Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:100
https://doi.org/10.28919/cmbn/8833
ISSN: 2052-2541

EXTREME VALUE ANALYSIS WITH NEW GENERALIZED EXTREME VALUE DISTRIBUTIONS: A CASE STUDY FOR RISK ANALYSIS ON PM2.5 AND PM10 IN PATHUM THANI, THAILAND

KITTIPONG KLINJAN, TIPAT SOTTIWAN, SIRINAPA ARYUYUEN*

Department of Mathematics and Computer Science, Rajamangala University of Technology Thanyaburi,

Pathum Thani, 12110, Thailand

Abstract: The paper emphasizes the development and application of the power Garima-generalized extreme value distribution for analyzing extreme values of PM2.5 and PM10 in Pathum Thani, Thailand. A new distribution is derived from the power Garima-generated family, and the generalized extreme value distribution provides a continuous framework for modeling extreme events. Additionally, a discrete version of the proposed distribution, namely the discrete power Garima-generalized extreme value distribution, is provided to handle discrete analog data. The maximum likelihood method is used to estimate the parameters when fitting the model to empirical data. The discrete power Garima-generalized extreme value model was utilized in a study to forecast the highest levels of PM2.5 and PM10 (measured in micrograms per cubic meter) for different return periods, including 2, 5, 10, 15, 20, 25, 30, 50, and 100 years. Both PM2.5 and PM10 show increasing return levels as the return period increases. This work's importance lies in its contribution to understanding and predicting extreme PM2.5 and PM10 values, which is critical for meteorologists and policymakers. By providing tools grounded in extreme value theory, the paper supports informed decision-making, planning, and mitigation strategies against the health impacts associated with these particulate matters.

Keywords: extreme value distribution; power Garima-generated family; extreme value analysis; risk analysis.

2020 AMS Subject Classification: 62G32, 62P12, 60E05, 60G70.

---

*Corresponding author

E-mail address: sirinapa_a@rmutt.ac.th

Received August 12, 2024

1. INTRODUCTION

The extreme value theorem is a theory that describes the properties of any random variable that is classified as an "extreme value," which may be the highest or lowest value in any period, and studies the probability distribution of these random variables. Data analysis often ignores extreme values due to their complexity and difficulty. However, we are interested in determining the probability of an event with a high-low value, particularly in heavy-tailed data where the value is very small. Examples of extreme values include the highest or lowest rainfall per day, the highest or lowest values of particulate matter (PM) in a month, the highest or lowest wind speeds in a month, and the highest or lowest temperatures in a day. Extreme value theory is concerned with analyzing and predicting rare and extreme events in random variables. It has wide-ranging applications in many fields, including climate change, coastal engineering, finance, hydrology, meteorology, and insurance. Examples of specific applications of extreme value analysis (EVA) in environmental and climate change can be found, for example, in Zhou et al. [1], Hazarika et al. [2], Pornsopin et al. [3], Warsono et al. [4], Aryuyuen and Bodhisuwan [5], Aguirre-Salado et al. [6], and Tanprayoon et al. [7].

There are generally two approaches to identifying and modeling the extrema of a random process, such as the block-maxima and peak-over-threshold approaches. The block-maxima approach uses the generalized extreme value (GEV) distribution to find and model the extrema of a random process. The peak-over-threshold approach, on the other hand, uses a generalized Pareto distribution to find and model the extrema. The theory derives extreme value Fréchet, Weibull, and Gumbel distributions for the block maxima approach, then develops the GEV distribution within the extreme value theory to combine these distributions [8]. However, by designing a fitted distribution that accurately represents the actual data, we can achieve improved model accuracy. This is especially true when using more versatile distributions that can accommodate a wide range of data types. A multitude of scholars employ novel generalizations to enhance the scope and adaptability of distribution. The application of new generalizations for continuous distributions has gained increased appeal due to its potential to enhance the goodness of fit and ascertain tail features. Many researchers have proposed distributions for extreme values to provide flexibility in describing them. Some examples of extensions to the GEV distribution are the uniform-GEV [9], Topp-Leone GEV [5], Gompertz-GEV [7], and compound GEV [10] distributions.

In this paper, we provide a new distribution for extreme value analysis, which is derived from the power Garima-generated (PGa-G) family of distributions [11] by using the GEV distribution as a baseline distribution. This family of distributions derived from the T-X family of distributions

[12], using the power Garima (PGa) random variable as a generator. The proposed distribution is called the power Garima-generalized extreme value distribution, which is the continuous distribution. Additionally, we propose the discretization of the proposed distribution for the analysis of discrete analog data [13, 14]. In recent decades, researchers have been interested in deriving discrete analogues (discretization) of continuous distributions. Because most of the original life-time data in the real world are continuous, they are discrete values in observation. The parameters of the proposed distribution are estimated by the maximum likelihood method. Finally, the proposed distributions are used for extreme value analysis and risk analysis of the maximum value of the PM that have a diameter of less than 2.5 micrometers in diameter (PM2.5) and the PM that have a diameter of less than 10 micrometers in diameter (PM10) in Pathum Thani, Thailand. Therefore, it is crucial to estimate and anticipate the performance of PM2.5 and PM10 in Pathum Tani. Both PM2.5 and PM10 are respirable and can deposit in various parts of the airways, with the size of the particles determining the specific locations of particle deposition in the lung. The PM2.5 particles have a higher tendency to penetrate and settle in the innermost areas of the lungs, whereas the PM10 particles are more prone to settling on the surfaces of the bigger airways in the upper section of the lungs. Particles that accumulate on the surface of the lungs have the potential to cause harm to the tissue and trigger inflammation in the lungs. Therefore, it is crucial to model and predict the performance of PM2.5 and PM10 in Pathum Tani. This is because both PM2.5 and PM10, when inhaled, can be deposited in various parts of the airways. However, the specific locations of particle deposition in the lung depend on the particle size. The PM2.5 is more likely to accumulate on the deeper lung wall, whereas the PM10 is more likely to accumulate on the upper lung airways. Particles deposited on the lung surface can induce tissue damage and lung inflammation. Meteorologists and policymakers in Thailand need to understand extreme PM2.5 and PM10 patterns and future behaviors for effective decision-making, planning, and mitigation purposes. This article's focus is on extreme value theory, which provides us with relevant tools for modeling and predicting extreme PM2.5 and PM10 in Pathum Thani.

2. PRELIMINARIES

In this section, we provide an overview of the theoretical background regarding the PGa-G family of distributions, the GEV distribution, and the concept of discrete extension of continuous probability distributions.

2.1 The PGa-G family of distributions

Let X be a random variable following the PGa-G family of distributions with the positive parameters a, b and c and a vector parameter $\boldsymbol{\omega}$ denoted by $X \sim$PGa-G$(a,b,c,\boldsymbol{\omega})$. Its cumulative density function (cdf) and probability density function (pdf) are respectively

$$F_{\text{PGa-G}}(x)=1-\left[1+\frac{b}{b+2}\left(\frac{G^{a}(x;\boldsymbol{\omega})}{1-G^{a}(x;\boldsymbol{\omega})}\right)^{c}\right]\exp\left\{-b\left(\frac{G^{a}(x;\boldsymbol{\omega})}{1-G^{a}(x;\boldsymbol{\omega})}\right)^{c}\right\},\ \text{and} \tag{1}$$

$$f_{\text{PGa-G}}(x)=\frac{abc}{b+2}\left[1+b+b\left(\frac{G^{a}(x;\boldsymbol{\omega})}{1-G^{a}(x;\boldsymbol{\omega})}\right)^{c}\right]\exp\left\{-b\left(\frac{G^{a}(x;\boldsymbol{\omega})}{1-G^{a}(x;\boldsymbol{\omega})}\right)^{c}\right\}$$

$$\times g(x;\boldsymbol{\omega})\left[\frac{G^{ac-1}(x;\boldsymbol{\omega})}{\left(1-G^{a}(x;\boldsymbol{\omega})\right)^{c+1}}\right]\quad \text{for}\quad x>0 \tag{2}$$

where $g(x;\boldsymbol{\omega})$ and $G(x;\boldsymbol{\omega})$ are the pdf and cdf of a baseline distribution with a vector parameter $\boldsymbol{\omega}$ [11]. Its quantile function (qf) is

$$Q_{\text{PGa-G}}(u)=G^{-1}\left\{\left[1+\left\{-\frac{1}{b}\left[W_{-1}\left\{-(1-u)(2+b)\exp(-2-b)\right\}+b+2\right]\right\}^{-1/c}\right]^{-1/a}\right\}, \tag{3}$$

for $0<u<1$ and $G^{-1}\{\cdot\}$ represents the inverse cdf of the baseline distribution. When $W_{-1}(\cdot)$ is the lower branch Lambert W function [15], and it is obtained by W function on LambertW package in R [16].

2.2 Extreme value identify with block maxima and the GEV distribution

The block maxima model, introduced by Fisher and Tippett in 1928, is a traditional method used to discover extreme values. The concept of extreme values refers to the identification of the highest values within a given dataset, sometimes referred to as block maxima. The GEV distribution can model these extreme values. The block maxima model, introduced by Fisher and Tippett in 1928, is a traditional method used to discover extreme values. The concept of extreme values refers to the identification of the highest values within a given dataset, sometimes referred to as block maxima.

Let X be a random variable distributed the GEV distribution with the parameters μ, σ, and ξ, denoted by $X\sim\text{GEV}(\mu,\sigma,\xi)$, which was first introduced by [17]. Then its pdf and cdf are respectively

$$g_{\text{GEV}}(x)=\frac{1}{\sigma}\left[\tau(x)\right]^{\xi+1}\exp\left\{-\tau(x)\right\}\ \text{and}\ G_{\text{GEV}}(x)=\exp\left\{-\tau(x)\right\}, \tag{4}$$

where

$$\tau(x)=\begin{cases}\left[1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-1/\xi} & \text{if } \xi\neq 0,\\ \exp\left(\dfrac{x-\mu}{\sigma}\right) & \text{if } \xi=0,\end{cases}\quad \text{and}\quad x\in\begin{cases}[\mu-\sigma/\xi,\infty) & \text{if } \xi>0,\\ (-\infty,\infty) & \text{if } \xi=0,\\ (-\infty,\mu-\sigma/\xi] & \text{if } \xi<0,\end{cases}\tag{5}$$

for $-\infty<\mu<\infty$, $\sigma>0$, and $-\infty<\xi<\infty$. From (4), if $\xi<0$ then the GEV distribution is reduce to the Weibull distribution; if $\xi>0$ then the GEV distribution is reduce to the Fréchet distribution; and the GEV distribution is called the Gumbel distribution when $\xi\to 0$ [18]. Since the quantile function is invertible cdf, thus the qf of the GEV distribution has an explicit expression,

$$Q_{\text{GEV}}(p)=\begin{cases}\mu-\dfrac{\sigma}{\xi}\left\{1-\left[-\log(p)\right]^{-\xi}\right\} & \text{if } \xi\neq 0,\\ \mu-\sigma\log\{-\log(p)\} & \text{if } \xi=0,\end{cases}\tag{6}$$

where $0<p<1$. Based on the extreme value theory that derives from the GEV distribution, we can fit a sample of extremes to the GEV distribution to obtain the parameters that best explain the probability distribution of the extremes [19]. Based on $P(Z>Z_p)=1-p$ where $p=1/T$, the return level (Z_p) at period T from the GEV model is

$$Z_T^{\text{GEV}}=\begin{cases}\mu-\dfrac{\sigma}{\xi}\left[1-\left\{-\log\left(1-\dfrac{1}{T}\right)\right\}^{-\xi}\right] & \text{if } \xi\neq 0,\\ \mu-\sigma\log\left\{-\log\left(1-\dfrac{1}{T}\right)\right\} & \text{if } \xi=0.\end{cases}\tag{7}$$

2.3 Extension of continuous probability distributions

In practical scenarios, even though we might deal with measurements on a continuous scale, the observed values are often discrete due to limitations in measurement precision. Measurements typically record only up to a finite number of decimal places or units, preventing them from representing all possible points in a continuous distribution, leading to this discrete value. Therefore, when modeling such data, using discrete models might be more appropriate and realistic than assuming perfect continuity. This acknowledges the practical constraints of data collection and measurement in real-world scenarios. In many practical situations, continuous variables are often measured in discrete terms due to the precision limitations of measuring instruments or to conserve space. This is especially evident in fields like survival analysis, where continuous variables like survival time are often recorded as discrete counts (e.g., number of days or weeks). These examples demonstrate that continuous lifetimes are frequently represented as discrete random variables, reflecting the practical necessity and convenience of discrete measurement in real-world applications [20]. In this paper, the survival discretization technique is used for extension of continuous probability distributions [12, 13], which a discretization of continuous probability distributions is obtained from the transforming a continuous random variable X with the survival function as $S_X(x) = \mathrm{P}(X > x).$ The probability mass function (pmf) of a discrete random variable Y is

$$f_Y(y) = S_X(y) - S_X(y+1) \quad , \tag{8}$$

where a discrete observed random variable $Y = \lfloor X \rfloor$ is equal to the greatest integer less than or equal to X. The discrete version of the GEV distribution called the discrete GEV (DGEV) distribution with the parameters $\mu,$ $\sigma,$ and $\xi,$ denoted by $X \sim \mathrm{DGEV}(\mu,\sigma,\xi),$ its pmf and cdf are respectively

$$f_{\mathrm{DGEV}}(y) = \begin{cases} \exp\left\{-\left[1+\xi\left(\dfrac{y-\mu+1}{\sigma}\right)\right]^{-1/\xi}\right\} - \exp\left\{-\left[1+\xi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{if } \xi \neq 0, \\ \exp\left\{-\exp\left(\dfrac{y-\mu+1}{\sigma}\right)\right\} - \exp\left\{-\exp\left(\dfrac{y-\mu}{\sigma}\right)\right\} & \text{if } \xi = 0, \end{cases} \tag{9}$$

$$G_{\mathrm{DGEV}}(y) = \begin{cases} \exp\left\{-\left[1+\xi\left(\dfrac{y-\mu+1}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{if } \xi \neq 0, \\ \exp\left\{-\exp\left(\dfrac{y-\mu+1}{\sigma}\right)\right\} & \text{if } \xi = 0. \end{cases} \tag{10}$$

3. MAIN RESULTS

3.1 The power Garima-generalized extreme value distribution

Based on the PGa-G family of distribution proposed by [11] and the GEV distribution as the baseline distribution [17], we have the power Garima-generalized extreme value (PGa-GEV) distribution with parameters $a,b,c,\mu,\sigma,$ and $\xi,$ denoted by $\text{PGa-GEV}(a,b,c,\mu,\sigma,\xi).$ Let

$X \sim \text{PGa-GEV}(a,b,c,\mu,\sigma,\xi),$ then its cdf and pdf can be presented as

$$F_X(x) = 1 - \left[1 + \frac{b}{b+2}\left(\frac{\exp\{-a\tau(x)\}}{1-\exp\{-a\tau(x)\}}\right)^c\right]\exp\left\{-b\left(\frac{\exp\{-a\tau(x)\}}{1-\exp\{-a\tau(x)\}}\right)^c\right\}, \tag{11}$$

$$f_X(x) = \frac{abc}{b+2}\left[1 + b + b\left(\frac{\exp\{-a\tau(x)\}}{1-\exp\{-a\tau(x)\}}\right)^c\right]\exp\left\{-b\left(\frac{\exp\{-a\tau(x)\}}{1-\exp\{-a\tau(x)\}}\right)^c\right\}$$
$$\times \frac{[\tau(x)]^{\xi+1}\exp\{-ac\tau(x)\}}{\sigma[1-\exp\{-a\tau(x)\}]^{c+1}}, \text{ for } -\infty < \mu < \infty \text{ and } -\infty < \xi < \infty, \tag{12}$$

where $\tau(x)$ as equation (5). Scale parameters $a,b,c,$ and σ are the positive value, and μ denotes a location parameter. For the shape parameter $c,$ the PGa-GEV distribution reduces to the Garima-GEV distribution for $c=1.$ The parameter ξ is called a shape parameter and may be used to define three-sub distribution, i.e., the PGa-GEV distribution is reduce to the PGa-Weibull distribution, the PGa-Fréchet distribution, and the PGa-Gumbel distribution when $\xi < 0,$ $\xi > 0$ and $\xi \to 0$ respectively.

The qf of the PGa-GEV distribution is

$$Q_X(u) = \begin{cases} \mu - \dfrac{\sigma}{\xi}\left[1 - \left(-\log\left\{\left(1 + \left[-\dfrac{1}{b}\delta_u\right]^{-1/c}\right)^{-1/a}\right\}\right)^{-\xi}\right] & \text{if } \xi \neq 0, \\ \mu - \sigma\log\left\{-\log\left\{\left(1 + \left[-\dfrac{1}{b}\delta_u\right]^{-1/c}\right)^{-1/a}\right\}\right\} & \text{if } \xi = 0, \end{cases} \tag{13}$$

where $0 < u < 1$ and $\delta_u = W_{-1}\{-(1-u)(2+b)\exp(-2-b)\} + b + 2.$ Its corresponding return level at period T is

$$Z_T^{\text{PGa-GEV}} = \begin{cases} \mu - \dfrac{\sigma}{\xi}\left\{1 - \left[-\log\left(1 + \left\{-\dfrac{1}{b}\delta_T\right\}^{-1/c}\right)^{-1/a}\right]^{-\xi}\right\} & \text{if } \xi \neq 0, \\ \mu - \sigma\log\left\{-\log\left[\left(1 + \left\{-\dfrac{1}{b}\delta_T\right\}^{-1/c}\right)^{-1/a}\right]\right\} & \text{if } \xi = 0, \end{cases} \tag{14}$$

where $\delta_T = W_{-1}\left\{-\frac{1}{T}(2+b)\exp(-2-b)\right\}+b+2.$

Probability density plots of the PGa-GEV distribution are shown in Figures 1-3. For $\xi < 0$, its pdf has a left-skewed shape (see Figure 1). For $\xi > 0$, its pdf has a right skewed shape (see Figure 2). Figure 3 shows the pdf's shape of the PGa-GEV distribution with the specified parameters, and its pdf has various shapes, i.e., symmetric and right-skewed.

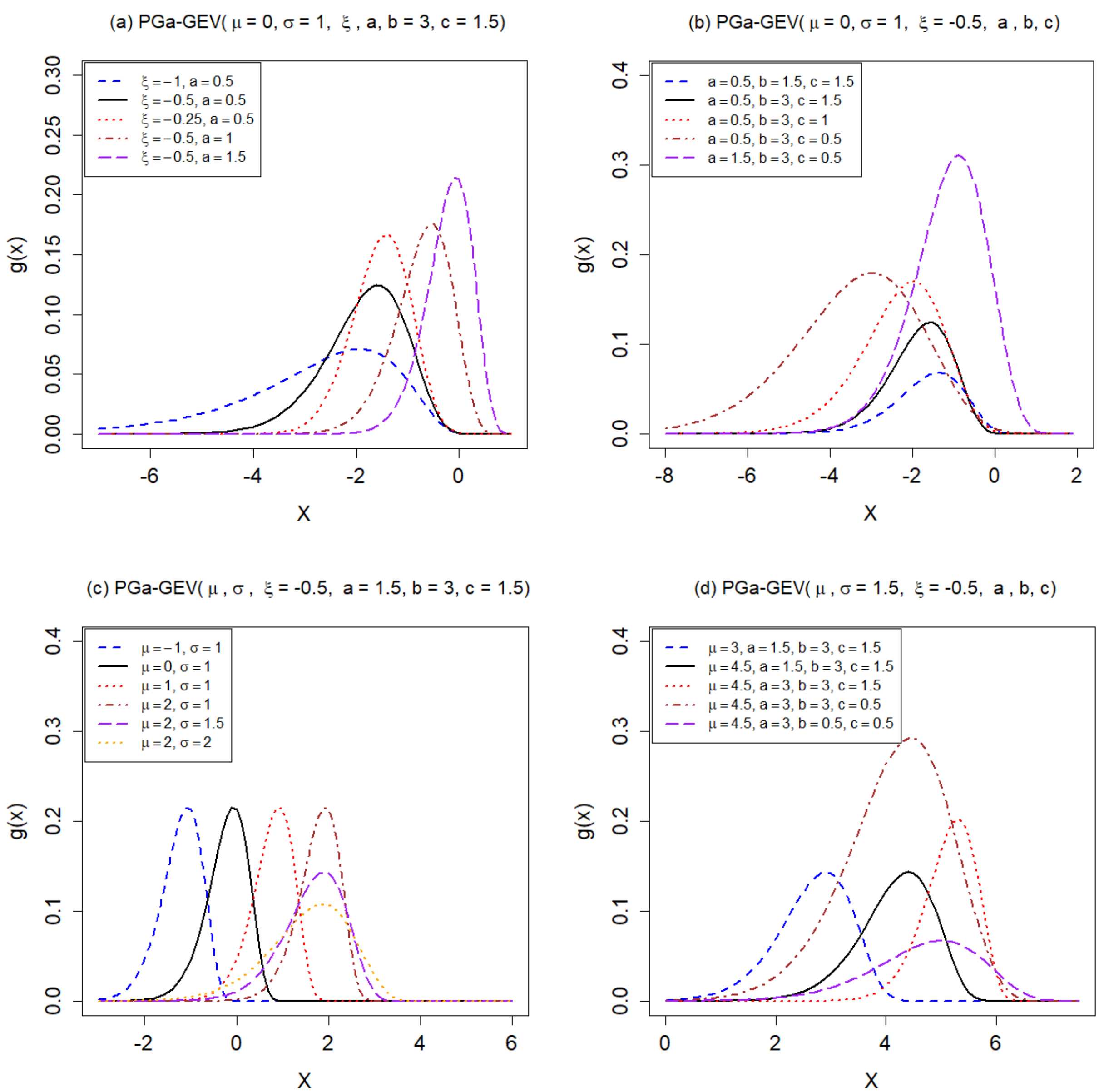

Figure 1. Probability density plots of the PGa-GEV distribution for $\xi < 0$.

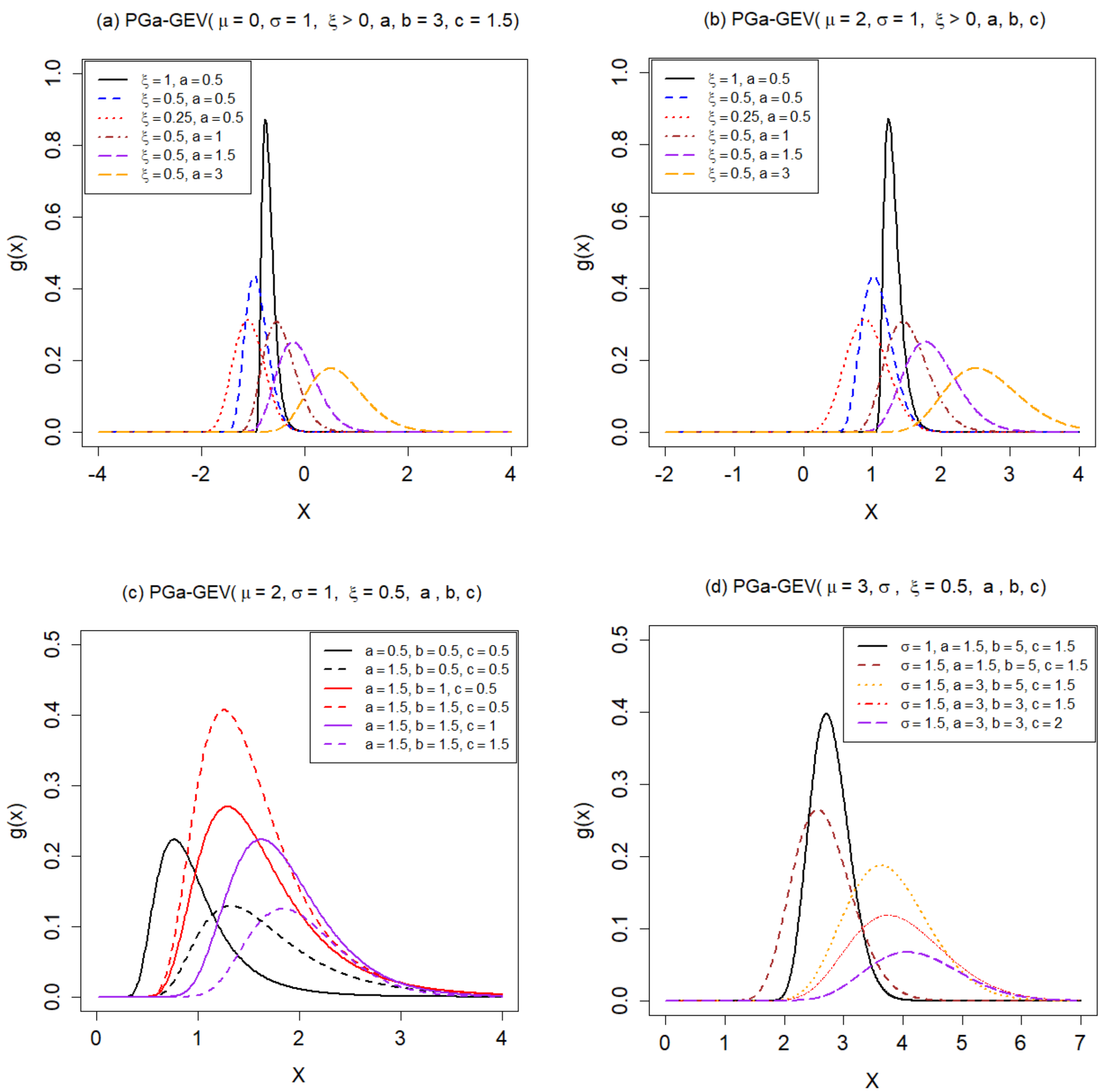

Figure 2. Probability density plots of the PGa-GEV distribution for $\xi > 0$.

3.2 The discrete power Garima-generalized extreme value distribution

Based on the survival discretization technique [12, 13], a discretization of continuous probability distributions is obtained from the transforming a continuous random variable X with the survival function. Let $X \sim$PGa-GEV(a,b,c,μ,σ,ξ) with the cdf in equation (11), and $Y = \lfloor X \rfloor$ is equal to the greatest integer less than or equal to X, then we have the discrete power Garima-generalized extreme value (DPGa-GEV) distribution, denoted by $Y \sim$DPGa-GEV(a,b,c,μ,σ,ξ).

Its cdf and pdf are respectively

$$f_Y(y)=\left[1+\frac{b}{b+2}\left(\frac{\exp\{-a\tau(y)\}}{1-\exp\{-a\tau(y)\}}\right)^c\right]\exp\left\{-b\left(\frac{\exp\{-a\tau(y)\}}{1-\exp\{-a\tau(y)\}}\right)^c\right\}$$
$$-\left[1+\frac{b}{b+2}\left(\frac{\exp\{-a\tau(y+1)\}}{1-\exp\{-a\tau(y+1)\}}\right)^c\right]\exp\left\{-b\left(\frac{\exp\{-a\tau(y+1)\}}{1-\exp\{-a\tau(y+1)\}}\right)^c\right\}, \qquad (15)$$

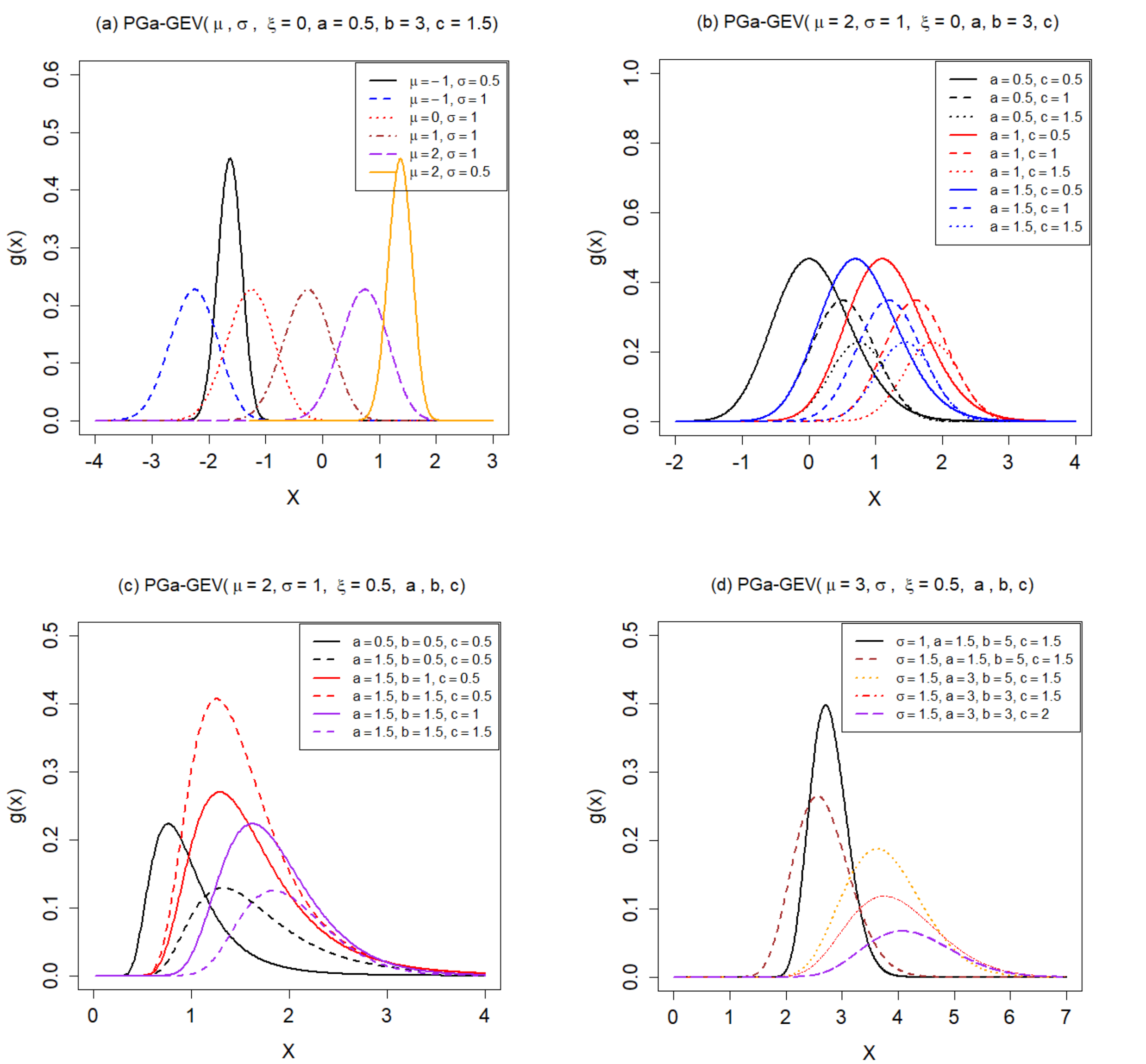

Figure 3. Probability density plots of the PGa-GEV distribution for $\xi = 0$.

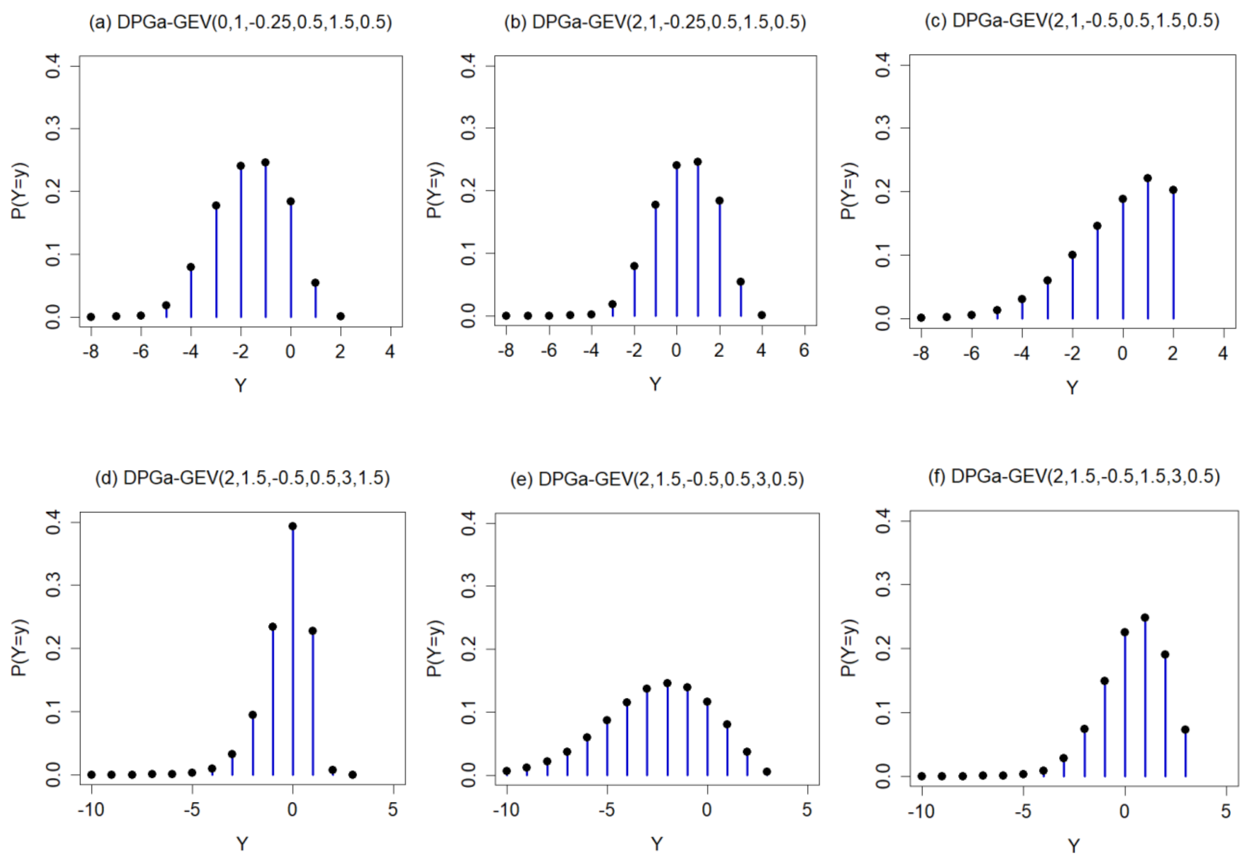

Figure 4. The pmf plots of the DPGa-GEV distribution for $\xi < 0$.

$$F_Y(y) = 1 - \left[1 + \frac{b}{b+2}\left(\frac{\exp\{-a\tau(y+1)\}}{1-\exp\{-a\tau(y+1)\}}\right)^c\right]\exp\left\{-b\left(\frac{\exp\{-a\tau(y+1)\}}{1-\exp\{-a\tau(y+1)\}}\right)^c\right\}, \tag{16}$$

where $\tau(y) = \left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi}$ if $\xi \neq 0$ and $\tau(y) = \exp\left(\frac{y-\mu}{\sigma}\right)$ if $\xi = 0$. The pmf plots of the DPGa-GEV distribution are shown in Figures 4-6. For $\xi < 0$, its pmf has a left-skewed shape (see Figure 1). For $\xi > 0$, its pmf has a right skewed shape (see Figure 5). Figure 6 shows the pmf's shape of Y which has symmetric and right skewed.

Its corresponding qf and return level at period T are respectively

$$Q_Y(u)=\begin{cases}\left[\mu-\dfrac{\sigma}{\xi}\left\{1-\left[-\log\left(1+\left\{-\dfrac{1}{b}\delta_u\right\}^{-1/c}\right)^{-1/a}\right]^{-\xi}-1\right\}-1\right] & \text{if } \xi\neq 0,\\ \left[\mu-\sigma\log\left\{-\log\left[\left(1+\left\{-\dfrac{1}{b}\delta_u\right\}^{-1/c}\right)^{-1/a}\right]\right\}-1\right] & \text{if } \xi=0,\end{cases} \tag{17}$$

$$Z_T^{\text{DPGa-GEV}}=\begin{cases}\left[\mu-\dfrac{\sigma}{\xi}\left\{1-\left[-\log\left(1+\left\{-\dfrac{1}{b}\delta_T\right\}^{-1/c}\right)^{-1/a}\right]^{-\xi}\right\}-1\right] & \text{if } \xi\neq 0,\\ \left[\mu-\sigma\log\left\{-\log\left[\left(1+\left\{-\dfrac{1}{b}\delta_T\right\}^{-1/c}\right)^{-1/a}\right]\right\}-1\right] & \text{if } \xi=0.\end{cases} \tag{18}$$

3.3 Parameter estimation

Several techniques have been suggested for parameter estimation, but the maximum likelihood (ML) method is the most frequently used. ML estimators provide favorable characteristics for the model's parameters. Therefore, the ML method is used to estimate the parameters of the PGa-GEV and DPGa-GEV distributions.

Let X_t be independent and identically distributed (iid) random variables of size n following the pdf (12) with a vector of parameters $\boldsymbol{\theta}=(a,b,c,\mu,\sigma,\xi),$ then the log-likelihood can be written as follows

$$\ell(\boldsymbol{\theta})=n\log a+n\log b+n\log c-n\log(b+2)-b\sum_{t=1}^{n}\left(\frac{\exp\{-a\tau(x_i)\}}{1-\exp\{-a\tau(x_i)\}}\right)^{c}$$
$$+\sum_{t=1}^{n}\log\left[1+b+b\left(\frac{\exp\{-a\tau(x_i)\}}{1-\exp\{-a\tau(x_i)\}}\right)^{c}\right]+\frac{1}{\sigma}\sum_{t=1}^{n}\log\left\{\frac{\tau_{x_i}^{\xi+1}\exp\{-ac\tau(x_i)\}}{\left[1-\exp\{-a\tau(x_i)\}\right]^{c+1}}\right\}$$

The ML estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained numerical from the nonlinear equations

$$\frac{\partial\ell(\boldsymbol{\theta})}{\partial\mu}=\frac{\partial\ell(\boldsymbol{\theta})}{\partial\sigma}=\frac{\partial\ell(\boldsymbol{\theta})}{\partial\xi}=\frac{\partial\ell(\boldsymbol{\theta})}{\partial a}=\frac{\partial\ell(\boldsymbol{\theta})}{\partial b}=\frac{\partial\ell(\boldsymbol{\theta})}{\partial c}=0\,. \tag{19}$$

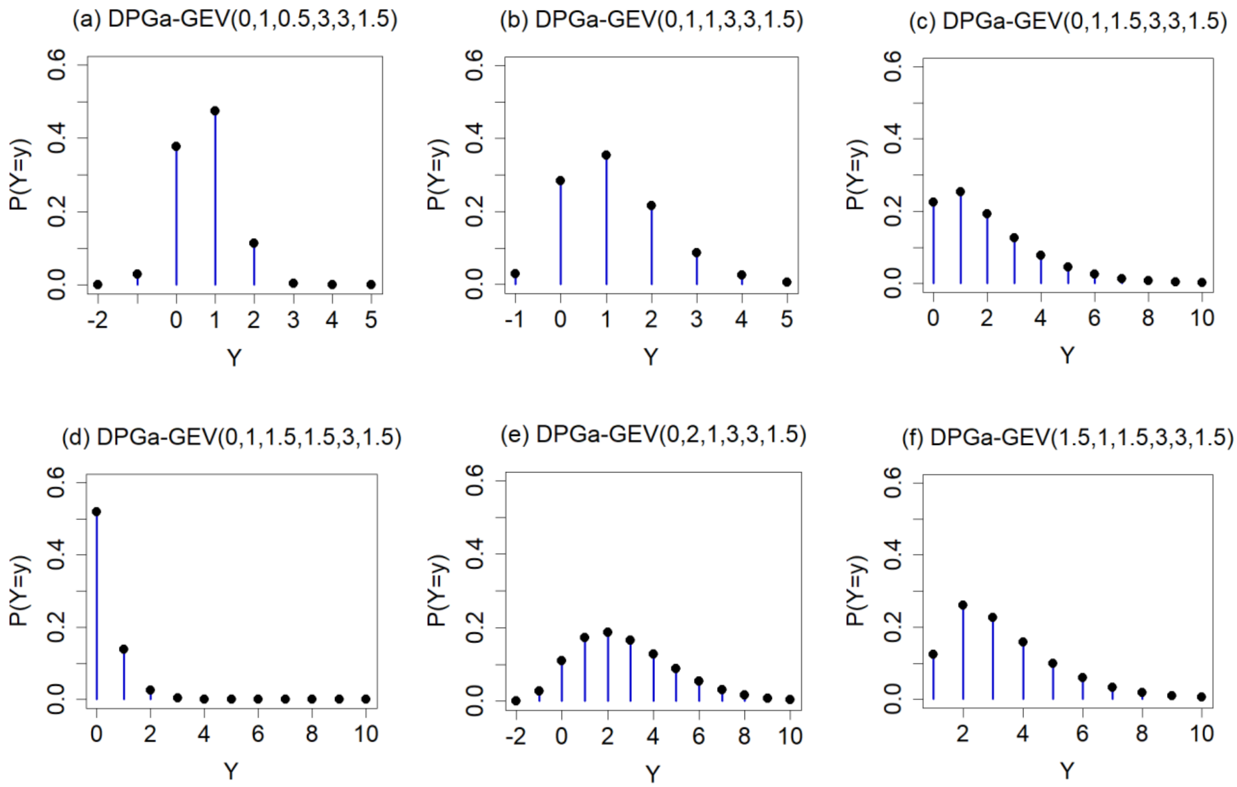

Figure 5. Thp pmf plots of the DPGa-GEV distribution for $\xi > 0$.

Let Y_t be iid random variables of size n following the pmf (15) with $\mathbf{\Theta} = (a, b, c, \mu, \sigma, \xi)$, then its log-likelihood is

$$l(\mathbf{\Theta}) = \sum_{t=1}^{n} \log\left\{ \left[1 + \frac{b}{b+2}\left(\frac{\exp\{-a\tau(y_t)\}}{1-\exp\{-a\tau(y_t)\}}\right)^c\right] \exp\left\{-b\left(\frac{\exp\{-a\tau(y_t)\}}{1-\exp\{-a\tau(y_t)\}}\right)^c\right\} \right.$$
$$\left. - \left[1 + \frac{b}{b+2}\left(\frac{\exp\{-a\tau(y_t+1)\}}{1-\exp\{-a\tau(y_t+1)\}}\right)^c\right] \exp\left\{-b\left(\frac{\exp\{-a\tau(y_t+1)\}}{1-\exp\{-a\tau(y_t+1)\}}\right)^c\right\} \right\}.$$

The ML estimator $\hat{\mathbf{\Theta}}$ of $\mathbf{\Theta}$ is obtained numerical from the nonlinear equations

$$\frac{\partial l(\mathbf{\Theta})}{\partial \mu} = \frac{\partial l(\mathbf{\Theta})}{\partial \sigma} = \frac{\partial l(\mathbf{\Theta})}{\partial \xi} = \frac{\partial l(\mathbf{\Theta})}{\partial a} = \frac{\partial l(\mathbf{\Theta})}{\partial b} = \frac{\partial l(\mathbf{\Theta})}{\partial c} = 0. \tag{20}$$

Since (19) - (20) cannot be derived in a closed form, thus the numerical method was used. The ML estimators of $\mathbf{\theta}$ and $\mathbf{\Theta}$ were obtained by the nlm function in the stats package [21].

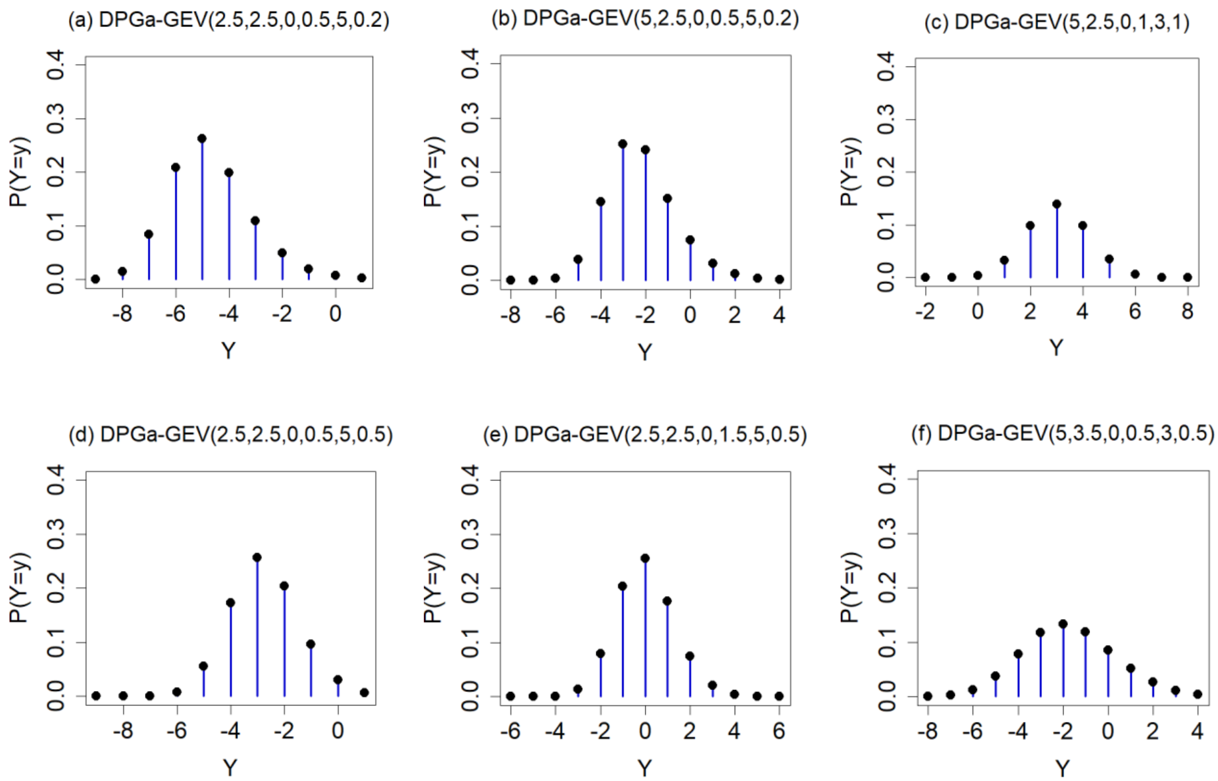

Figure 6. The pmf plots of the DPGa-GEV distribution for $\xi = 0$.

3.4 Extreme value analysis of the PM2.5 and PM10 in Pathum Thani, Thailand

In this study, we used two sets of real data on which we apply the proposed distributions that were developed in the preceding Sections 3.1 and 3.2. The data sets are related to air pollution, such as the PM2.5 and PM10 in Pathum Thani province, Thailand, from 2019 to 2023. Observations of daily PM are the 24-hour averages (unit: micrograms per cubic meter) of PM2.5 and PM10, which are report by Air Quality and Noise Management Office, Pollution Control Department, Ministry of Natural Resources and Environment, Thailand [22].

Let $X_t = \max(X_{t1}, \ldots, X_{tj}, \ldots, X_{tn_t})$ where X_{tj} is the 24-hour averages of PM2.5 (or PM10) in month $t = 1, 2, 3, \ldots, 60$ and day $j = 1, 2, \ldots, n_t$, X_t is a maximum value of PM2.5 (or PM10) in month t, and n_t is the number of days in month t. The maximum value of PM2.5 and PM10 in each month, Pathum Thani province from 2019 to 2023 are show in Table 1 and Figure 7. The distributions of a maximum value of PM2.5 and PM10 are right-skewed and heavy-tailed.

Table 1. Summary of empirical data concerning the maximum value of PM2.5 and PM10 in each month, Pathum Thani province from 2019 to 2023 [22].

Months	PM2.5					PM10				
	2019	2020	2021	2022	2023	2019	2020	2021	2022	2023
Jan	88	67	82	34	68	97	84	94	94	120
Feb	60	74	65	31	111	75	105	79	96	181
Mar	43	36	45	27	87	64	63	62	85	140
Apr	32	35	30	33	51	48	50	45	122	93
May	37	26	28	14	32	76	54	38	55	62
Jun	22	17	21	12	22	40	31	32	53	46
Jul	25	19	20	10	18.1	44	31	45	43	40
Aug	18	22	25	12	21.2	33	39	48	45	45
Sep	62	22	25	12	22.2	84	36	44	47	43
Oct	40	36	31	20	40.6	62	53	55	81	60
Nov	59	40	36	23	44.5	77	54	61	84	79
Dec	58	55	63	27	56.5	80	69	103	83	83

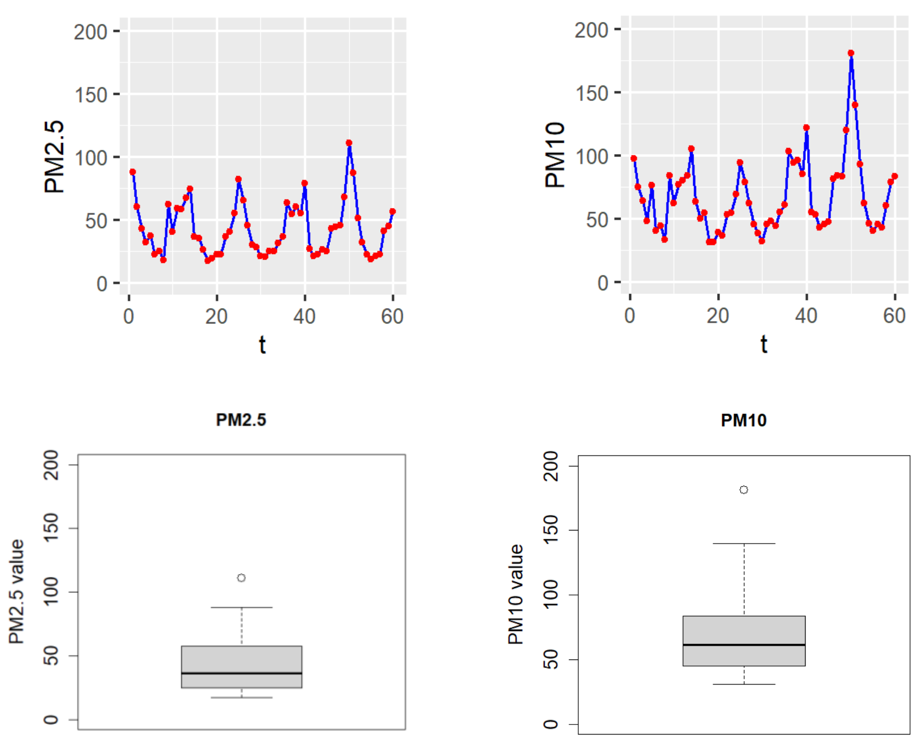

Figure 7. Line and box plots of the maximum value of PM2.5 and PM10.

Table 2. The values of ML estimates of distributions, log-likelihood (logL), and KS test for extreme value analysis of PM2.5 in Pathum Thani province, Thailand, from 2019 to 2023.

ML estimates	$\xi \neq 0$				$\xi = 0$			
	GEV	DGEV	PGa-GEV	DPGa-GEV	GEV	DGEV	PGa-GEV	DPGa-GEV
$\hat{\mu}$	30.1970	30.7053	24.1772	17.9468	34.3981	35.3079	38.1509	1.0106
$\hat{\sigma}$	12.7719	12.7770	49.8722	3.0204	35.8066	50.7084	28.9258	3.9847
$\hat{\xi}$	0.3461	0.3452	6.2319	0.0715	-	-	-	-
$\hat{a}$	-	-	103.7740	20.7395	-	-	89.2176	1.0533
$\hat{b}$	-	-	1090.042	0.6892	-	-	5.1726	0.2302
$\hat{c}$	-	-	0.0813	0.1715	-	-	0.0215	0.1423
-logL	259.06	259.07	255.10	252.28	283.50	300.27	264.15	281.51
KS	0.1054	0.1098	0.0975	0.0693	0.2238	0.3022	0.1192	0.2186
(p-value)	0.5180	0.4649	0.6181	0.9354	0.0049	<0.0001	0.3609	0.0064

In this study, we analyze the extreme value of the PM2.5 and PM10 as an application of the PGa-GEV distribution and compare it to the GEV distribution for cases $\xi \neq 0$ and $\xi = 0$. The PM data sets are continuous value for X_t, but it is recorded in discrete value. Thus, we apply the DPGa-GEV distribution to analysis these data, and compare it with the PGa-GEV distribution, GEV distribution, and DGEV distribution. In this study, the Kolmogorov-Smirnov (KS) test is used as criteria for the goodness of fit, where the model that gives the smaller values of KS test is the better fit to the data.

Extreme value analysis of the PM2.5 and PM10 in Pathum Thani, Thailand:

Let X_t is a maximum value of PM2.5 (or PM10) in month t from 2019-2023 for $t = 1, 2, 3, \ldots, 60$, and $\lfloor X_t \rfloor$ is equal to the greatest integer less than or equal to X_t. The parameter estimates and the goodness of fit test for PM2.5 data are summarized in Table 2. The DPGa-GEV distribution gives the lower KS values than other distributions, i.e., PGa-GEV, DGEV, and GEV distributions. We conclude that the DPGa-GEV distribution with $\xi \neq 0$ is an appropriate distribution to fit the PM2.5 data (KS = 0.0639, p-value = 0.9354), see Figure 8. For

PM10 data, the ML estimates and the goodness of fit test are shown in Table 3. The DPGa-GEV distribution gives the smallest value of AIC and KS, then it is the best distribution to describe the PM10 data (KS = 0.0902, p-value = 0.7131), see Figure 8.

Table 3. The values of ML estimates of distributions, log-likelihood (logL), and KS test for extreme value analysis of PM10 in Pathum Thani province, Thailand, from 2019 to 2023.

ML estimates	$\xi \neq 0$				$\xi = 0$			
	GEV	DGEV	PGa-GEV	DPGa-GEV	GEV	DGEV	PGa-GEV	DPGa-GEV
$\hat{\mu}$	52.8034	53.3046	39.1634	32.4299	25.8032	29.2679	44.2194	2.0571
$\hat{\sigma}$	18.7401	18.7385	57.8950	62.7781	113.3222	93.9364	38.3736	4.9957
$\hat{\xi}$	0.1789	0.1789	4.8773	41.7813	-	-	-	-
$\hat{a}$	-	-	48.8382	0.7245	-	-	127.8754	1.0393
$\hat{b}$	-	-	909.4893	0.2928	-	-	5.3544	0.2300
$\hat{c}$	-	-	0.1803	34.7365	-	-	0.0240	0.1165
-logL	276.75	276.75	275.13	274.10	348.60	338.64	280.07	304.01
KS	0.1107	0.1160	0.1020	0.0902	0.3847	0.3786	0.1000	0.2386
(p-value)	0.4544	0.3951	0.5599	0.7131	<0.0001	<0.0001	0.5861	0.0022

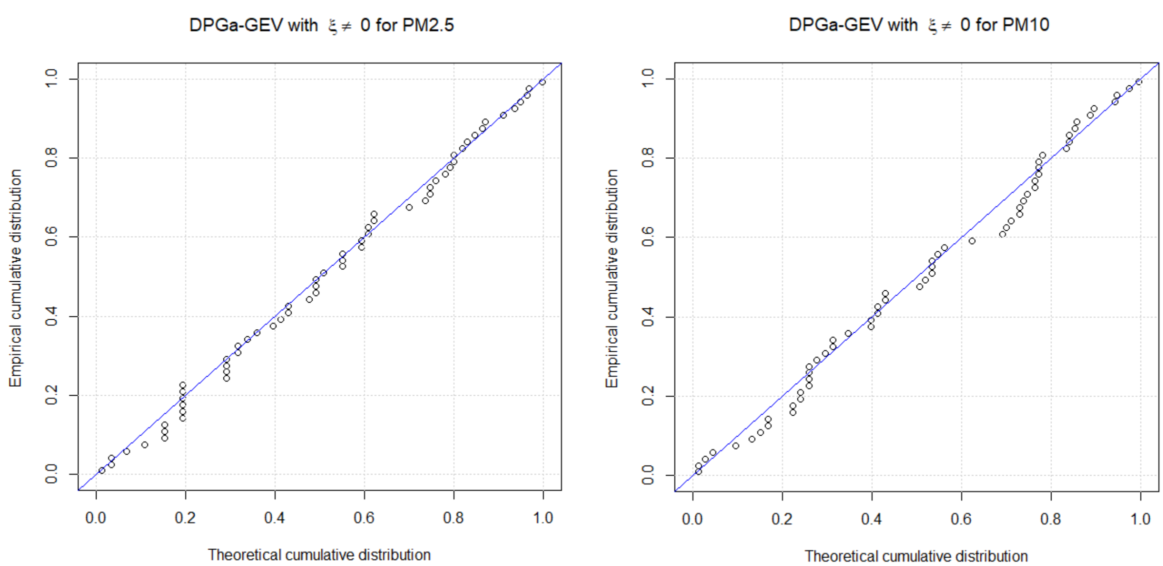

Figure 8. Empirical cumulative distribution and DPGa-GEV cumulative distribution for the maximum value of PM2.5 and PM10 in Pathum Thani province

Return level of PM2.5 and PM10 in Pathum Thani province

In our present work, we first calculate the monthly maximum PM2.5 (or PM10). These extreme-related quantities are, respectively, fitted to the DPGa-GEV distribution. From the fitted distribution, we can estimate how often the extreme quantiles occur with a certain return level. We define the return value as the average value that we expect to equal or exceed once every interval of time (T), with a probability of $\frac{1}{T}$. The return value can be calculated by solving this equation

$$\hat{Z}_T^{\text{DPGa-GEV}} = \left| \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left\{ 1 - \left[-\log\left(1 + \left\{ -\frac{1}{\hat{b}} \hat{\delta}_T \right\}^{-1/\hat{c}} \right)^{-1/\hat{a}} \right]^{-\hat{\xi}} \right\} - 1 \right|,$$

where $\hat{\delta}_T = W_{-1}\left\{ -\frac{1}{T}(2+\hat{b})\exp(-2-\hat{b}) \right\} + 2 + \hat{b}.$ For calculating $W_{-1}\{z\}$, the function "W(z,branch=-1)" in the LambertW package in R [20].

From Table 2, the ML estimates of the DPGa-GEV distribution are $\hat{\mu}$ =17.9468, $\hat{\sigma}$ =3.0204, $\hat{\xi}$ =0.0715, $\hat{a}$ =20.7395, $\hat{b}$ =0.6892, and $\hat{c}$ =0.1715. The expression of return levels at period T of the maximum value of PM2.5 in Pathum Thani is

$$\hat{Z}_T^{\text{PM2.5}} = \left| 17.9468 - 42.24336 \left\{ 1 - \left[-\log\left(1 + (-1.450958\hat{\delta}_T)^{-5.830904} \right)^{-0.04821717} \right]^{-0.0715} \right\} - 1 \right|$$

where $\hat{\delta}_T = W_{-1}\left\{ -\frac{1}{T}(2.6892)\exp(-2.6892) \right\} + 2.6892.$

From the results in Table 3, the ML estimates of the DPGa-GEV distribution are $\hat{\mu}$ =32.4299, $\hat{\sigma}$ =62.7781, $\hat{\xi}$ =41.7813, $\hat{a}$ =0.7245, $\hat{b}$ =0.2928, and $\hat{c}$ =34.7365. Then the expression of return levels at period T of the maximum value of PM10 in Pathum Thani is

$$\hat{Z}_T^{\text{PM10}} = \left| 32.4299 - 1.502541 \left\{ 1 - \left[-\log\left(1 + \left\{ -3.415301\hat{\delta}_T \right\}^{-0.02878816} \right)^{-1.380262} \right]^{-41.7813} \right\} - 1 \right|$$

where $\hat{\delta}_T = W_{-1}\left\{ -\frac{1}{T}(2.2928)\exp(-2.2928) \right\} + 2.2928.$

Table 4. The values of the DPGa-GEV return level at period T of PM2.5 and PM10 in Pathum Thani province, Thailand

Data	Return level at period T (micrograms per cubic meter)								
	2 years	5 years	10 years	15 years	20 years	25 years	30 years	50 years	100 years
PM2.5	36	59	72	78	82	85	87	94	101
PM10	59	87	106	116	124	129	133	145	161

Table 4 shows the predicted maximum PM2.5 and PM10 return levels (in micrograms per cubic meter) with the DPGa-GEV model for the return periods of 2, 5, 10, 15, 20, 25, 30, 50, and 100 years along. The results show an increasing return level of PM2.5 and PM10 as T increases. The probability of occurring an PM2.5 that exceeds the magnitude 36 micrograms per cubic meter is 0.5 for the next two years; the probability of occurring an PM2.5 that exceeds the magnitude 72 micrograms per cubic meter is 0.1 for the next ten years; and the probability of occurring an PM2.5 that exceeds the magnitude 82 micrograms per cubic meter is 0.05 for the next twenty years. For PM10 data, the probability of occurring an PM10 that exceeds the magnitude 59 micrograms per cubic meter is 0.5 for the next two years; the probability of occurring an PM10 that exceeds the magnitude 106 micrograms per cubic meter is 0.1 for the next ten years; and the probability of occurring an PM2.5 that exceeds the magnitude 124 micrograms per cubic meter is 0.05 for the next twenty years.

4. CONCLUSIONS

This study presents the development and application of the power Garima-generalized extreme value (PGa-GEV) distribution, alongside its discrete counterpart, for analyzing extreme PM2.5 and PM10 values in Pathum Thani, Thailand. The research highlights the robustness of these models, particularly the DPGa-GEV distribution, in fitting empirical data and predicting extreme events over extended return periods. The findings underscore the higher return levels for PM10 compared to PM2.5, especially over 10- and 20-year periods, which is critical for decision-making in public health and infrastructure planning. The study provides essential insights for Thai meteorologists and policymakers, enabling them to make decisions that mitigate the adverse effects of extreme PM2.5 and PM10 on human health and the environment. The research emphasizes the need for continuous preparedness and adaptation to enhance community resilience in the face of climate change. This research demonstrates the application and significance of extreme value theory in describing extreme PM2.5 and PM10 events in Pathum Thani, Thailand.

The DPGa-GEV, PGa-GEV, DGEV, and GEV models are considered for maximum monthly data in Pathum Thani from 2019 to 2023. We estimated the model parameters using maximum likelihood (ML) estimation and fit the distributions using the block maxima approach. Our findings reveal that the DPGa-GEV distribution is the optimal model from the GEV family for the monthly maximums of daily PM2.5 and PM10 data. We discovered that as return periods increase, so do the return levels. The average value of PM2.5 exceeding 36, 72, and 80 micrograms per cubic meter has a 50%, 10%, and 5% probability in the next two years, the next ten years, and the next twenty years, respectively. The probability of the PM10 average exceeding 59, 106, and 124 micrograms per cubic meter is 50%, 10%, and 5% in the next two years, the next ten years, and the next twenty years, respectively. When comparing the return levels for PM2.5 and PM10, our results show that PM10 has higher return levels for 10 and 20 years compared to PM2.5. The model diagnostics showed that the models were reasonable for modeling the PM2.5 and PM10 data. This study will equip Pathum Thani's decision-makers with insights into extreme PM2.5 and PM10 events during the specified return periods, empowering them to make informed choices that mitigate the harm extreme PM2.5 and PM10 inflict on people, infrastructure, and lives. As climate change persists, continuous preparedness and adaptation measures are essential for the Pathum Thani community's resilience. Therefore, this research will aid in developing strategies for early warning, management, preparedness, response, and mitigation of PM2.5 and PM10 risks. However, future studies can model and predict extreme rainfall in PM2.5 and PM10 with respect to specific regions of the country. Additionally, modeling extreme PM2.5, PM10, rainfall, and temperature in Pathum Thani is a possible research direction.

ACKNOWLEDGMENTS

This research was supported by the National Science, Research and Innovation Fund, Thailand Science Research and Innovation (TSRI), through Rajamangala University of Technology Thanyaburi (FRB67E0619) (Grant No.: FRB670027/0168). The authors would like to thank the anonymous reviewers for their comments and suggestions.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] S. Zhou, Q. Deng, W. Liu, Extreme air pollution events: Modeling and prediction, J. Cent. South Univ. Technol. 19 (2012), 1668–1672. https://doi.org/10.1007/s11771-012-1191-2.

[2] S. Hazarika, P. Borah, A. Prakash, The assessment of return probability of maximum ozone concentrations in an urban environment of Delhi: A generalized extreme value analysis approach, Atmos. Environ. 202 (2019), 53–63. https://doi.org/10.1016/j.atmosenv.2019.01.021.

[3] J.J. Pornsopin, P. Busababodhin, T. Phoophiwfa, et al. Risk analysis of PM2.5 and PM10: A case study at Khon Kaen City, J. Appl. Sci. 20 (2021), 157–172. https://doi.org/10.14416/j.appsci.2021.02.012.

[4] W. Warsono, Y. Antonio, S.B. Yuwono, et al. Modeling generalized statistical distributions of PM2.5 concentrations during the COVID-19 pandemic in Jakarta, Indonesia, Decision Sci. Lett. 10 (2021), 393–400. https://doi.org/10.5267/j.dsl.2021.1.005.

[5] S. Aryuyuen, W. Bodhisuwan, The Topp-Leone generalized extreme value distribution: extreme value analysis and return level estimation of the PM2.5 in Chiang Mai, Thailand, Songklanakarin J. Sci. Technol. 44 (2022), 1450-1461.

[6] A.I. Aguirre-Salado, S. Venancio-Guzmán, C.A. Aguirre-Salado, et al. A novel tree ensemble model to approximate the generalized extreme value distribution parameters of the PM2.5 maxima in the Mexico City Metropolitan Area, Mathematics. 10 (2022), 2056. https://doi.org/10.3390/math10122056.

[7] E. Tanprayoon, U. Tonggumnead, S. Aryuyuen, A new extension of generalized extreme value distribution: extreme value analysis and return level estimation of the rainfall data, Trends Sci. 20 (2022), 4034. https://doi.org/10.48048/tis.2023.4034.

[8] T.G. Bali, The generalized extreme value distribution, Econ. Lett. 79 (2003), 423–427. https://doi.org/10.1016/s0165-1765(03)00035-1.

[9] C.T. Guloksuz, N. Celik. An extension of generalized extreme value distribution: Uniform-GEV distribution and its application to earthquake data, Thail. Stat. 18 (2020), 491-506.

[10] M.A. Esfeh, L. Kattan, W.H.K. Lam, et al. Compound generalized extreme value distribution for modeling the effects of monthly and seasonal variation on the extreme travel delays for vulnerability analysis of road network, Transp. Res. Part C: Emerg. Technol. 120 (2020), 102808. https://doi.org/10.1016/j.trc.2020.102808.

[11] S. Aryuyuen, W. Bodhisuwan, T. Ngamkham, Power Garima-generated family of distributions: Properties and application, Lobachevskii J. Math. 42 (2021), 287–299. https://doi.org/10.1134/s1995080221020050.

[12] A. Alzaatreh, C. Lee, F. Famoye, A new method for generating families of continuous distributions, METRON. 71 (2013), 63–79. https://doi.org/10.1007/s40300-013-0007-y.

[13] A.W. Kemp, Classes of discrete lifetime distributions, Commun. Stat. – Theory Methods. 33 (2004) 3069–3093. https://doi.org/10.1081/sta-200039051.

[14] D. Roy, Discrete Rayleigh distribution, IEEE Trans. Rel. 53 (2004), 255–260. https://doi.org/10.1109/tr.2004.829161.

[15] D. Veberič, Lambert W function for applications in physics, Computer Phys. Commun. 183 (2012), 2622–2628. https://doi.org/10.1016/j.cpc.2012.07.008.

[16] G.M. Goerg, LambertW: An R package for Lambert W x F random variables, R package version 0.6.9-1, (2023).

[17] A.F. Jenkinson, The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Quart. J. R. Meteorol. Soc. 81 (1955), 158–171. https://doi.org/10.1002/qj.49708134804.

[18] S. Kotz, S. Nadaraja, Extreme value distribution: theory and applications, Imperial College Press, Singapore, 2000.

[19] S. Coles, An introduction to statistical modeling of extreme values, Springer, London, 2001. https://doi.org/10.1007/978-1-4471-3675-0.1

[20] S. Chakraborty, Generating discrete analogues of continuous probability distributions-A survey of methods and constructions, J. Stat. Distrib. Appl. 2 (2015), 6. https://doi.org/10.1186/s40488-015-0028-6.

[21] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org.

[22] Air Quality and Noise Management Office, Department of Pollution Control, Air quality historical data, (2024). http://air4thai.pcd.go.th/webV3/#/History.