$).

For the continuous *cdf* F to belong to the sufficient conditions on the density of a distribution, it must belong to the DARFD. This is expressed as $F \in D_{\min}(RF)$ if and only if:

$$\lim_{e \rightarrow 0} \frac{Q(e) - Q(2e)}{Q(2e) - Q(4e)} = 2^{-\beta}, \beta < 0 \quad 7$$

where, e is the base of the natural logarithm, $Q = F^{\leftarrow}(y) = \inf\{x : F(x) > y\}$ and β is the shape parameter of the associated limit RFD. The *cdf* satisfying Eq. 7 is called the DARFD.

Theorem-2: (Sufficient conditions of de Haan and Ferreira for $F \in D_{\min}(RF)$)

A *cdf* F is in the DARFD if $\beta > 0$ and is expressed as $F \in D_{\min}(RF)$ if and only if

$$\lim_{t \rightarrow \infty} \frac{u(\infty) - u(tx)}{u(\infty) - u(t)} = \frac{F^{\leftarrow}(1) - F^{\leftarrow}(1-1/tx)}{F^{\leftarrow}(1) - F^{\leftarrow}(1-1/t)} = x^\beta \quad 8$$

$$\lim_{t \rightarrow \infty} \frac{u(tx)}{u(t)} = \frac{F^{\leftarrow}(1-1/tx)}{F^{\leftarrow}(1-1/t)} = x^\beta \quad 9$$

$$\lim_{t \rightarrow \infty} \frac{tu'(t)}{u(t)} = \frac{t \left(F^{\leftarrow}(1-1/t)\right)'}{F^{\leftarrow}(1-1/t)} = \beta \quad 10$$

where, u is the quantile function of the tail and expressed as $U(y) = F^{\leftarrow}(1-1/y)$. A d.f. F is considered a part of the DARFD and is expressed as $F \in D_{\min}(RF)$. A simple proof of this result is given in [9]. The d.f. F , which is considered a part of the DARFD, is called a heavy-tailed distribution. Lastly, a_n and b_n are potential options for norming

constants. These constants are not exclusive and rely on the sort of DA. The μ and σ can be chosen in the DARFD as follows:

4.2 Norming Constant of Domain Attraction of RF distribution

Under the conditions of $F \in D_{\min}(RF)$, the a_n and b_n norming constants are defined as follows:

$$a_n = 0, b_n = |Q(n^{-1})| \quad 11$$

For more information on the a_n and b_n norming constants, see [17].

4.3 Examples of Theoretical Applications

Examples-1: Let F be the cdf of the Cauchy distribution is written as:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} x, -\infty < x < \infty$$

and its inverse (F^{\leftarrow}) or quantile (Q) function is:

$$x_p = F^{\leftarrow}(p) = \tan[(p - 1/2)\pi]$$

Direct calculations applied to the condition presented in Eq-7 show that the limiting Cauchy distribution as follows:

$$\lim_{e \rightarrow 0} \frac{\tan[(e - 0.5)\pi] - \tan[(2e - 0.5)\pi]}{\tan[(2e - 0.5)\pi] - \tan[(4e - 0.5)\pi]} = \lim_{e \rightarrow 0} \frac{-1/(2e)}{-1/(4e)} = 2^1 \Rightarrow \beta = 1$$

which shows that the Cauchy distribution belongs to the Reversed Fréchet domain of attraction $F \in D_{\min}(RF)$. The limiting distribution for the minima in this case is the RF distribution and by symmetry the same conclusion holds for the maxima. Also we will try to check whether the de Haan sufficient condition (Eq-8) holds. For that purpose, let us note that:

$$\lim_{t \rightarrow \infty} \frac{u(\infty) - u(tx)}{u(\infty) - u(t)} = \frac{F^{\leftarrow}(1) - F^{\leftarrow}(1 - 1/tx)}{F^{\leftarrow}(1) - F^{\leftarrow}(1 - 1/t)} = x^\beta$$

$$x_p = F^{\leftarrow}(p) = \tan[(p - 1/2)\pi]$$

$$F^{\leftarrow}(1) = \tan[(1 - 1/2)\pi] = \tan[(1/2)\pi] =$$

$$F^{\leftarrow}(1 - 1/t) = \tan[(1 - 1/t - 1/2)\pi] = \tan[(1/2 - 1/t)\pi] =$$

$$F^{\leftarrow}(1 - 1/tx) = \tan[(1 - 1/tx - 1/2)\pi] = \tan[(1/2 - 1/tx)\pi] =$$

$$\lim_{t \rightarrow \infty} \frac{\tan[(1/2)\pi] - \tan[(1/2 - 1/tx)\pi]}{\tan[(1/2)\pi] - \tan[(1/2 - 1/t)\pi]} =$$

This means that the condition of de Haan hold and we conclude that $F \in D_{\min}(W_m)$ When the domain of attraction condition is satisfied. A possible selection of the constants, according to (11), is

$$a_n = 0, b_n = \tan[\pi(\frac{1}{2} - \frac{1}{n})]$$

5 Parameter Estimation

In this section, we describe the maximum likelihood estimators (MLE) for estimating the parameters of RF distribution are considered in case of location known. If x_1, x_2, \dots, x_n is a random sample from the RF distribution, then the likelihood function corresponding to this sample is given by:

$$L = \prod_{i=1}^n f(x_i; \hat{\mu}, \hat{\sigma}, \hat{\beta}) = \prod_{i=1}^n \beta \alpha x_i^{-(\beta+1)} \exp[-\alpha x_i^{-\beta}] = \beta^n \sigma^n (\prod_{i=1}^n x_i^{-(\beta+1)}) \exp[-\sigma \sum_{i=1}^n x_i^{-\beta}] \quad 12$$

On taking the logarithms of (12),

$$\log L = n \log \beta + n \log \sigma - (\beta + 1) \sum_{i=1}^n \log x_i - \sigma \sum_{i=1}^n x_i^{-\beta} \quad 13$$

differentiating (Eq-13) with respect to σ and β respectively and equating to zero, (from $\frac{\partial \ln L}{\partial \sigma} = 0$ and $\frac{\partial \ln L}{\partial \beta} = 0$)

we get the likelihood equations:

$$\frac{n}{\sigma} + \sum_{i=1}^n x_i^{-\beta} = 0 \quad 14$$

And

$$\frac{n}{\beta} - \sum_{i=1}^n \log x_i + \sigma \sum_{i=1}^n x_i^{-\beta} \log x_i = 0 \quad 15$$

Whose solutions provide $\hat{\sigma}_{MLE}$ and $\hat{\beta}_{MLE}$. After some algebraic manipulations, the estimate $\hat{\beta}_{MLE}$ can be obtained by solving the following non-linear equation:

$$\frac{n}{\beta} - \sum_{i=1}^n \log x_i + \frac{n \sum_{i=1}^n x_i^{-\beta} \log x_i}{\sum_{i=1}^n x_i^{-\beta}} = 0 \quad 16$$

The estimate $\hat{\sigma}_{MLE}$ can be obtained by substituting $\hat{\beta}_{MLE}$ in

$$\hat{\sigma}_{MLE} = \frac{n}{\sum_{i=1}^n x_i^{-\beta}} \quad 17$$

This can be accomplished by the use of standard iterative procedures (i.e., Newton-Raphson method), for more details, see, [18].

6 Diagnostic plots

Graphical nature methods are usually used to evaluate whether a distribution, such as F , behaves in accordance with the Reversed Fréchet (RF) distribution. The RF distribution is notable for its unlimited upper bound and heavy tail features. These graphical methods are crucial tools in extreme value analysis, providing a way to visualize the behavior of the distribution's tail. They offer valuable, immediate insights that can direct the model selection process. Importantly, many commonly-used evaluation techniques in extreme value theory (EVT) are directly obtained from these graphical tools. By using them, we can more precisely decide on the most appropriate model to describe the underlying data. In this section, we explore some of the most efficient graphical diagnostic plots, which help evaluate whether the chosen model fits the used dataset appropriately.

6.1 Probability Paper Plot of Reversed Fréchet (PPPRF)

One of the most frequently used graphical techniques in statistical analysis is the Probability Paper Plot (PPP). The basic concept behind PPP, especially for a bi-parameter family of distributions, is to transform the Cumulative Distribution Function (CDF) to a family of straight lines when plotted. This linearization process allows for an easier evaluation of how well the data aligns with selected distribution. However, when the PPP is used for the Reversed Fréchet distribution. In order to produce a meaningful probability plot, we target a suitable transformation that reformulates the original equation, presenting the data in a way that discloses whether or not it obeys the expected behavior of the RF distribution. This graphical visualization tool simplifies the analysis of multifaceted data, especially when it comes to identifying the distribution's tail characteristics, which is vital in extreme value theory.

$$P = F(x; \mu, \sigma, \beta) \quad 18$$

in the form of a straight line,

$$g(P) = g(F(x; \alpha_0, \alpha_1)) = \alpha_0 h(x) + \alpha_1 \quad 19$$

or equivalently,

$$y = \alpha_0 u + \alpha_1 \quad 20 .$$

Now, we derive the probability paper plots of Reversed Fréchet (PPPRF), for more details, see,[19]. When the cdf of RF is given in (5) can be written as:

$$P = F(x; \mu, \sigma, \beta) = 1 - \exp\left[-((\sigma/(\lambda - x))^\beta)\right] \quad 20$$

and taking logarithms twice we get:

$$-\log[-\log(P)] = -\beta \log(\lambda - x) + \beta \log(\sigma) \quad 21$$

In the form of straight line, we get $y = g(p) = \log[-\log(1 - P)]$, $u = h(x) = -\log(\lambda - x)$, $\alpha_0 = \beta$ and $\alpha_1 = \beta \log(\sigma)$. which

shows that the family of straight lines is written by:

$$y = -\beta \log(\lambda - x) + \beta \log(\sigma) \quad 22$$

The scatter plot of y_i , versus u_i $i = 1, 2, \dots, n$, where $y = g(p) = \log[-\log(1 - P)]$ and $u = h(x) = -\log(\lambda - x)$ are called the PPPRF. In this way, when the trend as approximately linear is an indication of the sample coming from the corresponding

family (RFPPP). If the deviation from the straight line is too strong we conclude that the sample comes from a different distribution, for more details, see [19].

6.2 Return Level Plot (RLP)

The fitted distribution may be used to forecast the frequency of the extreme quantiles within a specified return level (RL). The theory indicates that, on average, the observations may be equal to or greater than the return value for every time period (T) (having a likelihood of $1/T$). The return value may be ascertained by solving for the following expression (i.e., by inverting Eq. 5 in RF) and the quantile RF is specified as:

$$X_p = \mu - \sigma(-\log(1 - p))^{1/\beta} \quad 23$$

The maximum likelihood estimation (MLE) technique is employed to predict the values of these variables. The parameters (β, μ, σ) are substituted by the estimated values $(\hat{\beta}, \hat{\mu}, \hat{\sigma})$ to obtain \hat{X}_p . The RLP indicates the $(\log y_p, x_p)$ plot. Moreover, confidence intervals are typically specified to enhance the information provided by the graph (refer to [12, 17]). The DARFD has concave properties

6.3 Return Levels Estimation (RLE)

It is critical to assess the RL pertaining to extreme events using shorter-term data, for example, the extremes that occur in 10-, 20-, 50-, 100-year or more extended periods; refer [12-17]. The T-year is when the event occurs; RL may be determined simply after estimating the model variables. For instance, consider $\hat{\sigma}$ and $\hat{\beta}$ as the CDF parameters of RF determined using MLE. Based on Eq. 5, the following expression is derived:

$$x_p = \hat{\sigma}[\ln \frac{1}{f(x_i; \hat{\sigma}, \hat{\beta})}]^{1/\hat{\beta}} \quad 24$$

and substituting $f(x_i; \hat{\mu}, \hat{\sigma}, \hat{\beta}) = 1 - 1/T$ from the RF distribution we get:

$$x_p = \hat{\sigma}[\ln \frac{T}{T-1}]^{1/\hat{\beta}} \quad 25$$

Where T denotes the return duration and x indicates the conceptual RL for a specified period. The usual terminology designates x_p as the RL linked with the return duration $1/T$ because we may expect with a reasonable predictability level that the x_p level may be breached once every $1/T$ years. A plot of x_p and $\log y_p$ (or, equivalently, y_p on a logarithmic scale plotted against x_p) is linear for $\beta = 0$. The plot has a convex shape for $\beta < 0$, whereas the plot is concave with indeterminable bounds for $\beta > 0$. The plot represents an RLP, which is specifically helpful for presenting and verifying the model.

6.4 Quantile Quantile Plot (Q-Q Plot)

A visual assessment of the QQ graphs based on the β of the extreme event distributions also helped validate if the datasets were consistent with specified distributions (had fat tails). Distinguishing between heavy, ordinary, and light tail scenarios helps ascertain the distributions used for modelling such events. For this graphical plot, we mark the quantiles of one *d.f.* against another (refer to [8-19]). The quantile function Q is the generalized inverse expression (5) pertaining to the *cdf* F :

$$Q(p) = F^{\leftarrow}(p), \text{ for } p \in (0,1) \quad 26$$

$x_p = F^{\leftarrow}(p)$ is a quantity denoting the p^{th} quantile of the distribution. Suppose that x_1, x_2, \dots, x_n is a sample selected from a specified set and \hat{F} estimates F . In this case, the points may be represented by the scatter plot specified using the following expression:

$$(\hat{F}^{-1}(p_{in}), x_{in}, i = 1, 2, \dots, n) \quad 27$$

This represents a Q-Q plot. A great fit between the model and data is indicated by the Q-Q plot points appearing as a diagonal that follows the distribution. It must be noted that the points comprising a Q-Q plot are within the square $[\hat{F}^{-1}(p_{1n}), \hat{F}^{-1}(p_{nn})] \times [x_{1n}, x_{nn}]$.

7 Hypothesis Testing for the Reversed Fréchet Model

To test the data from the *d.f.* of RF model defined in Eq. 4, the hypothesis testing can be written as:

$$H_0 : \beta = 0 \text{ versus } H_1 : \beta > 0 \text{ (Reversed Frechet)}$$

where μ and σ are unknown. Hence, the Gumbel distributions are validated against a different EV RFD for a specified data vector x_1, x_2, \dots, x_n . The likelihood ratio test (LRT) or Devison (Dev) statistic test is:

$$LRT = -2 \log \left(\frac{\prod_{i=1}^n f(x_i; \hat{\mu}, \hat{\sigma}, \hat{\beta})}{\prod_{i=1}^n G(x_i; \hat{\mu}, \hat{\sigma})} \right) = -2 \log \left(\frac{L_1}{L_0} \right) \approx \chi_{1, 1-\alpha}^2 \quad 28$$

Here, $(\hat{\mu}, \hat{\sigma})$ and $(\hat{\mu}, \hat{\sigma}, \hat{\beta})$ denote the Gumbel frameworks and the MLEs in the EV (i.e., RF), respectively. Since the dimensions of the parameter sets are 2 and 3, it can be said that LR-statistic (LRS) has an asymptotical distribution based on $\chi_{1, 1-\alpha}^2$ having a single degree of freedom based on the null hypothesis (H_0). Therefore, the p-value is specified as:

$$P - \text{value}(LRT) = 1 - \chi_1^2(LRT) \quad 29$$

The significance level is attained with a higher accuracy by employing the Bartlett correction when the LR-statistic (LRS) is replaced by $LRT / (1 + 2.8/n)$. In this case the *P-value* is:

$$P - \text{value} = 1 - \chi_1^2(LRT) / (1 + 2.8/n) \quad 30$$

There is a general approach to substitute an LRS by $LRT / (1 + 2.8/n)$ to attain better accuracy while approximating χ_1^2 .

8 Data Description

To exemplify the use of several approaches applicable for estimating the minimum distribution in applications. This section is based on practical data sets indicating sulphur dioxide (SO₂) and particulate matter (PM) pollutants measured in 2008 to determine air quality during the 10th Ramadan. The pollutant levels were recorded hourly for all twenty-four-hour periods across the year 2008. To apply the block minima (BM), the first sample was divided in blocks, then, the minimum value in each block was collected and the RFD was fitted. The BM data of the two air pollutants, SO₂ and PM, are depicted in Figs. 1 and 2, respectively. The plots allowed us to identify some inconsistent values when the data was partitioned into of four different time periods (6, 12, 16, and 24 hours).

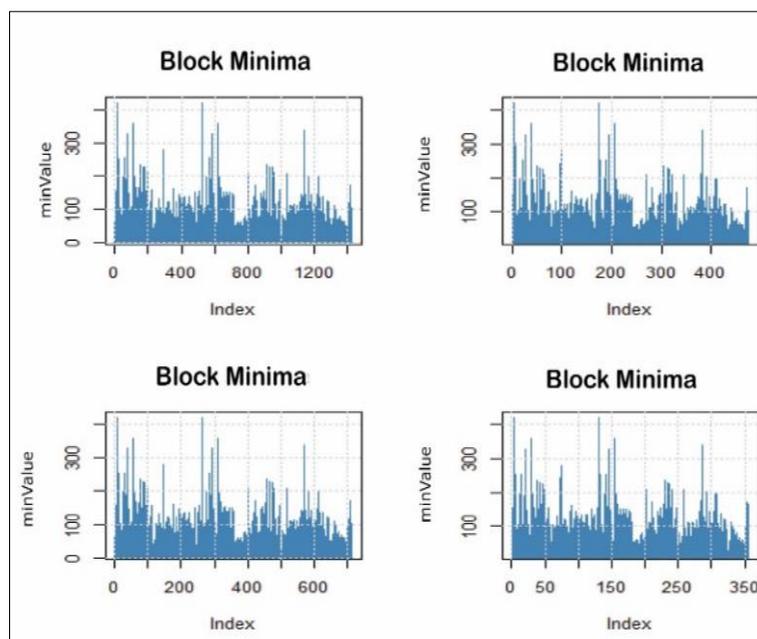


Fig. 1: Plot block minimum of four periods (6, 12, 16 and 24-Hours) from left panel to right respectively of SO2.

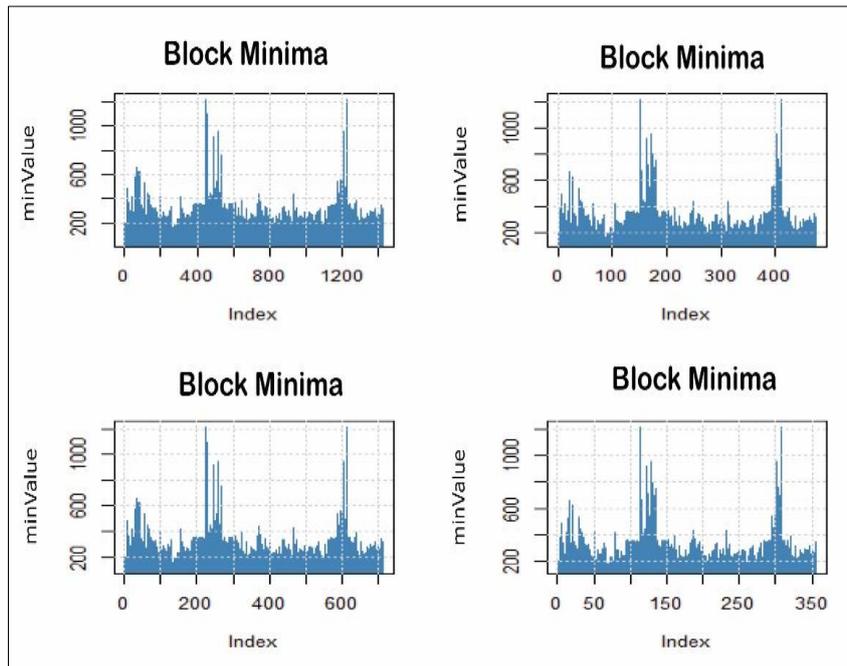


Fig. 2: Plot block minimum of four periods from left panel to right respectively of PM.

8.1 Descriptive Statistics

Firstly, it helps to have an overall understanding of the statistical properties of the specified dataset. Before using Gumbel's approach, it is suitable to gather an understanding of the tail index pertaining to the underlying *d.f. F*. Table-1 summarizes specific statistical characteristics of the two pollutants (SO₂ and PM) measured in the city during the 10th Ramadan. Moreover, Table-1 also indicates that the pollutant data pertaining to four durations exhibits a particular skewness indicative of heavy tails (i.e., $\beta > 0$). Hence, there is a compelling reason for such pollutants to be distributed based on the DARFD (fat tail $\beta > 0$).

Table 1: Descriptive statistics of two pollutants of 10th Ramadan city

Type of pollution	selection period	N 8760	Min	Max	Mean	Median	SD	SK
SO2	6-hours	1423	7.98	422.94	72.563	60.648	46.77	2.28
	12-hours	712	11.172	422.94	88.6019	77.42	52.67	2.13
	18-hours	475	18.088	422.94	101.888	90.052	56.96	2.07
	24-hours	356	18.886	422.94	108.611	95.494	60.22	1.97
PM10	6-hours	1423	49	1219	246.401	221	113.62	2.94
	12-hours	712	109	1219	274.415	247	124.77	3.25
	18-hours	475	128	1219	295.628	261	137.53	3.25
	24-hours	356	143	1219	306.78	273.5	142.31	3.30

8.2 Parameter Estimation

Assume that we have a probability distribution; estimating the parameters of the distribution requires the use of the MLE to evaluate these parameters ($\hat{\mu}, \hat{\sigma}, \hat{\beta}$) for BM considering four durations (6-, 12-, 16-, and 24-hours) for the two specified pollutants. MLE is utilized to fit the BM data and obtain the point estimates to present the results specified in Table-2.

Table 2: ML estimates for parameter of minimum data of two air pollutants

Type of pollution		SO2		PM10	
Period	Parameters	Estimation	Standard error	Estimation	Standard error
6-Hours	Location	50.67	0.83	196.86	2.03
	Scale	27.67	0.66	68.69	1.54
	Shape	0.18	0.02	0.12	0.01

Type of pollution		SO2		PM10	
Period	Parameters	Estimation	Standard error	Estimation	Standard error
12-Hours	Location	64.56	1.40	221.84	2.89
	Scale	33.22	1.07	69.39	2.21
	Shape	0.12	0.03	0.14	0.02
18-Hours	Location	76.02	1.83	237.64	3.44
	Scale	35.73	1.39	67.24	2.73
	Shape	0.13	0.03	0.21	0.03
24-Hours	Location	81.26	2.28	247.61	3.95
	Scale	38.51	1.72	67.11	3.14
	Shape	0.11	0.03	0.22	0.04

It can be seen that the β forecast for every pollutant is positive; hence, the BM distribution aligns with the RFD (right tail, $\beta > 0$). Here, it is clearly helpful to employ specific evaluation processes to ascertain how well the models align with the data. Specific processes are considered in this study.

8.3 Diagnostic Plot

The BM data fitting diagnostic graphs for SO₂ and PM are specified in Figs. 3 to 6 and 7 to 10, respectively. Further, four additional diagnostic graphs: Q-Q-plot, P-P-plot, RLP and density plot pertaining to SO₂ and PM are presented. The graphs provided indicate that the recommended model is a strong fit for the data. Moreover, the Q-Q plot fits well because the line is approximately straight. Furthermore, for every case, the RLP indicates that the RFD (i.e., $\beta > 0$) aligns well with the BM observations pertaining to the two pollutants. Lastly, Table-2 (column 3) indicates that the β is estimated as positive, indicating that the data have RFDs.

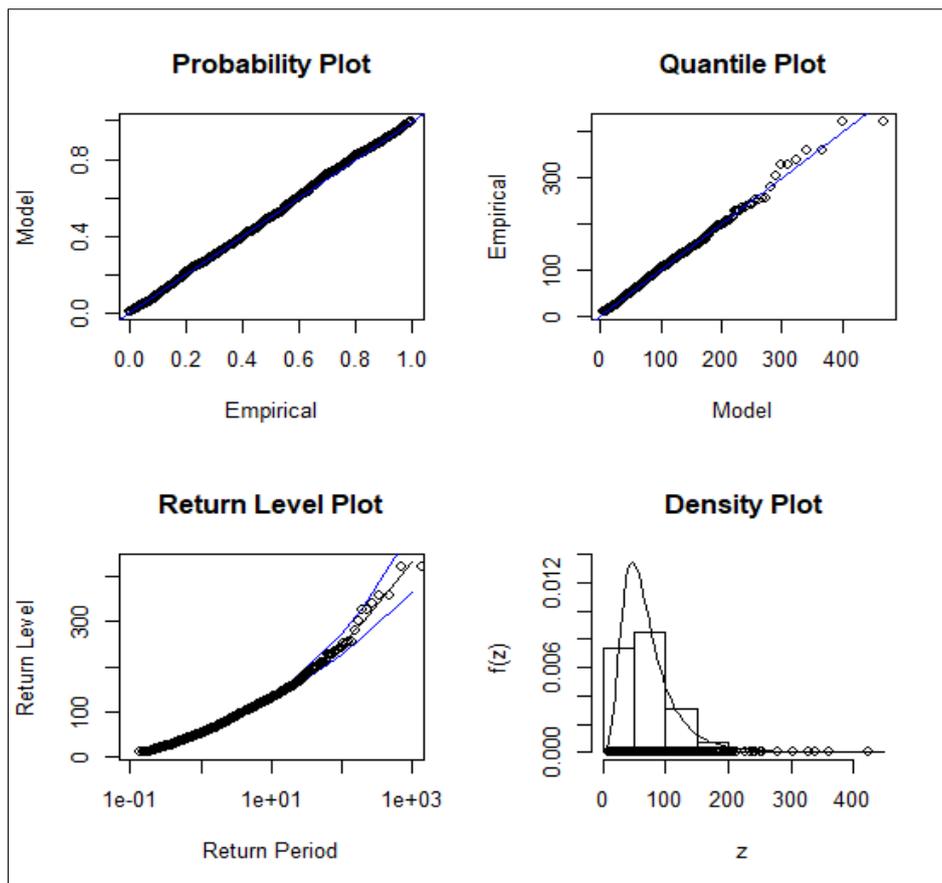


Fig. 3: Diagnostic plot of SO₂ dataset for the RF model fit of 6-Hours.

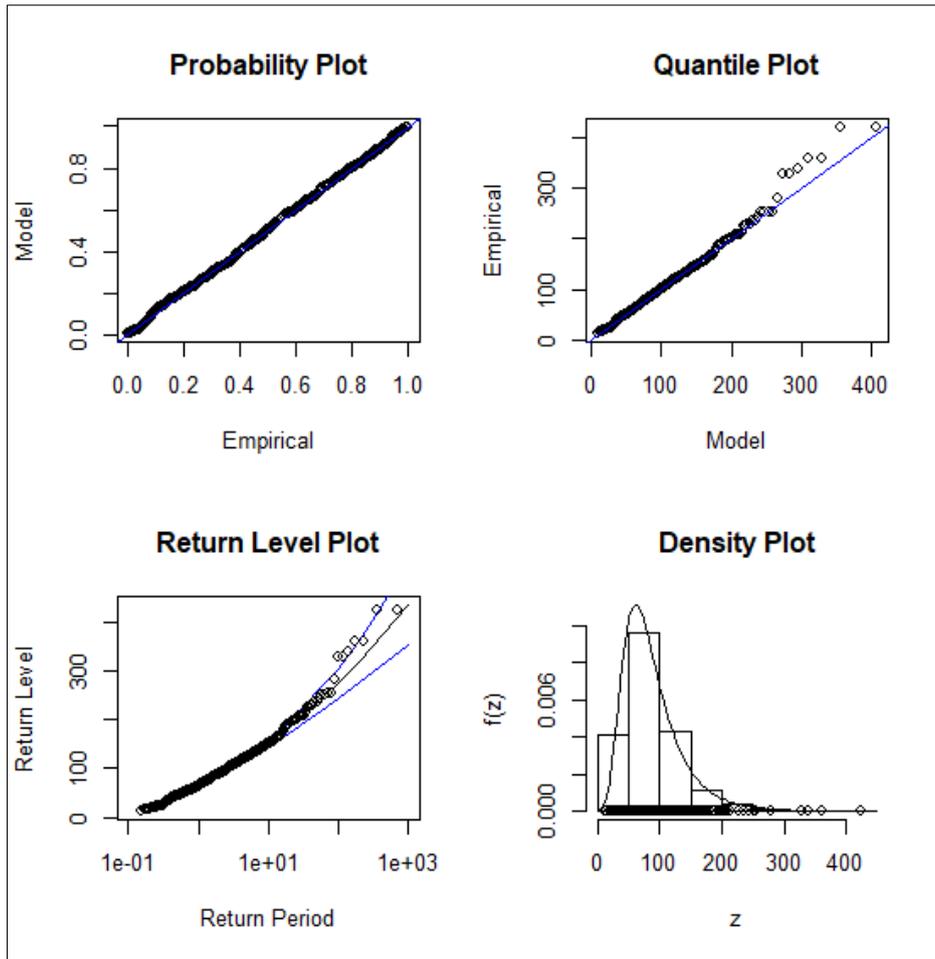


Fig. 4: Diagnostic plot of SO2 dataset for the RF model fit of 12-Hours.

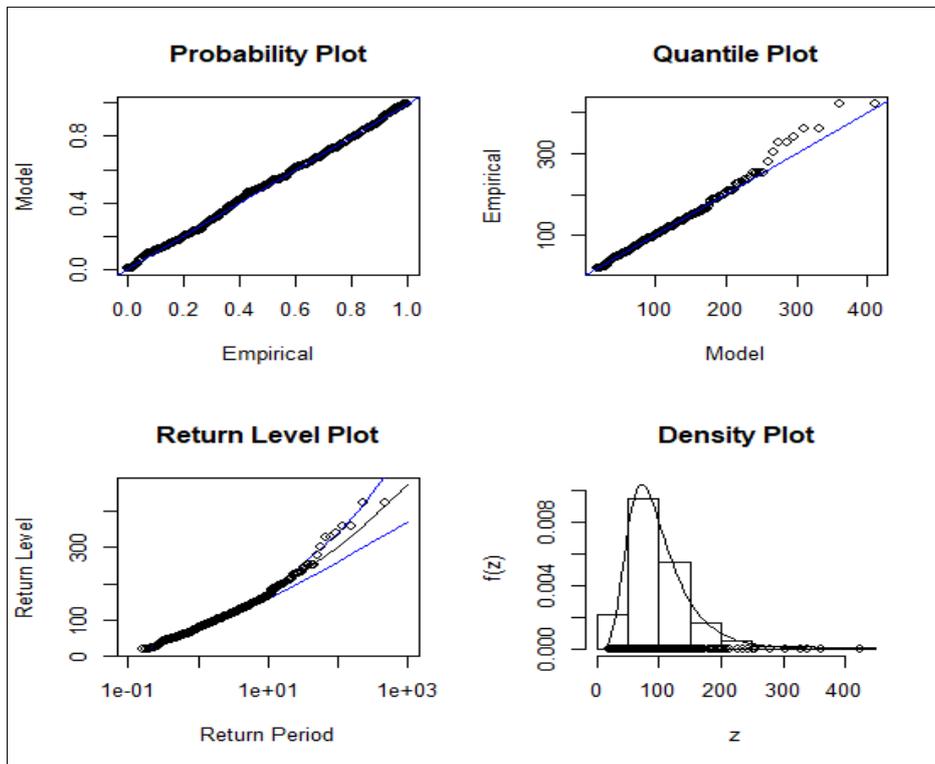


Fig. 5: Diagnostic plot of SO2 dataset for the RF model fit of 18-Hours.

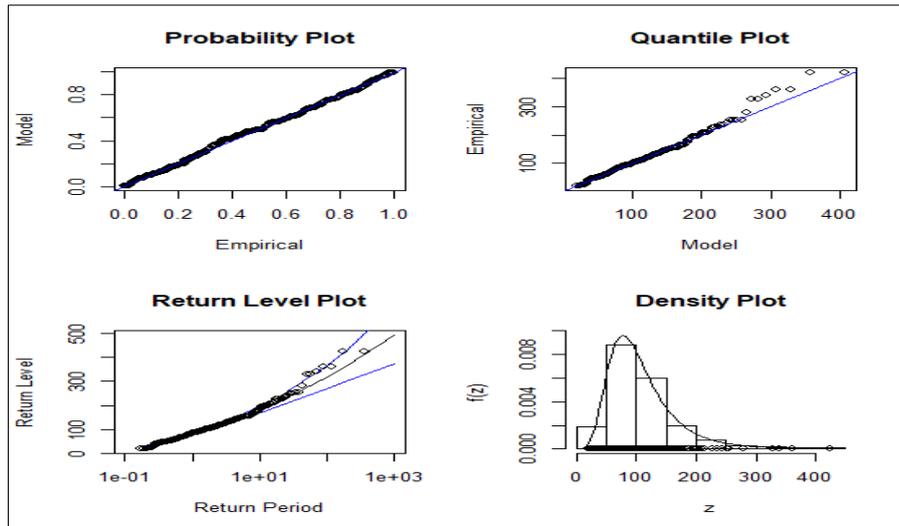


Fig. 6: Diagnostic plot of SO2 dataset for the RF model fit of 24-Hours.

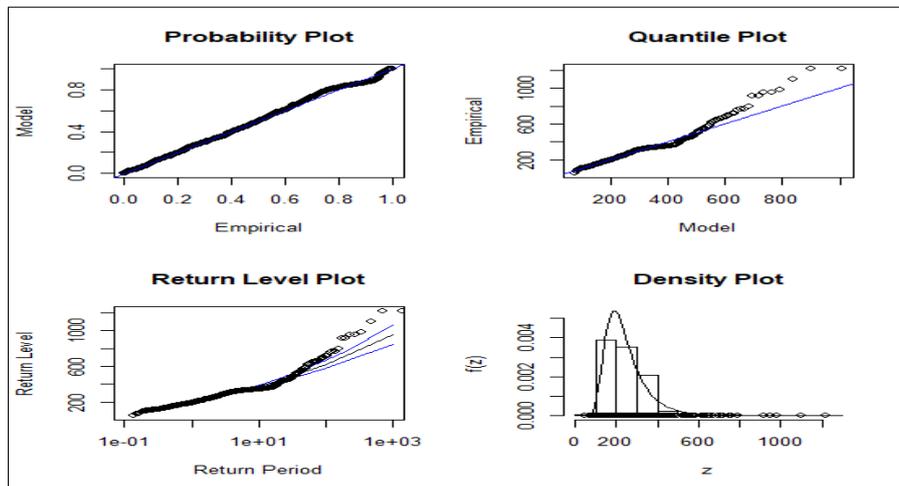


Fig. 7: Diagnostic plot of PM Dataset for the RF model fit of 6-Hours

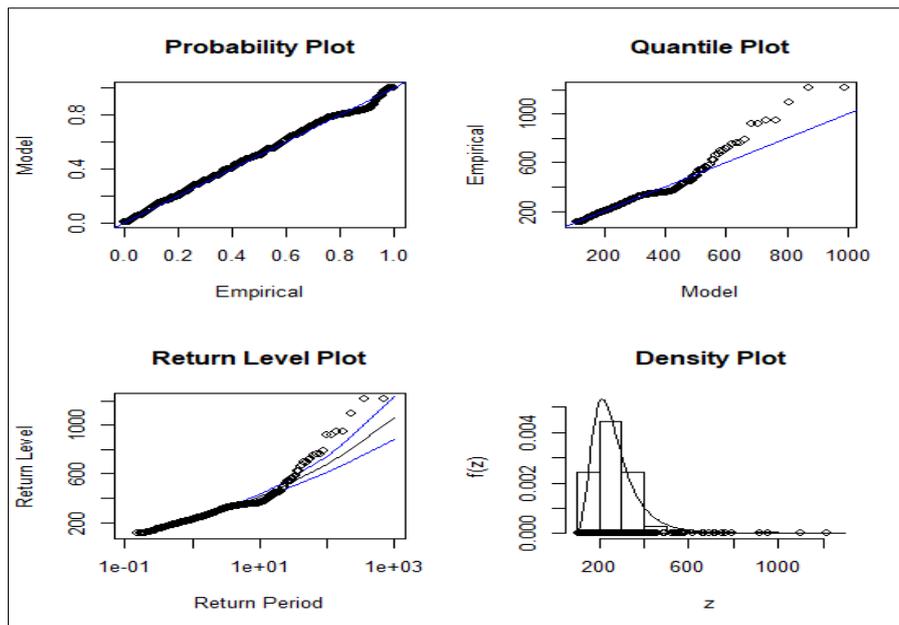


Fig. 8: Diagnostic plot of PM Dataset for the RF model fit of 12-Hours

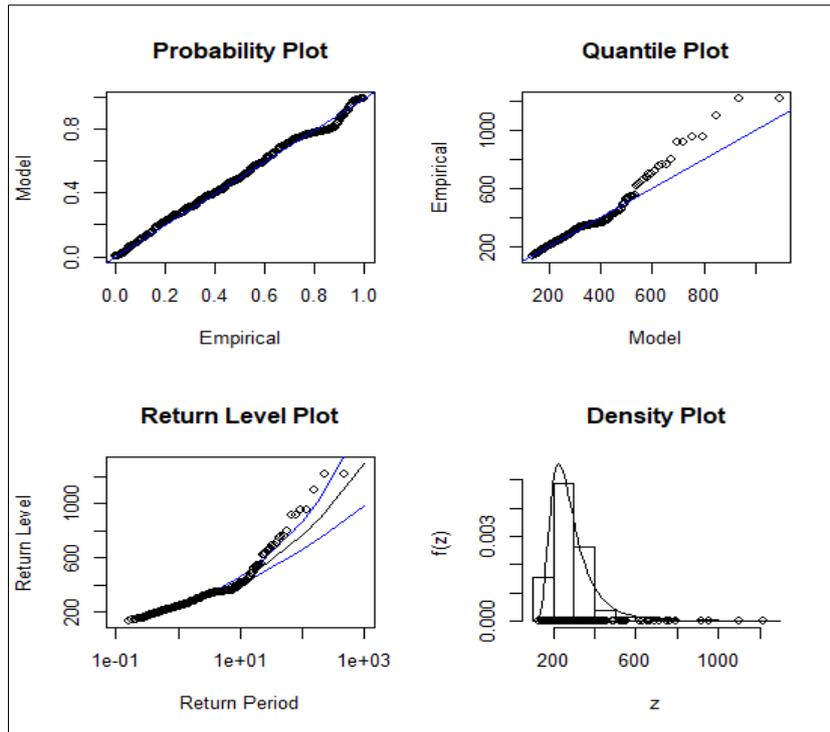


Fig. 9: Diagnostic plot of PM Dataset for the RF model fit of 18-Hours.

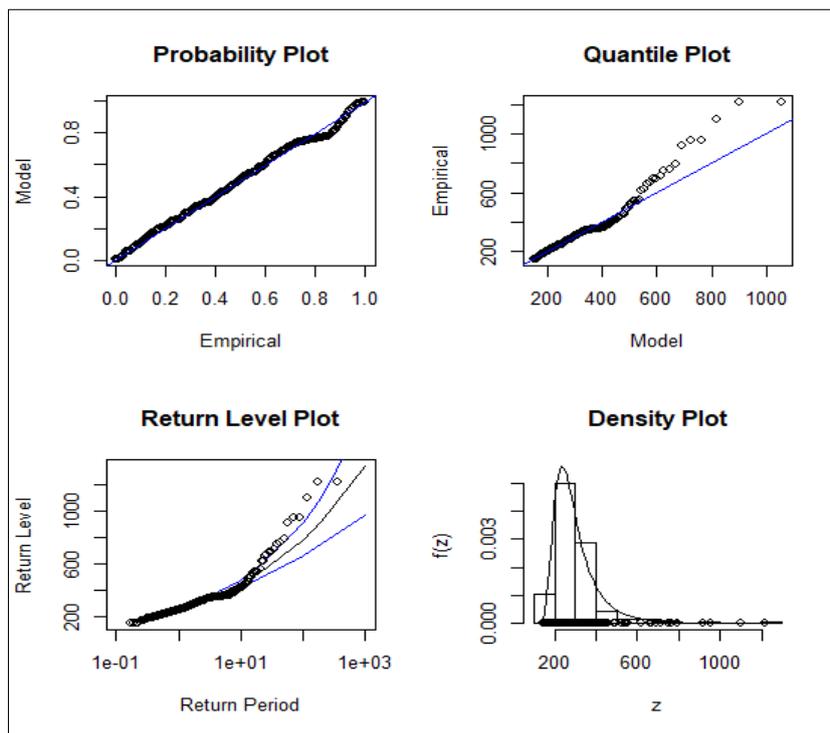


Fig. 10: Diagnostic plot of PM Dataset for the RF model fit of 24-Hours.

8.4 Hypothesis Testing for the Reversed Fréchet Model

Evaluating if the RF framework is applicable to the specified data requires a comparison of the RF and Gumbel families. The LRT is used to evaluate the distribution that fits better. The null hypothesis is stated as "the distribution is Gumbel ($\beta = 0$)", while the alternative hypothesis is stated as "the distribution is RF (i.e., $\beta > 0$)". The null hypothesis is rejected in both cases because the LRS specified in Table-3 is below 1%. Therefore, it is confirmed that the data aligns well with the Fréchet distribution.

Table 3: The LR test of four periods (6,12,16 and 24-Hours) and P-value

Type of Pollution	Period	6-Hours	12-Hours	18-Hours	24-Hours
SO2	Hypothesis	$H_0 : \beta = 0$ $H_1 : \beta > 0$	$H_0 : \beta = 0$ $H_1 : \beta > 0$	$H_0 : \beta = 0$ $H_1 : \beta > 0$	$H_0 : \beta = 0$ $H_1 : \beta > 0$
	Dev	93.04	27.42	19.17	12.12
	P-value	0.00	0.00	0.00	0.00
	Decision	Accept H1	Accept H1	Accept H1	Accept H1
Type of distribution		Fréchet ($\beta > 0$)			
PM	Hypothesis	$H_0 : \beta = 0$ $H_1 : \beta > 0$	$H_0 : \beta = 0$ $H_1 : \beta > 0$	$H_0 : \beta = 0$ $H_1 : \beta > 0$	$H_0 : \beta = 0$ $H_1 : \beta > 0$
	Dev	66.89	63.84	57.26	50.85
	P-value	0.00	0.00	0.00	0.00
	Decision	Accept H1	Accept H1	Accept H1	Accept H1
Type of distribution		Fréchet ($\beta > 0$)			

8.5 Return Levels Estimation (RLE)

Return levels (RL) are critical for estimating and planning. They may be devised to estimate the maximum level of air pollution in an extended future period. The RL is the level that may be reasonably expected to be breached on average once every T periods. For this research, we may not compute the RLs pertaining to the required return durations, i.e., 10, 20, 50, and 100-year. These must be determined for the two pollutants, considering specific selection periods to forecast using the parameters listed in Table-4.

Table 4: Estimated Return level of minimum extreme of two air pollutants

Type of Pollution	Block	Selection period (years)			
		T=10	T=20	T=50	T=100
SO2	6-Hours	127.73	159.87	208.28	250.37
	12-Hours	151.37	184.92	233.26	273.50
	18-Hours	169.31	205.34	257.22	300
	24-Hours	180.61	218.42	272.40	316.96
PM	6-Hours	287.54	336.55	401.98	452.54
	12-Hours	326.53	380.12	452.11	508.05
	18-Hours	346.60	405.45	487.64	554.01
	24-Hours	365.39	429.34	520.17	594.74

Estimates of the RL for the return periods 6, 12, 18, and 24 hours were calculated using 10, 20, 50, and 40 years of data for both the air pollutants. The results show that, for even shorter time series, there is a consistent increase from time to time for the next 100 years for each air pollutants, as with longer time series. In general, the results show that a 100-year event for SO₂ is between 250.37 to 300, while that of PM is between 452.54 to 594.74. The relative uncertainty also increases with time as the high return periods increase.

9. CONCLUSION

The present study investigated some theoretical and practical aspects of using RFD. This distribution is one of the special cases of RGEVD when modelling extreme events under linear normalizing. We applied the BM approach on the data of two air pollutants, namely, SO₂ and PM, and illustrated how EVT can be used to model extreme events using an RF model and MLE to the estimate its parameters. The LRT was used to determine which type of distribution best suited the study's data after it was partitioned into four different time periods, namely, 6, 12, 16, and 24 hours. Post-partitioning, as the selection period increased, the difference between the minimums and maximums increased. Furthermore, the sample means of the different selection periods were larger than the median. This indicates that the β of the distribution was positive. In addition, the skewness of all the selection periods was positive. This indicates that the distribution had a right tail that was relatively longer than the left. Meanwhile, the increasing skewness indicates that the right tail became heavier as the selection period increased. The MLE technique was used to estimate the parameters. It showed the estimates of the β at the four different time periods were positive. This indicates that the distribution of the BM was tail heavy (positive β) and the RFD was appropriate for the data of the two air pollutants. In all cases of the four different time periods, P-P, Q-Q, RLP, and density plots were used to assess the accuracy of the model when fitted to the data of the two air pollutants.

As seen, the scatter of points supported the assumption that the RFD model was better able to describe the behavior of this data. The figure also showed that the skewness of two air pollutants indicate that data is skewed to the right (heavy tail). For the LRT, the H_0 stated that the distribution was Gumbel, while H_1 stated that it was RFD. The P-value of the LRT was smaller than all the significance levels. This indicated that the data of the two air pollutants had an RFD. It also showed that the RF model is a good fit for our data. The diagnostic plot and goodness-of-fit results indicate that modelling using the different selection periods yielded fits that were almost identical. Therefore, the results of the present study indicate that, in the future, the extremes of the two air pollutants will increase from time to time at high return period for the next 100 years. Finally, we suggest some interesting directions for further research. First, it would be useful to deal with a multivariate continuous probabilistic model for the tail air pollutants distribution for other forms of insurance. Secondly, studying limiting distributions for extremes in cases where the observations are sample-dependent may be interesting.

REFERENCES

1. Frechet, M. (1927). Sur la loi de probabilité de l'écart maximum, Ann. SOC. Polon. Math. Cracovie, 6, 93-116.
2. Fisher, R. A., & Tippett, L. H. C. (1928). "Limiting forms of the frequency distribution of the largest or smallest member of a sample," in *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180-190.
3. Von Mises, R. (1936). "Über allgemeine Quadraturformeln.," J. für die reine und Angew. Math, 174, 56-67.
4. Gnedenko, B. (1943). "Sur la distribution limite du terme maximum d'une série aléatoire," Ann. Math, 423-453.
5. Hodges Jr, J. L. (1958). "The significance probability of the Smirnov two-sample test." *Arkiv för matematik*, 3(5), 469-486.
6. Balkema, A. A., & De Haan. L. (1972). "On R. von Mises' condition for the domain of attraction of $\exp(-x)$." *The Annals of Mathematical Statistics*, 1352-1354.
7. Rootzén, H. (1983). "The rate of convergence of extremes of stationary normal sequences." *Advances in Applied Probability*, 15(1), 54-80.
8. Castillo, E., Hadi, A. S., Balakrishnan, N., & Sarabia, J. M. (2005). Extreme value and related models with applications in engineering and science. *Wiley Hoboken, NJ*, P108, P148.
9. Galambos, J. (1994). "Extreme value theory for applications." *Extreme Value Theory and Applications: Proceedings of the Conference on Extreme Value Theory and Applications, Volume 1 Gaithersburg Maryland 1993*. Boston, MA: Springer US.
10. Falk, M., & Frank, M. (1993). "Von Mises conditions revisited." *The Annals of Probability*, 1310-1328.
11. Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. L. (2006). Statistics of extremes: theory and applications. *John Wiley & Sons*, PP56-59.
12. De Haan, L., & Ferreira, A. (2007). Extreme value theory: an introduction. *Springer Science & Business Media*, PP17-21.
13. Omev, E., Mallor, F., & Nualart, E. (2009). "An introduction to statistical modelling of extreme values. Application to calculate extreme wind speeds,".
14. Wuertz, D. (2009). "Package 'fExtremes'." (Missing)
15. Gilleland, E. M. R., & Alec, G. (2013). Stephenson. "A software review for extreme value analysis." *Extremes*, 16, 103-119.
16. Gilleland, M. E. (2018). "Package 'ismev'." (Missing).
17. Coles, S. G. (2001). An Introduction to Statistical Modeling of Extreme Values. New York: Springer, (Vol. 208, p. 208). London: Springer.
18. Bücher, A., & Segers, J. (2016). Supplement to "Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series". DOI:10.3150/16-BEJ903SUPP.
19. Alaswed, H. A., & Alsaidi, M. A. (2018). "Probability Paper and Plotting Position of Extreme Value Distribution for distribution selection and parameter estimation." *Journal of Pure & Applied Sciences*, 17(1).
20. Mansoor, M. (2016). "An Extended Frechet Distribution: Properties and Applications." *Journal of Data Science*, 14(1), 167-188.
21. Yousof, H. M., Emrah, A., & Hamedani. G. G. (2018). "A new extension of Fréchet distribution with regression models, residual analysis and characterization." *Journal of Data Science*, 16(4).
22. Ramos, P. L. (2020). "The Fréchet distribution: Estimation and application-An overview." *Journal of Statistics and Management Systems*, 23(3), 549-578.
23. Deka, D. (2021). "Some properties on Fréchet-Weibull distribution with application to real life data." *Math Stat*, 9(1), 8-15.