Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:60
https://doi.org/10.28919/cmbn/9914
ISSN: 2052-2541

A NOVEL UNIT GARIMA-GENERALIZED EXTREME VALUE DISTRIBUTION FOR STATIONARY AND NON-STATIONARY MODELING: APPLICATION TO PM2.5 MAXIMA IN THAILAND

SIRINAPA ARYUYUEN[1,*], ISSARAPORN THIAMSORN[1], SUPAWADEE PATATHANONE[2]

[1]Department of Mathematics and Computer Science, Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110, Thailand

[2]Department of Chemistry, Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110, Thailand

Abstract: This presented an extended generalized extreme value (GEV) distribution called the unit Garima-generalized extreme value (UGa-GEV) model for block maxima (BM) analysis. Its properties, including survival, hazard, quantile, and return level functions, are proposed. The parameters of the proposed models were estimated by the maximum likelihood method, which was applied to analyze the maximum monthly particulate matter with a diameter of less than 2.5 microns (PM2.5) in Thailand. Climate change has undermined the reliability of stationary models to accurately record extreme environmental events over the last few decades with exploration now focusing on nonstationary factors like trends and shifts to better reflect changing environmental conditions. To address this problem, this study proposed non-stationary maxima models using covariates in meteorology (PM10, ozone, carbon monoxide, nitrogen dioxide, and sulfur dioxide) and time. The optimal model between the GEV and UGa-GEV was compared using real data analysis in Bangkok and Chiang Rai, Thailand with our results demonstrating that the

*Corresponding author
E-mail address: sirinapa_a@rmutt.ac.th
Received April 7, 2026

proposed UGa-GEV model provided more flexibility than the traditional GEV model.

Keywords: extreme value analysis; unit Garima-GEV; non-stationary block maxima method; PM2.5.

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

1. INTRODUCTION

The extreme value theorem (EVT) provides a robust theoretical framework for characterizing the statistical properties of random variables classified as extreme values (the maximum or minimum observations within a specific period). Traditional data analysis often neglects these extremes due to their inherent complexity and the difficulty of modeling rare events, but understanding their probability distributions is crucial, especially in the context of heavy-tailed data. In environmental science, these extremes manifest as critical events such as peak daily rainfall, maximum monthly wind speeds, or hazardous levels of particulate matter (PM). The two main approaches to analyzing extreme values are the block maxima (BM) method, which selects only the highest (or lowest) values from each specified time period and analyzes them using the generalize extreme value (GEV) distribution, and the peak-over-threshold (POT) method, which focused on all data values that exceed a set limit using the generalized Pareto distribution. This study focuses on the BM approach because this is uniquely suited for analyzing environmental data with clear periodic patterns, such as annual rainfall or seasonal PM cycles. The dataset was partitioned into fixed time intervals (blocks) to effectively capture the most extreme event within each period while mitigating the noise from daily fluctuations. This study employed the BM approach as the primary method for identifying extrema within the dataset. The choice was motivated by the method's ability to systematically partition data into significant temporal blocks, such as months or years and capture the most critical peak values, while maintaining a robust connection to the GEV distribution theory [1, 2]. Unlike the POT method, which requires a subjective choice of threshold, the BM method provides a stable and objective framework for estimating return periods [3], essential for assessing long-term risks associated with extreme environmental phenomena, such as sea-level rise [4] or air pollution episodes [5].

The GEV distribution, systematized by [3], serves as a robust framework for modeling these extremes. The analysis of extreme values is indispensable for developing robust preparedness strategies against high-impact, damaging events, thereby mitigating potential degradation in future occurrences. Accurate cross-examination of extreme values is crucial to obtaining the most precise analysis results and using them as a basis for future data management. In recent years, the analysis of extreme events has intensified across various fields, including hydrology, finance, and coastal engineering. Within the realm of climate change and meteorology, the precise prediction of rare events is vital for mitigating socio-economic impacts and analyzing PM concentrations with a diameter of less than 2.5 microns (PM2.5) in urban areas has become a priority. Identifying the probability of high-value air pollution episodes is essential for public health protection because-high PM2.5 values lead to severe acute health impacts. Numerous studies have demonstrated the efficacy of extreme value analysis in environmental contexts [5, 6, 7, 8, 9, 10, 11].

Atmospheric pollution, particularly PM2.5, has now emerged as a critical public health crisis in many regions. The traditional GEV distribution with constant values of the location parameter μ, scale parameter σ, and shape parameter ξ, often assumes that the occurrence of peak pollution levels is stationary. However, PM2.5 extremes are highly sensitive to shifting meteorological conditions and evolving emission patterns, manifesting in clear non-stationary characteristics. Applying a non-stationary GEV distribution allows researchers to incorporate time-varying covariates, such as wind speed or seasonal indices, into the model parameters. This approach provides a more realistic estimation of return periods for hazardous air quality events. Understanding these non-stationary trends is essential for policymakers to develop adaptive air quality management strategies and to strengthen public health preparedness against increasingly frequent extreme pollution episodes. The GEV distribution, systematized by [3], serves as a robust framework for modeling these extremes. The critical challenge lies in the sensitivity of its parameters, particularly ξ. Minor fluctuations in ξ can radically alter the distribution's tail behavior, shifting from a bounded Weibull class to a heavy-tailed Fréchet class, which predicts much more severe flooding events for long return periods, as highlighted in recent studies of the

Baltic Sea [12, 13]. [9] studied non-stationary models for monthly PM2.5 data when using the GEV distribution on the covariables in μ and constant σ and ξ, while [14] derived non-stationary modeling of extremes performed on the BM of water level when the parameters of the GEV distribution were time-variable. A non-stationary GEV analysis was performed assuming that the parameters were functions of time. [15] proposed model averaging of non-stationary models for extreme value analyses, while [16] derived models for sea level using nonstationary GEV with time function parameters. However, understanding how these parameters evolve in time and space is paramount, and failure to account for non-stationarity in infrastructure design and coastal management could result in catastrophic socio-economic losses, making the development of advanced non-stationary models a vital necessity for climate adaptation.

An extreme model necessitates the development of tailored distributions that closely align with the empirical characteristics of the data. The adoption of versatile frameworks is particularly advantageous, as they possess the capacity to accommodate diverse data structures and complex stochastic behaviors. Consequently, there is a growing trend among researchers to utilize novel generalizations of continuous distributions to expand their adaptability and enhance their descriptive power. Such advancements are driven by the need for superior goodness-of-fit and a more precise characterization of extreme tail features. This research bridges the gap between theoretical GEV modeling and practical air quality management by employing advanced distributions to capture the volatility of extreme environmental data better. To address these requirements, several flexible extensions of the GEV distribution were introduced, including the Topp-Leone GEV [11], uniform-GEV [17], Gompertz-GEV [18], compound GEV [19], and power Garima-GEV [5] distributions.

Thailand continues to face protracted air pollution challenges, particularly within dense urban areas, construction sites, and industrial zones. National air quality indices have shown gradual long-term improvements over the past decade, but several regions consistently exceed safety thresholds for key pollutants, most notably particulate matter and ozone gas. These pollutants primarily stem from energy consumption in the transport sector, power generation, and industrial activities. In 2022-2023, the situation was relatively severe in Northern Thailand and Bangkok due to the full resumption of economic activity after COVID-19, with dry and stagnant weather leading to the accumulation of dust particles. PM2.5 values remain relatively high at certain times. In 2025, the situation stabilized and slightly decreased in some areas due to proactive measures and stricter controls on agricultural burning. However, air pollution in Bangkok has now reached a critical juncture due to the synergistic effects of stagnant atmospheric conditions and intensified local emissions. In early 2026, Bangkok recorded high levels of particulate matter during periods of air inversion. The northern region of Thailand consistently reports the PM2.5 maxima and the most days exceeding the nation standard. Chiang Rai is a heavily polluted area, with a high concentration of particulate matter due to poor ventilation [20]. The EVT was used to analyze datasets with more extreme values of PM2.5.

This paper presented a novel extension of the GEV distribution, developed within the framework of the unit Garima-generated (UGa-G) family of distributions [21]. This family was constructed using the T-X distribution framework [22], employing the unit Garima random variable as the generator. This proposed model provides more flexibility compared to the traditional GEV distribution as an alternative for BM analysis, with our non-stationary maxima models using covariates in meteorology and time. The maximum likelihood method was used to estimate the model parameters. Our proposed distributions were used for extreme value analysis to assess the monthly maximum PM2.5 values in Bangkok and in Chiang Rai in the northern region of Thailand.

2. Preliminaries

Theoretical foundations utilized in this study included the UGa-G family of distributions, the GEV distribution, and the methodologies for incorporating nonstationary factors into extreme value modeling.

2.1 The UGa-G Family of Distributions

The UGa-G family of distributions introduced by [21], is characterized by two positive parameters, a and θ. The mathematical framework for its cumulative distribution function (CDF) and probability density function (PDF) can be defined as follows:

$$F_{\text{UGa-G}}(y)=\left[1+\frac{\theta}{\theta+2}\left(G^{-a}(y;\boldsymbol{\omega})-1\right)\right]\exp\left\{-\theta\left(G^{-a}(y;\boldsymbol{\omega})-1\right)\right\}, \tag{1}$$

$$f_{\text{UGa-G}}(y)=\frac{a\theta g(y;\boldsymbol{\omega})}{\theta+2}\left[\frac{\theta+G^{a}(y;\boldsymbol{\omega})}{G^{2a+1}(y;\boldsymbol{\omega})}\right]\exp\left\{-\theta\left(G^{-a}(x;\boldsymbol{\omega})-1\right)\right\}, \tag{2}$$

where $g(y;\boldsymbol{\omega})$ and $G(y;\boldsymbol{\omega})$ represent the PDF and CDF of a baseline distribution, and $\boldsymbol{\omega}$ denotes the vector of baseline parameters. Its quantile function (QF) is

$$Q_{\text{UGa-G}}(u)=G^{-1}\left\{\left[\frac{\theta}{-2-W_{-1}\left\{-u(2+\theta)\exp(-2-\theta)\right\}}\right]^{1/a}\right\},\quad \text{for}\quad 0<u<1 \tag{3}$$

where $G^{-1}\{\cdot\}$ denotes the inverse CDF of a baseline distribution [21], and $W_{-1}(\cdot)$ is the lower branch of the Lambert W function [23].

2.2 Block Maxima Model

The block maxima model approach involves partitioning a dataset into equal-sized blocks and identifying the maximum value within each process, known as extracting block maxima. These extreme observations are typically modeled using the GEV distribution, which was first introduced by [24]. Let Y be a GEV random variable, then its CDF and PDF are respectively

$$G_{\text{GEV}}(y)=\begin{cases}\exp\left(-\left[1+\xi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{-1/\xi}\right), & \text{if } \xi\neq 0\\[2ex] \exp\left(-\exp\left(\dfrac{y-\mu}{\sigma}\right)\right), & \text{if } \xi=0\end{cases} \tag{4}$$

$$g_{\text{GEV}}(y)=\begin{cases}\dfrac{1}{\sigma}\left[1+\xi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{-(1+1/\xi)}\exp\left\{-\left[1+\xi\left(\dfrac{y-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, & \text{if } \xi\neq 0\\[2ex] \dfrac{1}{\sigma}\left[\exp\left(\dfrac{y-\mu}{\sigma}\right)\right]^{\xi+1}\exp\left\{-\exp\left(\dfrac{y-\mu}{\sigma}\right)\right\}, & \text{if } \xi=0\end{cases} \tag{5}$$

where $-\infty<\mu<\infty, \sigma>0,$ and $-\infty<\xi<\infty.$ The GEV distribution reduces to the Fréchet distribution if $\xi>0$ for $y\in[\mu-\sigma/\xi,+\infty),$ and the Weibull distribution if $\xi<0$ for

$y \in (-\infty, \mu - \sigma/\xi]$. When $\xi = 0$, it is referred to as the Gumbel distribution for $y \in (-\infty, +\infty)$. The QF of the GEV distribution is

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

for $0 < u < 1$ [25]. The application of the GEV distribution for estimating extreme values and their corresponding return level (Z_p) at return period T is based on the principle that independent and identically distributed (iid) block maxima z_p. Based on $\mathrm{P}(Z > Z_p) = 1 - p$, the return period for a particular $\hat{z}$ is given by $T(\hat{z}) = \left[1 - G(\hat{z})\right]^{-1}$. The return level z_p, which is expected to be exceeded once every $T = 1/p$ years, can be estimated using the following formula [3]

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

2.3 Non-Stationary GEV Distribution

A sliding-window technique was used, where stationary GEV distribution was iteratively fitted to block maximum to estimate non-stationarity levels. Each fit estimated a fixed set of parameters μ, σ, and ξ. Data such as PM2.5 or global temperature often involve trends or seasonal patterns. The location, scale and shape parameters of a GEV distribution change over time (t) or other independent variables (covariates) $\mathbf{x}$ [3]. If the parameter is revealed to be time-variable, a nonstationary-GEV analysis is used. It is assumed that the parameters are functions of t such that $\mu = \mathrm{M}(t)$, $\sigma = \Sigma(t)$, and $\xi = \Xi(t)$. The CDF of nonstationary-GEV is [14]

$$G_{\text{GEV}}(y;t) = \begin{cases} \exp\left(-\left[1 + \Xi(t)\left(\dfrac{y - \mathrm{M}(t)}{\Sigma(t)}\right)\right]^{-1/\Xi(t)}\right), & \text{if } \xi \neq 0 \\ \exp\left(-\exp\left(\dfrac{y - \mathrm{M}(t)}{\Sigma(t)}\right)\right), & \text{if } \xi = 0. \end{cases} \tag{8}$$

Various specifications were employed to capture the non-stationarity of the parameters, including linear, quadratic, and piecewise constant step functions, to ensure a robust representation of time variability. The time-dependent non-stationarity GEV models considered that the parameter μ corresponded to constant, linear, second order, and exponential functions, σ constant and exponential function, and constant ξ [15]. Several research studies have incorporated covariates $X_1, X_2, \ldots, X_k$ into parameters to more accurately describe the occurrence of extremes for $\mu = \mathrm{M}(\mathbf{x}) = \mu_0 + \mu_1 X_1 + \mu_2 X_2 + \cdots + \mu_k X_k$ [9, 26, 27].

3. MAIN RESULTS

3.1 The Unit Garima-Generalized Extreme Value Distribution

A novel distribution was proposed, using the UGa-G family and the GEV distribution as the baseline for extreme value analysis, to improve the flexibility and comprehensiveness of the GEV distribution for the BM model.

Theorem 1. Let Y be a random variable in the unit Garima-generalized extreme value (UGa-GEV) distribution with parameters $\mu, \sigma, \xi,$ and θ, denoted by $Y \sim$UGa-GEV(μ,σ,ξ,θ). Then the CDF and PDF of Y are, respectively

$$F(y) = \left(1 + \frac{\theta}{2+\theta}\left[\exp(\tau_y) - 1\right]\right)\exp\left\{-\theta\left[\exp(\tau_y) - 1\right]\right\}, \tag{9}$$

$$f(y) = \frac{\theta}{(2+\theta)\sigma}\left[\exp(-\tau_y) + \theta\right](\tau_y)^{\xi+1}\exp\left\{-\theta\left[\exp(\tau_y) - 1\right] + 2\tau_y\right\}, \tag{10}$$

where $-\infty < \mu < \infty,\ \sigma > 0,\ -\infty < \xi < \infty,$ and $\theta > 0$, and $1 + \xi\left(\frac{y-\mu}{\sigma}\right) > 0$ for

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

Proof. The CDF and PDF of the T-X family [22] are defined as follows

$$F(y)=\int_{a}^{W[G(y)]} r(s)ds, \text{ and } f(y)=\left\{\frac{d}{dy}W[G(y)]\right\} r\{W[G(y)]\}, \tag{11}$$

where S represents the unit Garima (UGa) distribution [21] with the PDF

$$r(s)=\frac{\theta(s+\theta)}{(2+\theta)s^3}\exp\left\{-\theta\left(\frac{1}{s}-1\right)\right\}, \text{ for } 0\le s\le 1.$$

The GEV distribution was used for the baseline distribution with the CDF (3). Subtracting $r(s)$ and $W[G(y)]=G(y)$ in (11) gives the CDF and PDF of Y as

$$\begin{aligned}F(y)&=\left[1+\frac{\theta}{2+\theta}\left(\frac{1}{\exp(-\tau_y)}-1\right)\right]\exp\left\{-\theta\left(\frac{1}{\exp(-\tau_y)}-1\right)\right\}\\&=\left(1+\frac{\theta}{2+\theta}\left[\exp(\tau_y)-1\right]\right)\exp\left\{-\theta\left[\exp(\tau_y)-1\right]\right\},\end{aligned}$$

$$\begin{aligned}f(y)&=\frac{1}{\sigma}(\tau_y)^{\xi+1}\exp(-\tau_y)\frac{\theta}{(2+\theta)}\frac{\exp(-\tau_y)+\theta}{\left[\exp(-\tau_y)\right]^3}\exp\left\{-\theta\left(\frac{1}{\exp(-\tau_y)}-1\right)\right\}\\&=\frac{\theta}{(2+\theta)\sigma}\left[\exp(-\tau_y)+\theta\right](\tau_y)^{\xi+1}\exp\left\{-\theta\left[\exp(\tau_y)-1\right]+2\tau_y\right\}.\end{aligned}$$

The existing UGa-GEV distribution was used to obtain the extended the GEV distribution using the UGa-G family presented by [21] with $W[G(y)]=G^a(y)$ for $a=1$. The density plots of the UGa-GEV distribution are shown in Figures 1-3. The UGa-GEV distribution reduces to the following unit distributions based on the value of the shape parameter ξ: (i) A UGa-Gumbel distribution occurs when the limit of ξ approaches zero. Figure 1 shows the right skewed density plot. For constant σ and θ, the distribution had the same distribution when μ changed and the peak of the curve changed location (Figure 1(a)). Thus, μ acted as a location parameter. For constant μ and the density shape remained unchanged, but the density curve became spikier and more spread out as the value of σ increased (Figure 1(b)). Therefore, σ acted as a scale parameter. When the values of μ and σ were constant, the density shape remained unchanged, but the density curve became less prominent and less dispersed as θ increased (Figure 1 (c)). (ii) The density shape had a right skew (Figure 2) for the UGa-Fréchet distribution when ξ was positive. In Figure 2 (d) – (f), if the other parameters are constant, when the θ value is changed, its density plot also changes shape. For $\theta \le 1$ it becomes a decreasing function and is right skewed for the case $\theta > 1$. Therefore, θ acted as a shape parameter. (iii) The UGa-Weibull distribution occurs when the shape parameter is negative. The density graph had a right skew distribution (Figure 3), with a location parameter μ and scale parameters σ and θ.

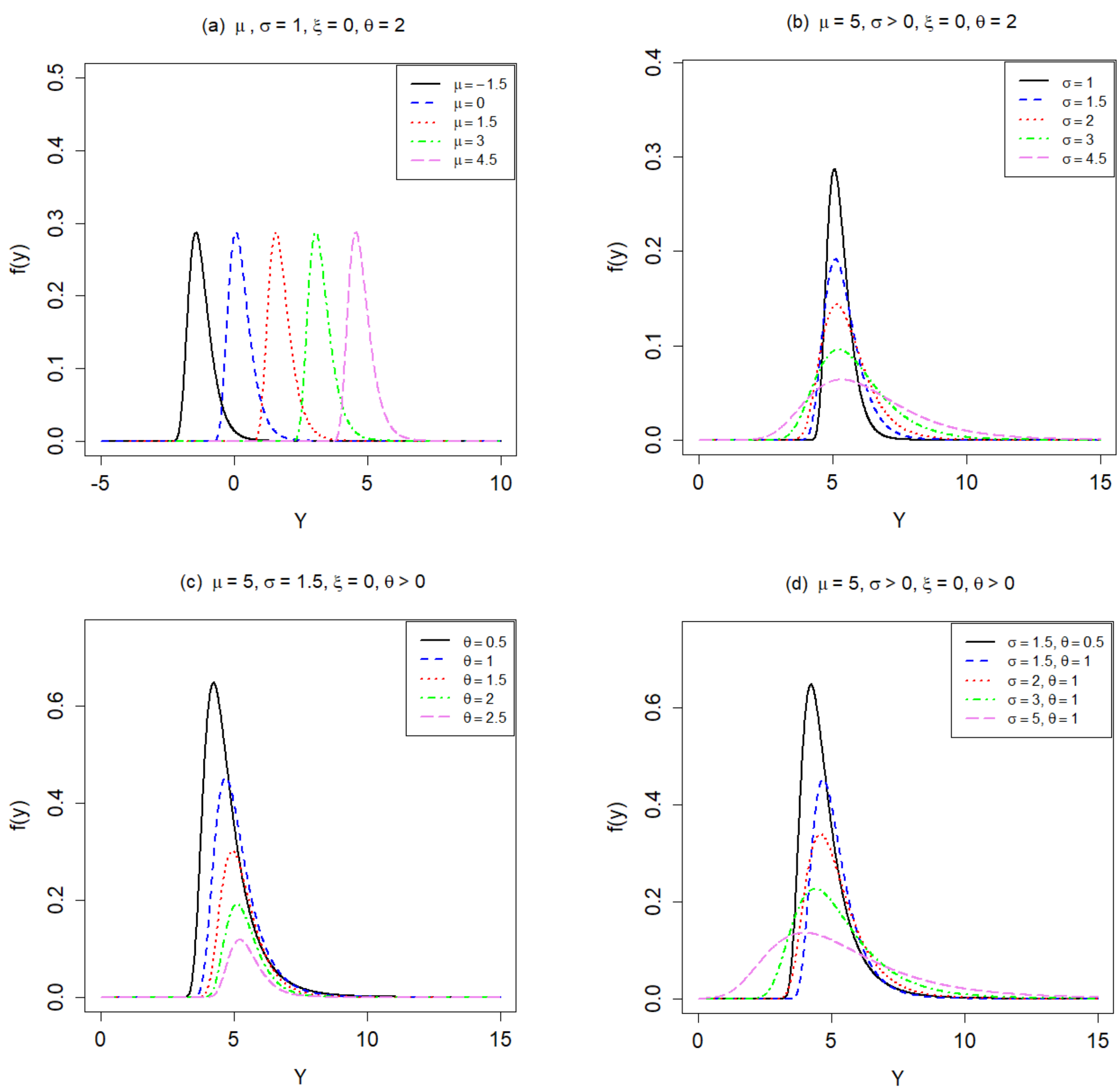

FIGURE 1. Density plots of the UGa-GEV distribution for $\xi = 0$

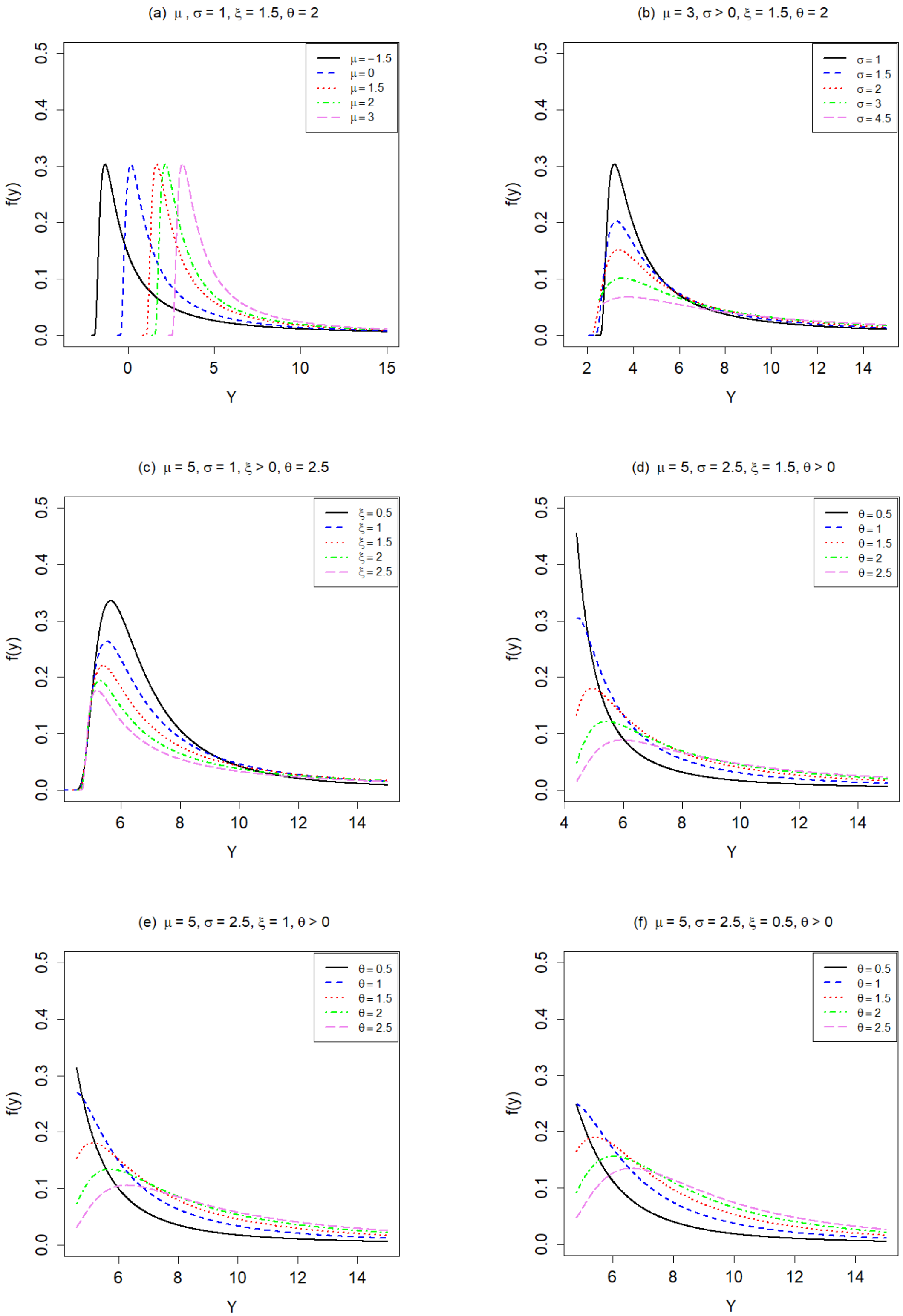

FIGURE 2. Density plots of the UGa-GEV distribution for $\xi > 0$

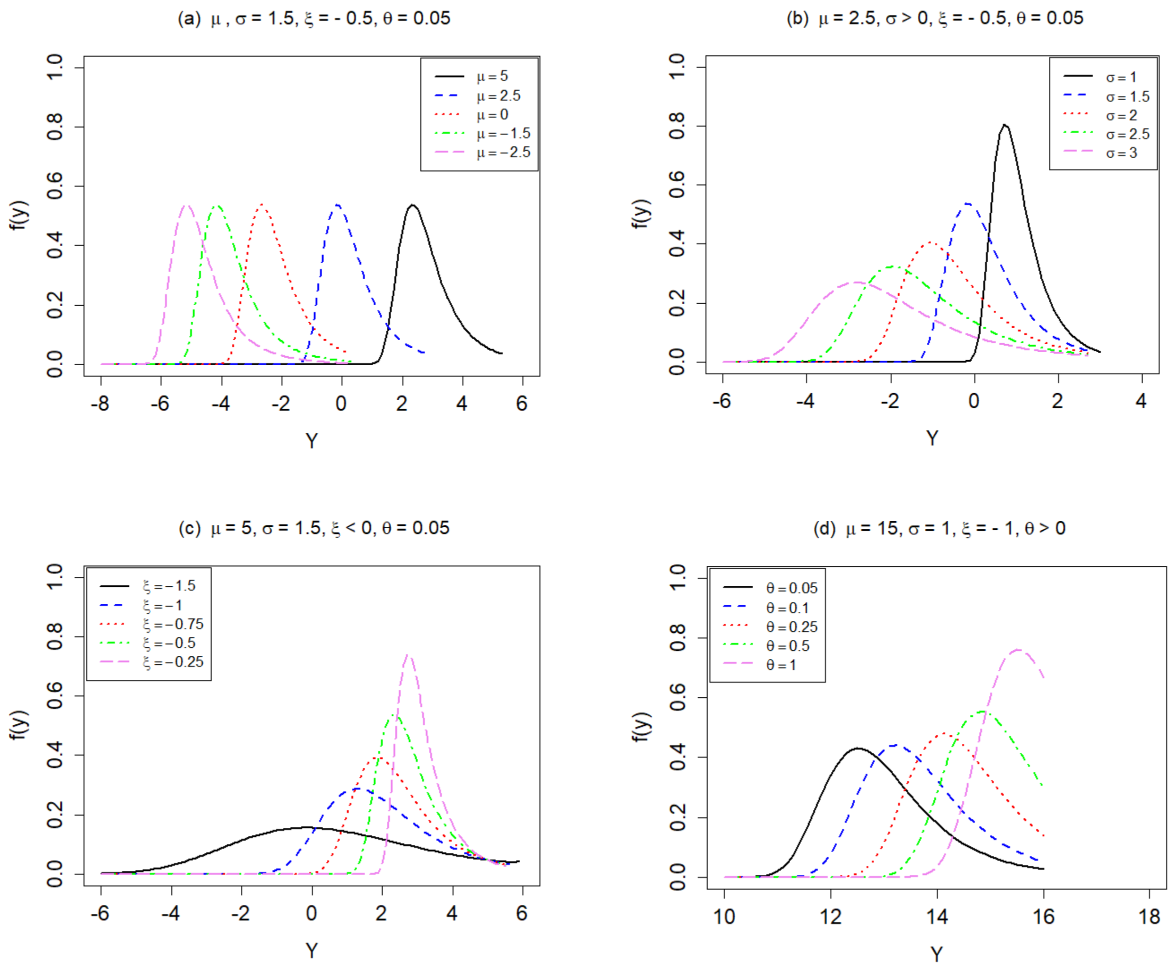

FIGURE 3. Density plots of the UGa-GEV distribution for $\xi < 0$

3.2 Properties of the UGa-GEV Distribution

Let $Y \sim$ UGa-GEV $(\mu, \sigma, \xi, \theta)$, its survival function $S(y) = 1 - F(y)$ is

$$S(y) = 1 - \left[1 + \frac{\theta}{2+\theta}\left(\exp(\tau_y) - 1\right)\right]\exp\left\{-\theta\left(\exp(\tau_y) - 1\right)\right\} \quad \text{for } \theta > 0 .$$

Its corresponding hazard function $h(y)$ can be written as

$$h(y) = \frac{1}{\sigma}\frac{\theta\left[\exp(-\tau_y) + \theta\right](\tau_y)^{\xi+1}\exp\left\{-\theta\left[\exp(\tau_y) - 1\right] + 2\tau_y\right\}}{2 + \theta - \left[2 + \theta + \theta(\exp(\tau_y) - 1)\right]\exp\left\{-\theta\left[\exp(\tau_y) - 1\right]\right\}} .$$

From the UGa-GEV distribution with the CDF of $F(y)$, let $F(y) = U$ where U is a uniform random variable on the interval (0, 1), then the QF, $Q_Y(u) = F^{-1}(y)$, is

$$Q_Y(u) = \begin{cases} \mu - \dfrac{\sigma}{\xi}\left(1 - \left[-\log\left(\dfrac{\theta}{-2 - W\left\{-u(2+\theta)\exp(-2-\theta)\right\}}\right)\right]^{-\xi}\right), & \text{if } \xi \neq 0 \\ \mu - \sigma\log\left[-\log\left(\dfrac{\theta}{-2 - W\left\{-u(2+\theta)\exp(-2-\theta)\right\}}\right)\right], & \text{if } \xi = 0 \end{cases}$$

where $W_{-1}(\cdot)$ is obtained by the "lambertWm1" function in the lamW package [28] in R. The return level z_p of the UGa-GEV distribution, which is expected to be exceeded once every $T = 1/p$ years can be represented as

$$z_p(\mu,\sigma,\xi,\theta) = \begin{cases} \mu - \dfrac{\sigma}{\xi}\left(1 - \left[-\log\left(\dfrac{\theta}{-2 - W\left\{-(1-p)(2+\theta)\exp(-2-\theta)\right\}}\right)\right]^{-\xi}\right), & \text{if } \xi \neq 0 \\ \mu - \sigma\log\left[-\log\left(\dfrac{\theta}{2 - W\left\{-(1-p)(2+\theta)\exp(-2-\theta)\right\}}\right)\right], & \text{if } \xi = 0 \end{cases} \tag{12}$$

3.3 Non-Stationary UGa-GEV Model

The non-stationary UGa-GEV models, covariates $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ or time t are provided as:

- Model 1, stationary data where $Y \sim$UGa-GEV(μ,σ,ξ,θ):
- Model 2, non-stationary data on time t where $Y \sim$UGa-GEV$(\mathrm{M}(t),\sigma,\xi,\theta)$:

  $\mu = \mathrm{M}(t) = \mu_0 + \mu_1 t,$ and the constant values of σ and ξ. (13)

- Model 3: non-stationary data on time t where $Y \sim$UGa-GEV$(\mathrm{M}(t),\Sigma(t),\xi,\theta)$:

  $\mu = \mathrm{M}(t) = \mu_0 + \mu_1 t,\ \sigma = \Sigma(t) = \exp(\sigma_0 + \sigma_1 t)$ and a constant ξ. (14)

- Model 4: non-stationary data on covariates $\mathbf{x}$ where $Y \sim$UGa-GE $(\mathrm{M}(\mathbf{x}),\sigma,\xi,\theta)$:

  $\mu = \mathrm{M}(\mathbf{x}) = \mu_0 + \mu_1 x_1 + \mu_2 x_2 + \cdots + \mu_k x_k,$ and the constant values of σ and ξ. (15)

- Model 5: non-stationary data on $\mathbf{x}$ and t where $Y \sim$UGa-GEV$(\mathrm{M}(\mathbf{x}),\Sigma(t),\xi,\theta)$:

  $\mu = \mathrm{M}(\mathbf{x}) = \mu_0 + \mu_1 x_1 + \mu_2 x_2 + \cdots + \mu_k x_k,\quad \sigma = \Sigma(t) = \exp(\sigma_0 + \sigma_1 t)$ and a constant ξ. (16)

3.4 Parameter Estimation and Model Selection

The maximum likelihood estimation (MLE) was used to find the value that maximized the likelihood function to ensure that the calculated parameters best matched and described the observed data under the given model. The MLE method is widely favored in academic research due to its desirable consistent and efficient asymptotic properties which facilitate the derivation of standard errors and the implementation of model selection criteria like the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Model selection was conducted by evaluating candidate models through the AIC, a widely recognized metric in statistical modeling. This criterion was formulated as $\mathrm{AIC} = -2\log L + 2m$, where L and m are the likelihood value and the number of parameters of the model, respectively. Another robust alternative is the BIC was formulated as $\mathrm{BIC} = -2\log L + m\log n$, where n is the number of observations. The optimal model was identified by the minimum AIC and BIC values.

3.5 Extreme Value Analysis of PM2.5 in Thailand

In this study, two real datasets related to air pollution, as the PM2.5 in Bangkok and Chiang Rai Provinces, Thailand, from 2020 to 2024, were analyzed using the proposed models and the baseline models. The datasets sourced from [20] and comprised 24-hour averages of PM2.5 (measured in micrograms per cubic meter (mg/m^3)) for Bangkok and Chiang Rai. Several secondary environmental variables were collected to supplement these observations as monthly averages, including PM10 (PM_{10}, µg/m^3), ozone (O_3, ppb), carbon monoxide (CO, ppm), nitrogen dioxide (NO_2, ppb), and sulfur dioxide (SO_2, ppb). Data collected via Air4Thai revealed that some stations did not report/record relevant data, which was considered as missing with data from nearby stations used as replacements. If no data were found at any nearby station, the average for that year was used. Let $Y_i = \max(Y_{i1}, \ldots, Y_{ij}, \ldots, Y_{in_i})$ represent the maximum PM2.5 concentration recorded during month i, where Y_{ij} is the 24-hour average of PM2.5 in month i and day $j = 1, 2, \ldots, n_i$, n_i denotes the total count of days within month i. The distribution parameters were estimated using the meteorological covariates and covariate instead of time (t) for $t = 1$ (January, 2020) to

$t = 60$ (December, 2024). However, the PM2.5 value is correlated with historical time. Therefore, as an approximation, time was added as a covariate for parameter estimation. The following steps were taken to analyze the data to study the probability of PM2.5 extreme events, as the maximum monthly PM2.5 from 2020 to 2024 (a total of 60 months):

1) Input data sets for month $t = 1, 2, 3, ..., 60$, such as the maximum monthly PM2.5 value (Y_t), and the monthly averages of PM_{10} (X_{t1}), O_3 (X_{t2}), CO (X_{t3}), NO_2 (X_{t4}) and SO_2 (X_{t5}).

2) The stationary data of Y_t were checked using the Augmented Dickey-Fuller (ADF) test (H_0: data is stationary, and H_1: data is non-stationary).

3) The co-variables $X_{t1}, X_{t2}, ..., X_{t5}$ that affected Y_t were analyzed using multiple regression analysis with the enter method. Only co-variables that significantly affected Y_t at the 0.05 level were selected to construct the models.

4) The BM models for stationary and non-stationary data were defined in subsection 3.3. The MLE parameter estimates of each model were obtained using the nlm function in the stats package [29].

5) The optimal model of Y_t, was selected as the model with the lowest AIC and BIC values.

6) The return level of Y_t was estimated over the next 2, 5, 10, and 15 years using an optimal model.

3.5.1 Extreme value analysis of PM2.5 values in Bangkok

The value of Y_t in Bangkok (Station 59T) between 2020 and 2024 [20] indicated that the PM2.5 value exceeded the standard limit in 21 months (more than 37.5 mg/m^3), with Y_t exceeding the standard between December and February of each year. The distribution of Y_t was right-skewed and heavy-tailed (Figure 4).

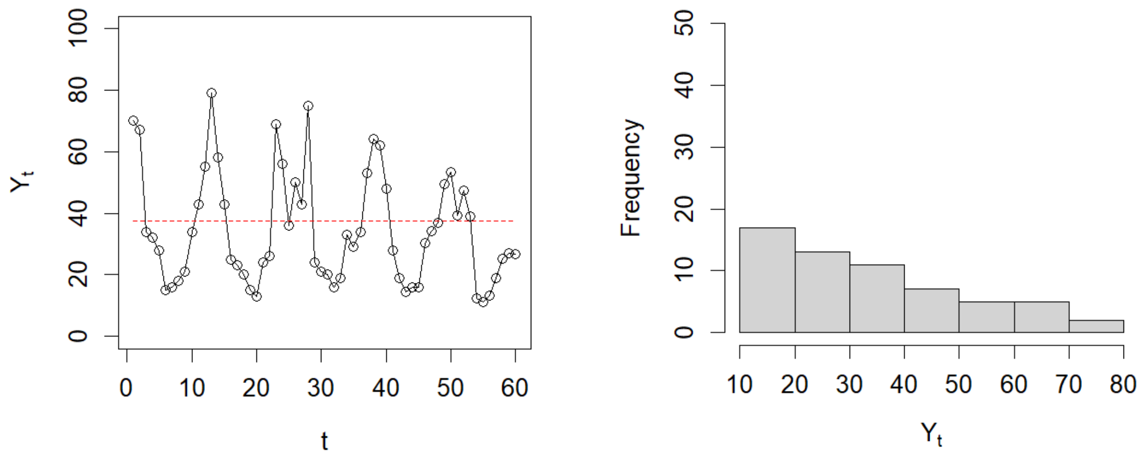

FIGURE 4. Maximum monthly PM2.5 values (Y_t) in Bangkok between 2020 and 2024

TABLE 1. Parameter estimates using the MLE of Y_t in Bangkok, assuming stationary data

Models	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\theta}$	KS test (p-value)
GEV	0.112	39.915	2.775	-	0.443 (< 0.001)
UGa-GEV	11.602	14.736	0.061	2.398	0.080 (0.840)

Table 1 shows that the UGa-GEV model was suitable for analyzing Y_t in Bangkok, and provided the return level of Y_t as

$$\hat{z}_p = 11.602 - \frac{14.736}{0.061}\left(1 - \left[-\log\left(\frac{2.398}{-2 - W\{-(1-p)(4.398)\exp(-4.398)\}}\right)\right]^{-\xi}\right), \quad p = \frac{1}{T}. \quad (17)$$

The return level of Y_t tended to increase when T increased (as shown in Figure 5). Return periods of 2, 5, 10, and 15 years, gave the return levels of Y_t in Bangkok as 29.445, 46.005, 58.334, and 65.680 mg/m^3, respectively.

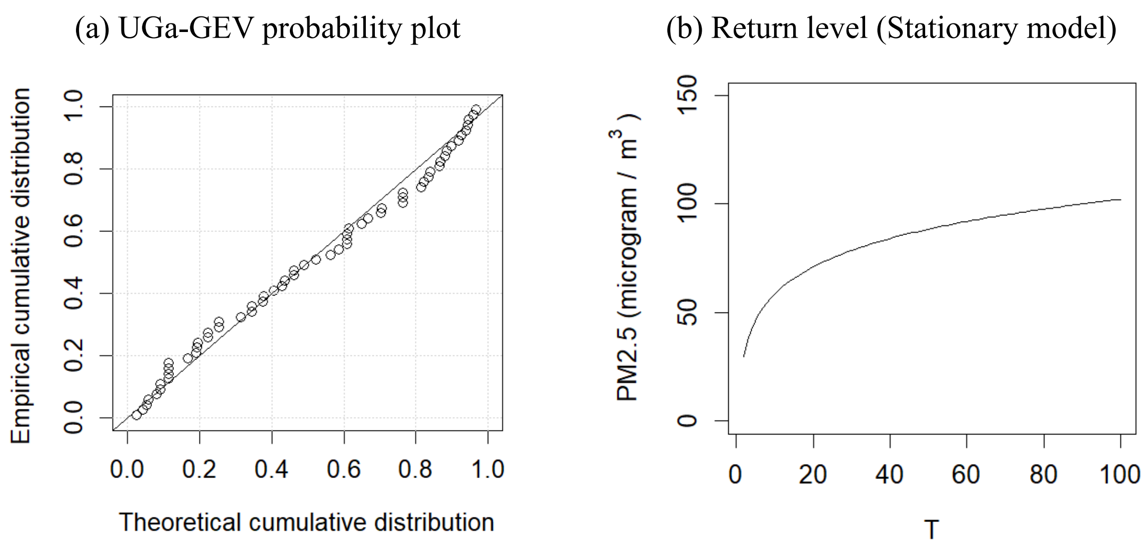

FIGURE 5. Probability plot and return levels of Y_t in Bangkok using the UGa-GEV model with $\hat{\mu} = 18.499$, $\hat{\sigma} = 17.894$ and $\hat{\theta} = 1.312$

TABLE 2. Analysis results of the factors affecting Y_t in Bangkok

Variables	Estimate	Standard error (S.E.)	t	p-value
Intercept	-15.916	4.814	-3.306	0.002
PM_{10}	0.922	0.202	4.561*	< 0.001
O_3	0.753	0.251	3.001*	0.004
CO	0.498	3.203	0.156	0.877
NO_2	-0.015	0.143	-0.104	0.917
SO_2	-0.319	1.138	-0.280	0.780
Residual S.E. = 8.499, R^2 = 0.796, Adjusted R^2 = 0.777, F = 42.03, p-value < 0.001				

*Statistically significant at the 0.05 level.

Based on the ADF test $(\text{ADF} = -3.839, \text{p-value} = 0.005)$, the data Y_t in Bangkok was non-stationary. From the multiple regression analysis in Table 2, the variables that influenced Y_t at a significance level of 0.05 $(F = 42.03, \text{p-value} < 0.001)$ were PM_{10} and O_3. The optimal model of Y_t was the non-stationary UGa-GEV model with $\hat{M}(\mathbf{x}) = 16.289 + 1.027\text{PM}_{10} + 0.325\text{O}_3$, $\hat{\Sigma}(t) = \exp(2.387 + 0.004t)$, $\hat{\xi} = -0.528$, and $\hat{\theta} = 0.021$ (Table 3).

TABLE 3. Parameter estimates of each model for Y_t in Bangkok

Models	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\theta}$	-log$L$	AIC	BIC
GEV	0.112	39.915	2.775	-	355.52	717.04	723.32
UGa-GEV	11.602	14.736	0.061	2.398	250.51	509.02	517.40
GEV	27.115 - 0.075t	11.871	0.227	-	250.96	509.91	518.29
UGa-GEV	26.981 - 0.080t	21.168	-0.178	0.868	249.53	509.05	519.52
GEV	29.088 - 0.130t	exp(-0.006 + 2.674t)	0.190	-	250.61	511.22	521.69
UGa-GEV	29.298 - 0.145t	exp(3.320 - 0.008t)	-0.253	0.828	248.93	513.22	525.79
GEV	- 4.692 + 0.552PM_{10} + 0.613O_3	6.227	0.155	-	207.82	425.64	436.11
UGa-GEV	-20.934+0.827PM_{10}+ 0.553O_3	7.567	-0.023	3.697	204.14	420.28	432.85
GEV	-11.967 + 0.865PM_{10} + 0.540O_3	exp(1.708 + 0.003t)	0.059	-	203.52	419.04	431.61
UGa-GEV	16.289 + 1.027PM_{10} + 0.325O_3	exp(2.387 + 0.004t)	-0.528	0.021	199.86	413.72	428.38

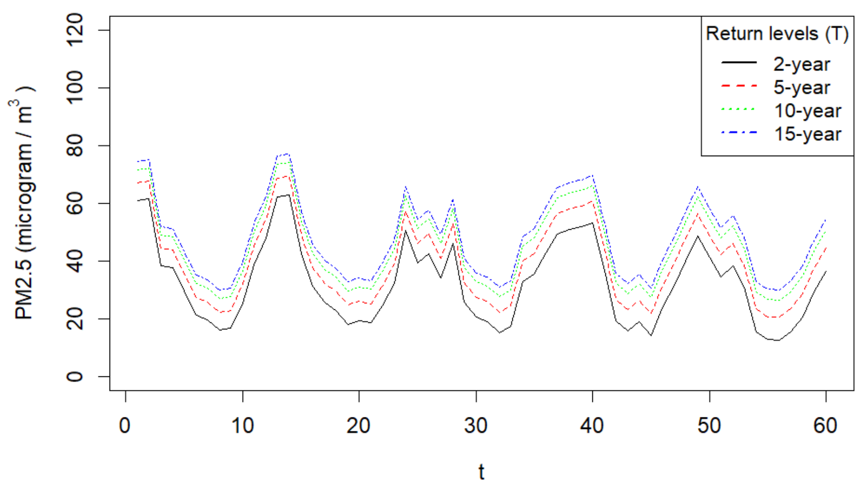

FIGURE 6. Return level of Y_t in Bangkok using the UGa-GEV model with $\hat{\xi}$ = - 0.528, $\hat{\mu}$ =16.289+1.027PM_{10}+0.325O_3, $\hat{\sigma}$ = exp(2.387+0.004t), and $\hat{\theta}$ = 0.021

The return level of Y_t was

$$\hat{z}_{t,p} = \hat{\mathrm{M}}(\mathbf{x}) + \frac{\hat{\Sigma}(t)}{0.528}\left(1 - \left[-\log\left(\frac{0.021}{-2 - W\{-(1-p)(2.021)\exp(-2.021)\}}\right)\right]^{-\xi}\right), \quad p = \frac{1}{T} \tag{18}$$

where $\hat{\mathrm{M}}(\mathbf{x}) = 16.289 + 1.027x_{t1} + 0.325x_{t2}, \hat{\Sigma}(t) = \exp(2.387 + 0.004t)$ for $t = 1, 2, \ldots, 60$. The return level of Y_t concentration increased when periods of T increased (Figure 6). The return levels of Y_t (mg/m³) concentration for 2 years, 5 years, 10 years, and 15 years were 62.77, 69.23, 74.09, and 77.02, respectively.

3.5.2 Extreme value analysis of PM2.5 values in Chiang Rai

The values of Y_t in Chiang Rai (Station 57T) between 2020 and 2024 exceeded the standard limit in 25 months (more than 37.5 mg/m^3). Frequency Y_t exceeded the standard between February and April each year, while Y_t had a right skewed and heavy-tailed distribution (Figure 7).

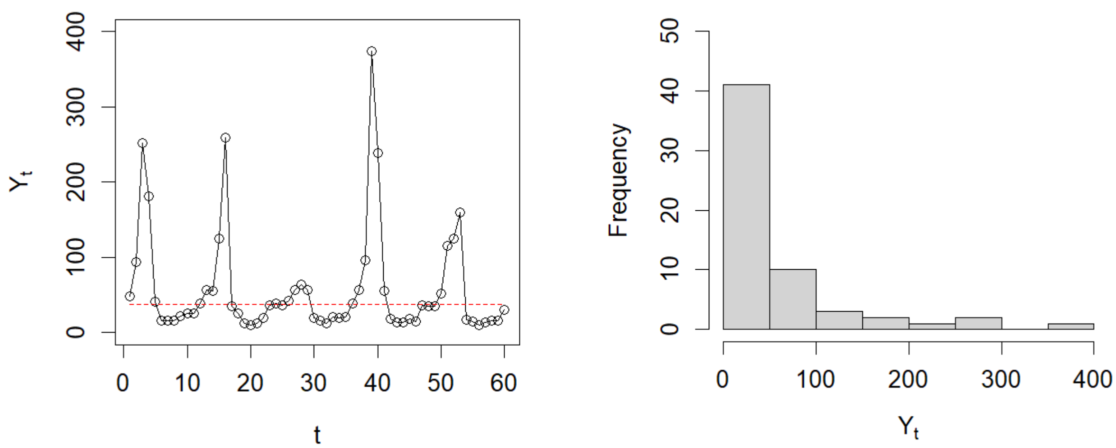

FIGURE 7. Maximum monthly PM2.5 values (Y_t) in Chiang Rai between 2020 and 2024

TABLE 4. Parameter estimates using the MLE of Y_t in Chiang Rai, assuming stationary data.

Models	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\theta}$	KS test (p-value)
GEV	0.112	39.911	2.844	-	0.435 (<0.001)
UGa-GEV	7.481	2.861	0.907	6.827	0.108 (0.488)

Table 4 shows that the UGa-GEV model was suitable for analyzing Y_t in Chiang Rai, with the expression of return level Y_t as

$$\hat{z}_p = 7.481 - \frac{2.861}{0.907}\left(1 - \left[-\log\left(\frac{6.827}{-2 - W\left\{-(1-p)(8.827)\exp(-8.827)\right\}}\right)\right]^{-\xi}\right), \quad p = \frac{1}{T}. \quad (19)$$

The return level Y_t tended to increase when T increased (Figure 8). Time periods (T) of 2, 5, 10, and 15 years gave return levels of Y_t in Chiang Rai as 28.100, 68.428, 129.745, and 187.918 mg/m^3, respectively.

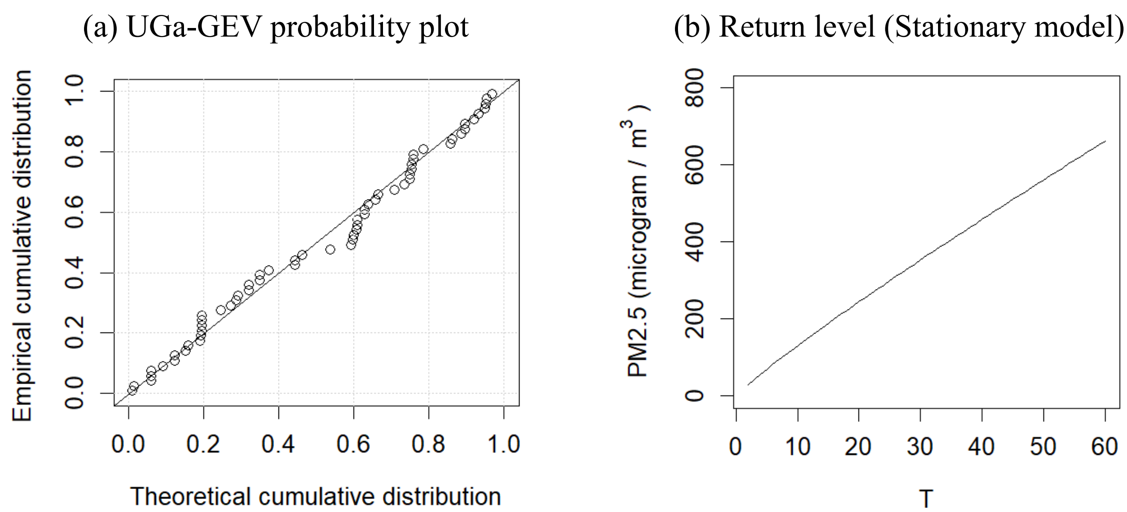

FIGURE 8. Probability plot and return level of Y_t in Chiang Rai using the UGa-GEV model with $\hat{\mu} = 7.481, \hat{\sigma} = 2.861, \hat{\xi} = 0.970$ and $\hat{\theta} = 6.827$

TABLE 5. Analysis results of the factors affecting Y_t in Chiang Rai

Variables	Estimate	S.E.	t	p-value
Intercept	-15.256	13.193	-1.156	0.253
PM_{10}	2.285	0.295	7.760*	< 0.001
O_3	0.175	0.784	0.223	0.825
CO	-25.232	37.001	-0.682	0.498
NO_2	0.394	2.271	0.174	0.863
SO_2	-5.711	4.867	-1.173	0.246
Residual S.E. = 34.58, $R^2 = 0.787$, Adjusted $R^2 = 0.768$, F = 39.95, p-value < 0.001				

*Statistically significant at the 0.05 level.

The Y_t data were non-stationary (ADF = –4.270, p-value = 0.001). The multiple regression analysis results in Table 5 showed that the variables influenced Y_t in Chiang Rai at a significance level of 0.05 (F = 39.95, p-value < 0.001) for PM_{10}.

TABLE 6. Parameter estimates of each model for Y_t in Chiang Rai.

Models	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\theta}$	-log$L$	AIC	BIC
GEV	0.112	39.911	2.844	-	367.60	741.20	747.48
UGa-GEV	7.481	2.861	0.907	6.827	286.18	580.36	588.74
GEV	22.909 - 0.034t	14.604	0.913	-	285.92	579.83	588.21
UGa-GEV	7.951 - 0.034t	1.747	0.905	11.329	285.91	581.83	592.30
GEV	27.917 - 0.184t	exp(3.139 - 0.015t)	0.889	-	284.96	579.91	590.38
UGa-GEV	4.977 + 0.033t	exp(- 0.041 - 0.014t)	0.889	35.422	284.96	581.91	594.48
GEV	- 14.875 + 1.523PM_{10}	13.263	0.226	-	257.64	523.28	531.66
UGa-GEV	46.304 + 1.479PM_{10}	49.500	0.005	0.056	256.17	522.34	532.81
GEV	- 6.864 + 1.338PM_{10}	exp(1.615 + 0.026t)	0.394	-	249.73	509.46	519.93
UGa-GEV	18.545 + 1.300PM_{10}	exp(3.313 + 0.010t)	0.363	0.170	246.36	504.72	517.29

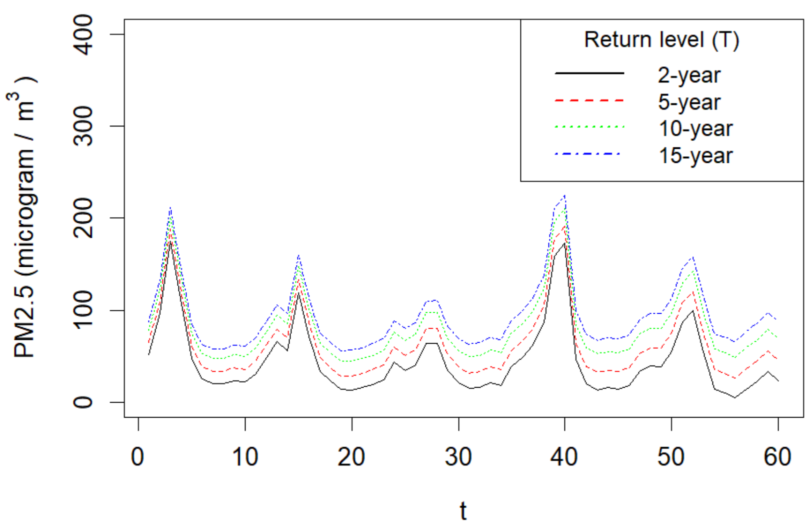

FIGURE 9. Return levels of Y_t in Chiang Rai using the UGa-GEV model with

$$\hat{\mu} = 18.545 + 1.300PM_{10},\ \hat{\sigma} = \exp(3.313 + 0.010t), \hat{\xi} = 0.363, \text{ and } \hat{\theta} = 0.170.$$

Table 6 shows that the non-stationary UGa-GEV model with $\hat{\mu} = 18.545 + 1.300PM_{10}$, $\hat{\sigma} = \exp(3.313 + 0.010t), \hat{\xi} = 0.363,$ and $\hat{\theta} = 0.170$ was the optimal model for Y_t in Chiang Rai. The estimated model of the return level Y_t was

$$\hat{z}_{t,p} = \hat{\mathrm{M}}(\mathbf{x}) + \frac{\hat{\Sigma}(t)}{0.363}\left(1 - \left[-\log\left(\frac{0.170}{-2 - W\left\{-(1-p)(2.170)\exp(-2.170)\right\}}\right)\right]^{-\xi}\right), \quad p = \frac{1}{T} \qquad (20)$$

where $\hat{\mathrm{M}}(\mathbf{x}) = 18.545 + 1.300x_{t1}, \hat{\Sigma}(t) = \exp(3.313 + 0.010t)$ for $t = 1, 2, ..., 60.$ The return level of Y_t increased with the return period T (Figure 9). The return levels of Y_t concentration in 2 years, 5 years, 10 years, and 15 years were 175.24, 191.12, 210.56, and 225.20, respectively.

4. CONCLUSIONS

This study proposed stationary and non-stationary BM models via a novel UGa-GEV distribution. as an alternative for BM analysis. Some properties of the UGa-GEV distribution were introduced. The MLE method was used to estimate the parameters of models. The non-stationary maxima models used covariates with meteorology in location parameter μ, and time in μ and σ. The proposed models were used for extreme value analysis and risk analysis of the monthly maximum PM2.5 value between 2020 and 2024 in Bangkok and Chiang Rai, Thailand. The common variables affecting PM2.5 in Bangkok were PM10 and ozone, while only ozone had an equal effect on PM2.5 in Chiang Rai. Analysis of the recurrence rate of PM2.5 in both provinces revealed that PM2.5 levels of return period increased. Our findings contribute to the development of robust strategies for PM2.5 risk management, including early warning and mitigation. Future research could expand this scope by modeling the impact of extreme rainfall on particulate matter in specific regions, or by investigating the interplay between PM levels, precipitation, and temperature across diverse geographic areas.

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 (FRB69E0622)(Grant No.:FRB690069/0168). 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 interest.

REFERENCES

[1] R.A. Fisher, L.H.C. Tippett, Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample, Math. Proc. Cambridge Philos. Soc. 24 (1928), 180-190. https://doi.org/10.1017/S0305004100015681.

[2] B. Gnedenko, Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire, Ann. Math. 44 (1943), 423. https://doi.org/10.2307/1968974.

[3] S. Coles, An Introduction to Statistical Modeling of Extreme Values, Springer, 2001. https://doi.org/10.1007/978-1-4471-3675-0.

[4] N. Bezak, M. Brilly, M. Sraj, Comparison between the Peaks-Over-Threshold Method and the Annual Maximum Method for Flood Frequency Analysis, Hydrol. Sci. J. 59 (2014), 959-977. https://doi.org/10.1080/02626667.2013.831174.

[5] K. Klinjan, T. Sottiwan, S. Aryuyuen, Extreme Value Analysis with New Generalized Extreme Value Distributions: A Case Study for Risk Analysis on PM2.5 and PM10 in Pathum Thani, Thailand, Commun. Math. Biol. Neurosci. 2024 (2024), 100. https://doi.org/10.28919/cmbn/8833.

[6] S. Zhou, Q. Deng, W. Liu, Extreme Air Pollution Events: Modeling and Prediction, J. Cent. South Univ. 19 (2012), 1668-1672. https://doi.org/10.1007/s11771-012-1191-2.

[7] 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.

[8] J. Pornsopin, P. Busababodhin, T. Phoophiwfa, M. Chiangpradit, P. Guayjarernpanishk, 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.

[9] W. Warsono, Y. Antonio, S.B. Yuwono, D. Kurniasari, E. Suroso, et al., Modeling Generalized Statistical Distributions of PM2.5 Concentrations during the COVID-19 Pandemic in Jakarta, Indonesia, Decis. Sci. Lett.

10 (2021), 393-400. https://doi.org/10.5267/j.dsl.2021.1.005.

[10] A.I. Aguirre-Salado, S. Venancio-Guzmán, C.A. Aguirre-Salado, A. Santiago-Santos, 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.

[11] 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.

[12] T. Soomere, M. Eelsalu, K. Pindsoo, Variations in Parameters of Extreme Value Distributions of Water Level along the Eastern Baltic Sea Coast, Estuar. Coast. Shelf Sci. 215 (2018), 59-68. https://doi.org/10.1016/j.ecss.2018.10.010.

[13] K. Viigand, M. Eelsalu, T. Soomere, Quantifying Exposedness of the Eastern Baltic Sea Shores with Respect to Extremely High and Low Water Levels, Estuar. Coast. Shelf Sci. 319 (2025), 109267. https://doi.org/10.1016/j.ecss.2025.109267.

[14] N. Kudryavtseva, T. Soomere, R. Männikus, Non-Stationary Analysis of Water Level Extremes in Latvian Waters, Baltic Sea, during 1961–2018, Nat. Hazards Earth Syst. Sci. 21 (2021), 1279-1296. https://doi.org/10.5194/nhess-21-1279-2021.

[15] T. Prahadchai, P. Busababodhin, Y. Shin, Y. Shin, Model Averaging of Nonstationary Extreme Value Models, Applied to Maximum Precipitation in Thailand, Commun. Stat. Appl. Methods 32 (2025), 455-473. https://doi.org/10.29220/CSAM.2025.32.4.455.

[16] K. Viigand, M. Eelsalu, T. Soomere, Inherent Non-Stationarity in the GEV Distribution for Extreme Sea Levels: Implications for Coastal Vulnerability in the Baltic Sea, Ocean Eng. 341 (2025), 122681. https://doi.org/10.1016/j.oceaneng.2025.122681.

[17] 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.

[18] 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.

[19] M. Ansari Esfeh, L. Kattan, W.H.K. Lam, R. Ansari Esfe, M. Salari, 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.

[20] Pollution Control Department, Air4Thai, Archived Data, http://air4thai.pcd.go.th/webV3/#/History.

[21] S. Ayuyuen, W. Bodhisuwan, A generating family of unit-Garima distribution: Properties, likelihood inference, and application. Pak. J. Stat. Oper. Res. 20 (2024), 69-84. https://doi.org/10.18187/pjsor.v20i1.4307.

[22] 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

[23] D. Veberič, Lambert W Function for Applications in Physics, Comput. Phys. Commun. 183 (2012), 2622-2628. https://doi.org/10.1016/j.cpc.2012.07.008.

[24] A.F. Jenkinson, The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements, Q. J. R. Meteorol. Soc. 81 (1955), 158-171. https://doi.org/10.1002/qj.49708134804.

[25] S. Kotz, S. Nadarajah, Extreme Value Distribution: Theory and Applications, Imperial College Press, 2000.

[26] R.W. Katz, M.B. Parlange, P. Naveau, Statistics of Extremes in Hydrology, Adv. Water Resour. 25 (2002), 1287-1304. https://doi.org/10.1016/S0309-1708(02)00056-8.

[27] I. Parseh, H. Teiri, Y. Hajizadeh, K. Ebrahimpour, Phytoremediation of Benzene Vapors from Indoor Air by Schefflera arboricola and Spathiphyllum wallisii Plants, Atmos. Pollut. Res. 9 (2018), 1083-1087. https://doi.org/10.1016/j.apr.2018.04.005.

[28] A. Adler, lamW: Lambert-W Function, CRAN, 2015. https://doi.org/10.32614/CRAN.package.lamW.

[29] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, 2026.