Songklanakarin J. Sci. Technol.
44 (6), 1450-1461, Nov. – Dec. 2022

http://www.sjst.psu.ac.th

Original Article

The Topp-Leone generalized extreme value distribution: Extreme value analysis and return level estimation of the PM2.5 in Chiang Mai, Thailand

Sirinapa Aryuyuen[1], and Winai Bodhisuwan[2*]

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

[2] *Department of Statistics, Faculty of Science, Kasetsart University, Chatuchak, Bangkok, 10900 Thailand*

Received: 14 June 2022; Revised: 18 September 2022; Accepted: 20 October 2022

Abstract

In this paper, an extension of the generalized extreme value (GEV) distribution called the Topp Leone-GEV (TL-GEV) distribution is applied. The TL-GEV distribution has four parameters (λ, μ, σ, ξ), and it has the three named sub-models TL-Gumbel (for $\xi = 0$), TL-Fréchet (for $\xi > 0$), and TL-reversed Weibull (for $\xi < 0$). Its properties and maximum likelihood estimation are discussed. A data set was used to demonstrate the efficiency of the proposed distribution. The TL-GEV distribution was employed for fitting the data and compared to some selected distributions. The two datasets represented PM2.5 data for Chiang Mai Province (Tambon Sri Phum and Tambon Chang Phueak), in Thailand. According to the Kolmogorov-Smirnov test, Akaike Information Criterion, and Bayesian Information Criterion, the TL-GEV distribution for $\xi > 0$ or TL-Fréchet distribution can be considered competitive. The TL-GEV distribution is an alternative flexible way to analyze any extreme values and to estimate the return level, because the additional parameter λ provides flexibility to the distribution affecting its skewness and kurtosis.

Keywords: extreme value theory, T-X family of distributions, topp leone-generalized extreme value distribution, PM2.5, return level

1. Introduction

Probability modelling of continuous data plays an essential role in many fields, for instance in engineering, medicine, biological science, management, and public health. Probability distributions may provide helpful information supporting conclusions and decisions. When there is a need for more cover and flexible distributions, many researchers will use the new ones that are more general. There are several methods for generating families of continuous distributions, such as differential equations, transformations, quantile functions, and adding extra parameters or combining existing distributions. Recently, applying new generators for continuous distributions has become more interesting. Some recent developments can improve the goodness of fit and determine tail properties. In general, many newly developed distributions mainly add more flexibility to existing distributions, which result from implanting a primary distribution into a more capable structure (Hamed & Alzaghal, 2021; Lee, Famoye & Alzaatreh, 2013). These features have been established by the results of many generators, such as beta-X family (Eugene, Lee, & Famoye, 2002), transformed transformer (T-X) family (Alzaatreh, Lee, & Famoye, 2013), and Topp Leone-generated (TL-G) family (Sangsanit & Bodhisuwan, 2016) of distributions. The TL-G family of distributions was proposed by Sangsanit and Bodhisuwan

*Corresponding author
Email address: fsciwnb@ku.ac.th

(2016). They show the Topp Leone-generalized exponential distribution to be an example of the TL-G distribution. Its density function has flexible shapes, such as decreasing function or unimodal shape that is right-tailed. It has a decreasing hazard function, or a constant or increasing hazard function. This study considers the TL-G distribution family to contribute a new model to analyze data.

One of the most outstanding modelling achievements is finding an optimal model for the data. If the extreme values are included, an analyst usually cuts them out from the data because of problems with complexity. In practice, if an analyst wants to know the extreme event's probability, the extreme values are in the tails. A popular statistical tool for extreme value analysis is the so-called "Extreme Value Theory". Typically, most analysts who analyze data where extreme values occur discard that portion of the data from consideration. But, in fact, often we want to know the probability of events with the largest or smallest value, which normally are located at the ends of the distribution tails. Some examples are the highest or lowest rainfall in the month, the highest wind speed, daily highest or lowest temperature, etc. Statistical analysis is used as a tool in decision-making and for finding ways to prevent and resolve various subsequent situations, such as droughts, floods, storms, earthquakes, etc., based on extreme value theory (Coles, 2001; Kotx & Nadaraja, 2000). The extreme value theory is a branch of statistics that focuses on the extreme events and especially the tail behaviors of a distribution. The theory uses the block maxima approach to derive extreme value distributions, including the Fréchet, Weibull, and Gumbel distributions. In addition, the generalized extreme value (GEV) distribution is developed within the extreme value theory to combine the Gumbel, Fréchet and Weibull distributions. The GEV distribution has been widely used.

The fine particulate matter (PM2.5) is one air pollutant that affects people's health when its levels are very high. The PM2.5 are tiny particles in the air that reduce visibility and cause the air to appear hazy when the levels are elevated. Outdoor PM2.5 levels are most likely to be elevated on days with little or no wind or air mixing. Since the last decade or so, the PM2.5 in Chiang Mai, Thailand has been dramatically increasing with an average of 54.14 ug/m^3 (ug/m^3: micrograms per cubic meter), and a maximum value of 266.00 ug/m^3, which represents an extreme value. There are much research and applications based on the extreme value theory in various fields, especially in hydrology (Castillo, Hadi, Balakrishnan, & Sarabia, 2004; Guloksuz & Celik, 2020; Nadarajah & Pogány, 2013; Shukla, Trivedi, & Kumar, 2012).

Coverage and flexibility are areas of great importance for developing quality improvement of the statistical model. A statistical model is an important and necessary tool in data analysis. The GEV distribution may prove somewhat inadequate in practice, and generalizations ought to provide greater flexibility for data modelling purposes. For example, extensions of the GEV distribution have been studied, such as the uniform-GEV distribution by Guloksuz and Celik (2020). It is used to analyze earthquake data. The results show that the uniform-GEV is a better fit than the GEV distribution.

This paper aims to propose new modifications to GEV models that incorporate an additional parameter, hoping that this will yield better results in certain practical situations. We will create a new extended GEV distribution for analyzing extreme values. The proposed distribution is developed using the TL-G family of distributions proposed by Sangsanit and Bodhisuwan (2016). This TL-G family is obtained using the method for generating a family of distributions, that is, the T-X family, which was proposed by Alzaatreh *et al.* (2013). The article is organized as follows. In Section 2, some methods are introduced. Some results and discussion are provided in Section 3, including a new extended GEV distribution, the proposed model parameter estimation, and a simulation study. Furthermore, the example of extreme value analysis including return level estimation of rainfall data in Thailand, is demonstrated. Finally, some conclusions are provided in the last section.

2. Materials and Methods

In this section, the GEV distribution, the T-X family of distributions, and the Topp-Leone distribution are introduced to derive the new extended GEV distribution.

2.1 The GEV distribution

Let X be a random variable which is distributed as the GEV distribution with parameters μ, σ and ξ, denoted by X ~GEV (μ, σ, ξ). The cumulative density function (cdf) and the probability density function (pdf) of X are respectively

$$G_{GEV}(x) = \begin{cases} \exp\left(-\left(1+\xi z\right)^{-1/\xi}\right), & \xi \neq 0, \\ \exp\left(-e^{-z}\right), & \xi = 0, \end{cases} \tag{1}$$

$$g_{GEV}(x) = \begin{cases} \dfrac{1}{\sigma}\left(1+\xi z\right)^{-(1+1/\xi)} \exp\left(-\left(1+\xi z\right)^{-1/\xi}\right), & \xi \neq 0, \\ \dfrac{1}{\sigma} e^{-z} \exp\left(-e^{-z}\right), & \xi = 0, \end{cases} \tag{2}$$

where $z = (x-\mu)/\sigma$ and parameters μ, σ and ξ represent location, scale, and shape, respectively, with $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. The corresponding quantile function is explicitly

$$Q_{GEV}(p) = \begin{cases} \mu + \dfrac{\sigma}{\xi}\left[\left(-\log(p)\right)^{-\xi} - 1\right], & \xi \neq 0, \\ \mu - \sigma \log\left(-\log(p)\right), & \xi = 0, \end{cases} \tag{3}$$

where $0 < p < 1$. The GEV distributions can be represented by a single shape parameter which controls the tail behavior. Its mean and variance are respectively,

$$E(X) = \mu - \frac{\sigma}{\xi} + \frac{\sigma}{\xi}\Gamma\left(1-\xi\right),$$

$$Var(X) = \frac{\sigma^2}{\xi^2}\left[\Gamma\left(1-2\xi\right) - \left(\Gamma\left(1-\xi\right)\right)^2\right],$$

where $\Gamma(\cdot)$ is the complete gamma function. The class of the GEV distribution has three sub-models (Bali, 2003), namely

(i) Gumbel (Gum) or type I extreme value distribution ($\xi = 0$):

$$G_{Gum}(x) = \exp\left(-e^{-z}\right), \text{ and } g_{Gum}(x) = \frac{1}{\sigma} e^{-z} \exp\left(-e^{-z}\right), \tag{4}$$

where $-\infty < x < \infty$ and $z = (x-\mu)/\sigma$.

(ii) Fréchet (Fr) or type II extreme value distribution ($\xi > 0$):

$$G_{Fr}(x) = \exp\left(-y^{-1/\xi}\right), \text{ and } g_{Fr}(x) = \frac{1}{\sigma} y^{-1/\xi-1} \exp\left(-y^{-1/\xi}\right), \tag{5}$$

where $y > 0$, $y = 1 + \xi z$ and $z = (x-\mu)/\sigma$.

(iii) Reversed Weibull (RW) or type III extreme value distribution ($\xi < 0$):

$$G_{RW}(x) = \exp\left(-(-y^{-1/\xi})\right), \text{ and } g_{RW}(x) = \frac{1}{\sigma} y^{-1/\xi-1} \exp\left(-(-y^{-1/\xi})\right), \tag{6}$$

where $y < 0$, $y = -(1+\xi z)$ and $z = (x-\mu)/\sigma$.

2.2 The T-X family of distributions

We consider the method for generating families of continuous distributions which was proposed by Alzaatreh *et al.* (2013). This method is transforming a random variable X to T, for $T \in [a,b]$ and $-\infty < a < b < \infty$, through the function $W[G(x)]$. The generated families of this method are called the T-X family of distributions. Its cdf and pdf are respectively

$$F_{T\text{-}X}(x) = \int_a^{W[G(x)]} r(t)dt\,, \tag{7}$$

$$f_{T\text{-}X}(x) = \left\{\frac{d}{dx} W[G(x)]\right\} r\left\{W[G(x)]\right\}, \tag{8}$$

where $r(t)$ and $G(x)$ are respectively, the pdf and cdf of parent distribution. $W[G(x)]$ is a function of $G(x)$, which satisfies the following conditions: (i) $W[G(x)] \in [a,b]$, (ii) $W[G(x)]$ is differentiable and monotonically non-decreasing, and (iii) $W[G(x)] \to a$ as $x \to -\infty$ and $W[G(x)] \to b$ as $x \to \infty$.

2.3 The Topp-Leone family of distributions

If a random variable T is distributed as the Topp Leone (TL) distribution with a parameter $\lambda > 0$, denoted by T~TL (λ), it has pdf $r(t) = 2\lambda t^{\lambda-1}(1-t)(2-t)^{\lambda-1}$, $0 < t < 1$. Let $W[G(x)] = G(x)$, where $G(x)$ is a cdf of any specified random variable X , and the TL distribution is a parent distribution. This generated family is called the Topp Leone-generated (TL-G) family of distributions, and was introduced by Sangsanit and Bodhisuwan (2016). Its cdf and pdf are respectively,

$$F_{TL\text{-}G}(x) = G^{\lambda}(x)\left[2 - G(x)\right]^{\lambda}, \tag{9}$$

$$f_{TL\text{-}G}(x) = 2\lambda g(x)\left[1 - G(x)\right] G^{\lambda-1}(x)\left[2 - G(x)\right]^{\lambda-1}, \tag{10}$$

where $g(x)$ is a pdf of X, which is distributed as a parent distribution. The quantile function of the TL-G family of distributions is

$$Q_{TL\text{-}G}(u) = Q_G\left[1 - (1 - u^{1/\lambda})^{1/2}\right], \tag{11}$$

where $0 < u < 1$, and $Q_G(u)$ is the quantile function of a specified parent distribution.

3. Results and Discussion

In this section, a new extended GEV distribution and its properties are discussed. Parameters of the proposed distribution are estimated using the maximum likelihood (ML) method, and its simulation is studied. An extreme value analysis and return level estimation of the PM2.5 in Chiang Mai, Thailand, are illustrated.

3.1 A new extended GEV distribution

In this section, we will develop a new extended GEV distribution based on the TL-G family with the GEV distribution as a parent distribution, namely the Topp Leone-generalized extreme value (TL-GEV) distribution.

Theorem 1. If X be a random variable which is distributed as the TL-GEV distribution with the parameters λ, σ, μ, and ξ, denoted by X~TL-GEV(λ, μ, σ, ξ), then its cdf and pdf are respectively

$$F_{\text{TL-GEV}}(x) = e^{-\lambda\tau(x)}\left[2 - e^{-\tau(x)}\right]^{\lambda}, \tag{12}$$

$$f_{\text{TL-GEV}}(x) = \frac{2\lambda}{\sigma}\left(1 - e^{-\tau(x)}\right)\left(2 - e^{-\tau(x)}\right)^{\lambda-1} e^{-\lambda\tau(x)}\tau(x)^{\xi+1}, \tag{13}$$

where $z = (x-\mu)/\sigma$, parameters $\lambda > 0$, $\sigma > 0$, $-\infty < \mu < \infty$, and $-\infty < \xi < \infty$, and

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

Proof. Based on the method of T-X family of distributions (Alzaatreh *et al.*, 2013), the cdf of the TL-G distribution (Sangsanit & Bodhisuwan, 2016) is obtained by replacing the pdf as $r(t) = 2\lambda t^{\lambda-1}(1-t)(2-t)^{\lambda-1}$, and $W\left[G(x)\right] = G(x)$ in Equation (7):

$$F_{\text{TL-X}}(x) = \int_0^{G(x)} 2\lambda t^{\lambda-1}(1-t)(2-t)^{\lambda-1}dt = G^{\lambda}(x)\left[2 - G(x)\right]^{\lambda}.$$

By replacing the GEV's cdf in Equation (1) as in above equation $F_{\text{TL-X}}(x)$, we have the cdf of the TL-GEV distribution as follows:

$$\begin{aligned} F_{\text{TL-GEV}}(x) &= e^{-\lambda\tau(x)}\left[2 - e^{-\tau(x)}\right]^{\lambda} \\ &= \begin{cases} \left[2 + \exp\left(-\left(1+\xi z\right)^{-1/\xi}\right)\right]^{\lambda} \exp\left(-\lambda\left(1+\xi z\right)^{-1/\xi}\right), & \xi \neq 0, \\ \left[2 - \exp\left(-e^{-z}\right)\right]^{\lambda} \exp\left(-\lambda e^{-z}\right), & \xi = 0. \end{cases} \end{aligned}$$

Its corresponding pdf is

$$\begin{aligned} f_{\text{TL-GEV}}(x) &= \frac{d}{dx}\left\{e^{-\lambda\tau(x)}\left[2 - e^{-\tau(x)}\right]^{\lambda}\right\} \\ &= \frac{2\lambda}{\sigma}\left(1 - e^{-\tau(x)}\right)\left(2 - e^{-\tau(x)}\right)^{\lambda-1} e^{-\lambda\tau(x)}\tau(x)^{\xi+1}. \end{aligned}$$

The class of the TL-GEV distributions can be represented by the shape parameter which controls the tail behavior. It has three sub-models as follows:

(i) If $\xi = 0$, the TL-GEV distribution reduces to the TL-Gumbel (TL-Gum) distribution with the cdf and pdf as

$$F_{\text{TL-Gum}}(x) = \left[2 - \exp\left(-e^{-z}\right)\right]^{\lambda} \exp\left(-e^{-\lambda z}\right), \tag{14}$$

$$\begin{aligned} f_{\text{TL-Gum}}(x) &= \frac{2\lambda}{\sigma}\left[1 - \exp\left(-e^{-z}\right)\right] e^{-z} \exp\left(-e^{-z}\right) \\ &\quad \times \left[2 - \exp\left(-e^{-z}\right)\right]^{\lambda-1} \exp\left(-(\lambda-1)e^{-z}\right), \end{aligned} \tag{15}$$

where $-\infty < x < \infty$ and $z = (x-\mu)/\sigma$. The TL-Gum and Gum distributions are right-tailed. The TL-Gum distribution is more long-tailed (heavy-tailed) than the Gum distribution. The TL-Gum curve has more kurtosis when λ is large (Figure 1).

(ii) If $\xi > 0$, the TL-GEV distribution reduces to the TL-Fréchet (TL-Fr) distribution with the cdf and pdf as

$$F_{\text{TL-Fr}}(x) = \left[2 - \exp\left(-y^{-1/\xi}\right)\right]^{\lambda} \exp\left(-\lambda y^{-1/\xi}\right), \tag{16}$$

$$\begin{aligned} f_{\text{TL-Fr}}(x) &= \frac{2\lambda}{\sigma} y^{-(1+1/\xi)}\left[1 - \exp\left(-y^{-1/\xi}\right)\right]\exp\left(-y^{-1/\xi}\right) \\ &\quad \times \left[2 - \exp\left(-y^{-1/\xi}\right)\right]^{\lambda-1} \exp\left(-(\lambda-1)y^{-1/\xi}\right), \end{aligned} \tag{17}$$

where $y = 1 + \xi z$ and $z = (x-\mu)/\sigma$. The TL-Fr and Fr distributions are right-tailed. The TL-Fr distribution is more long-tailed (heavy-tailed) than the Fr distribution. The curve of the TL-Fr has more kurtosis when λ is large (Figure 2).

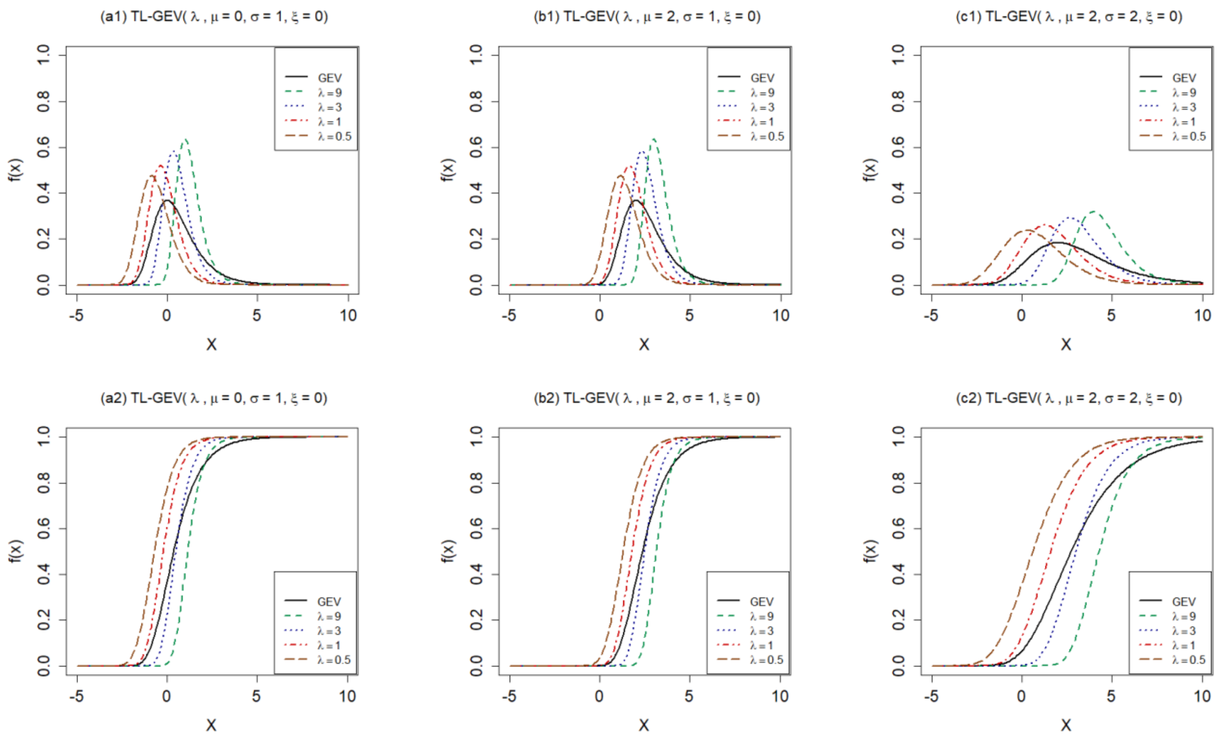

Figure 1. The pdf plots of the GEV and TL-GEV distributions with $\xi = 0$

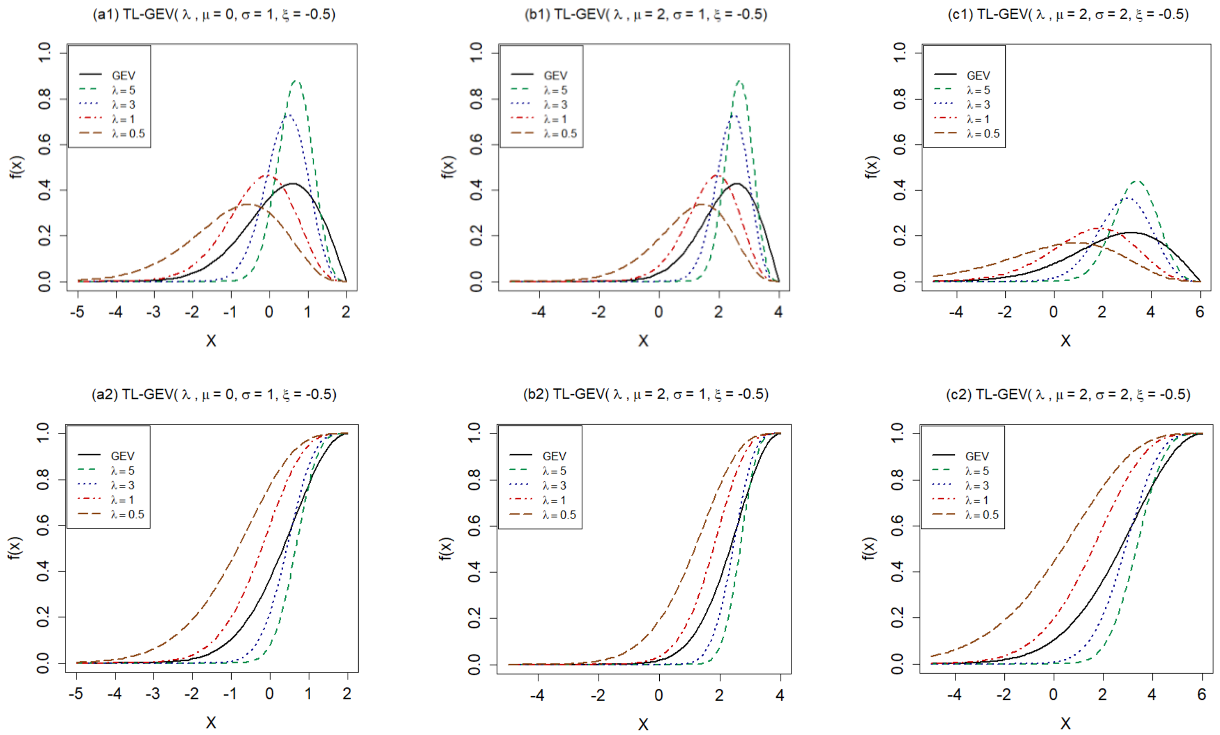

Figure 2. The pdf plots of the GEV and TL-GEV distributions with $\xi > 0$

(iii) If $\xi < 0$, the TL-GEV reduces to the TL-reversed Weibull (TL-RW) distribution with the cdf and pdf as

$$F_{TL\text{-}RW}(x) = \left[2 - \exp\left(-(-y^{-1/\xi})\right)\right]^{\lambda} \exp\left(-\lambda(-y^{-1/\xi})\right), \tag{18}$$

$$f_{TL\text{-}RW}(x) = \frac{2\lambda}{\sigma}\left[1 - \exp\left(-(-y^{-1/\xi})\right)\right] y^{-(1+1/\xi)} \exp\left(-(-y^{-1/\xi})\right) \times \left[2 - \exp\left(-(-y^{-1/\xi})\right)\right]^{\lambda-1} \exp\left(-(\lambda-1)(-y^{-1/\xi})\right), \tag{19}$$

where $y = -(1+\xi z)$ and $z = (x-\mu)/\sigma$. The TL-RW and RW distributions are left-tailed. The TL-RW distribution is less fat-tailed than the RW distribution when $\lambda \geq 3$, and it is more long-tailed than the RW distribution for $\lambda < 3$ (Figure 3).

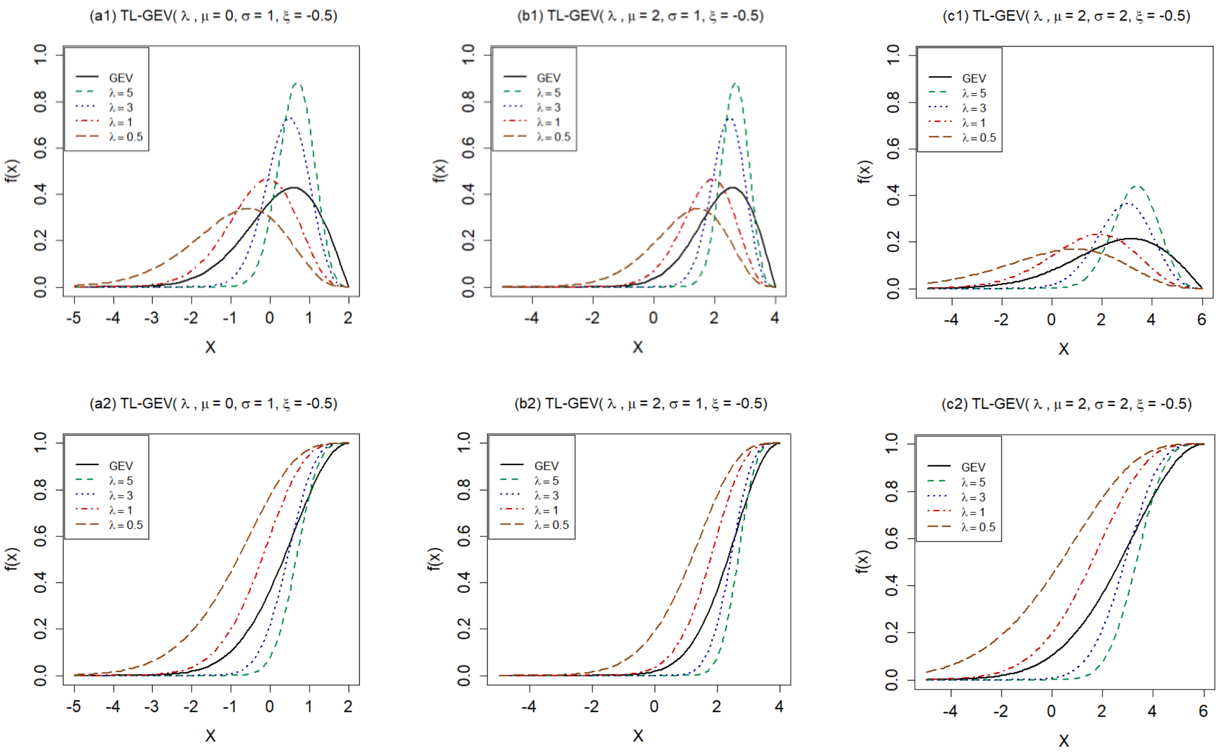

Figure 3. The pdf plots of the GEV and TL-GEV distributions with $\xi < 0$

3.2 Statistical properties of the TL-GEV distribution

In this section, expressions for the quantile function, moments, moment generating function, and order statistics, are derived.

3.2.1 Quantile function

If X is distributed as TL-GEV then its quantile function is

$$Q_{TL\text{-}GEV}(u) = \begin{cases} \mu + \dfrac{\sigma}{\xi}\left\{\left[-\log\left(1-\sqrt{1-u^{1/\lambda}}\right)\right]^{-\xi} - 1\right\}, & \xi \neq 0, \\ \mu - \sigma \log\left[-\log\left(1-\sqrt{1-u^{1/\lambda}}\right)\right], & \xi = 0, \end{cases} \tag{20}$$

where u is a value of a uniform random variable in the interval (0,1).

An estimate of the extreme quantiles of the annual maximum are obtained by $Q_{TL\text{-}GEV}(u)$. Let $F_{TL\text{-}GEV}(x_p)=1-p$, and x_p is called return level with the return period. This means that x_p is exceeded by the annual maximum in any particular year with probability p (Coles, 2001). For p=1/m, m is called a return period. The return level for the TL-GEV distribution is

$$x_p = \begin{cases} \mu + \dfrac{\sigma}{\xi}\left\{\left[-\log\left(\sqrt{1-p^{1/\lambda}}\right)\right]^{-\xi} - 1\right\}, & \xi \neq 0, \\ \mu - \sigma\log\left[-\log\left(\sqrt{1-p^{1/\lambda}}\right)\right], & \xi = 0. \end{cases} \quad (21)$$

3.2.2 Moments

The moments of the TL-G family of distribution (Sangsanit & Bodhisuwan, 2016) are

$$E(X^s) = \sum_{j,k=0}^{\infty}\sum_{r=0}^{k}\binom{\lambda}{j}\binom{\lambda+j}{k}\binom{k}{r} r(-1)^{j+k+r}\nu_{s,r-1}, \text{ for } \nu_{s,r} = \int_0^1 u^r Q_G^s(u)du$$

where $\lambda > 0$ and $\nu_{s,r}$ is (s,r)*th* probability weighted moment of X (Greenwood *et al.*, 1979). From the quantile function in Equation (20), the moment of the TL-GEV distribution is

$$E(X^s) = \sum_{j,k=0}^{\infty}\sum_{r=0}^{k}\binom{\lambda}{j}\binom{\lambda+j}{k}\binom{k}{r}(-1)^{j+k+r} r$$
$$\times \begin{cases} \int_0^1 \left(1-\sqrt{1-u^{1/\lambda}}\right)^r \left\{\mu + \dfrac{\sigma}{\xi}\left\{\left[-\log\left(1-\sqrt{1-u^{1/\lambda}}\right)\right]^{-\xi} - 1\right\}\right\}^s du, & \xi \neq 0, \\ \int_0^1 \left(1-\sqrt{1-u^{1/\lambda}}\right)^r \left\{\mu - \sigma\log\left[-\log\left(1-\sqrt{1-u^{1/\lambda}}\right)\right]\right\}^s du, & \xi = 0. \end{cases}$$

where $\lambda > 0$, $\sigma > 0$, $-\infty < \mu < \infty$, and $-\infty < \xi < \infty$.

3.2.3 Moment generating function

The moment generating function (mgf) of the TL-G family of distributions (Sangsanit & Bodhisuwan, 2016) is

$$E(e^{tX}) = \sum_{j=0}^{\lambda}\binom{\lambda}{j}(\lambda+j)(1)^j 2^{\lambda-1}\int_0^1 e^{tQ_G(u)}u^{\lambda+j-1}Q_G(u)du, \text{ for } \lambda > 0.$$

From the quantile function in Equation (20), the mgf of the TL-GEV distribution is

$$M(t) = \sum_{j=0}^{\lambda}\mathbf{b}_{\lambda,j}\int_0^1\left(1-\sqrt{1-u^{1/\lambda}}\right)^{\lambda+j-1}\exp\left\{t\left[\mu + \frac{\sigma}{\xi}\left\{\left[-\log\left(1-\sqrt{1-u^{1/\lambda}}\right)\right]^{-\xi} - 1\right\}\right]\right\}$$
$$\times\left\{\mu + \frac{\sigma}{\xi}\left\{\left[-\log\left(1-\sqrt{1-u^{1/\lambda}}\right)\right]^{-\xi} - 1\right\}\right\}du, \quad \text{for } \xi \neq 0,$$

$$M(t) = \sum_{j=0}^{\lambda}\mathbf{b}_{\lambda,j}\int_0^1\left(1-\sqrt{1-u^{1/\lambda}}\right)^{\lambda+j-1}\exp\left\{t\left[\mu - \sigma\log\left[-\log\left(1-\sqrt{1-u^{1/\lambda}}\right)\right]\right]\right\}$$
$$\times\left\{\mu - \sigma\log\left[-\log\left(1-\sqrt{1-u^{1/\lambda}}\right)\right]\right\}du, \quad \text{for } \xi = 0,$$

where $\mathbf{b}_{\lambda,j} = (\lambda+j)\binom{\lambda}{j}(1)^j 2^{\lambda-1}$ for $\lambda > 0$.

3.2.4 Order statistics

Let $X_1, X_2, X_3, ..., X_n$ be a random sample of size n from the TL-GEV distribution. Then the pdf of the i*th* order statistic, $1 \le i \le n$, can be obtained as follows

$$f_{i:n}(x) = \frac{f_{TL-GEV}(x)}{B(i, n-i-1)}F_{TL-GEV}^{i-1}(x)\left[1 - F_{TL-GEV}(x)\right]^{n-i},$$

where $B(a,b)$ is a beta function. Finally, we have the $f_{i:n}(x)$ as

$$f_{i:n}(x) = \frac{2\lambda}{\sigma B(i, n-i-1)}\left(1-e^{-\tau(x)}\right)\left(2-e^{-\tau(x)}\right)^{\lambda-1}e^{-\lambda\tau(x)}\tau(x)^{\xi+1}$$
$$\times\left[e^{-\lambda\tau(x)}\left(2-e^{-\tau(x)}\right)^{\lambda}\right]^{i-1}\left[1-e^{-\lambda\tau(x)}\left(2-e^{-\tau(x)}\right)^{\lambda}\right]^{n-i}.$$

The moments and mgf of the TL-GEV distribution are not available in closed form. Thus, the expressions for mean, standard deviation, skewness, and kurtosis are also not in closed form. However, these values are shown in a simulation study as follows.

Some simulations were performed for illustration purposes of how the skewness of the TL-GEV distribution changes with different λ values, and for comparison to the GEV distribution. Random samples with size 1,000 were generated 1,000 times for the GEV and TL-GEV distributions using Equation (3) and Equation (20), respectively, that is, X~TL-GEV(λ, μ, σ, ξ) Y~GEV(μ, σ, ξ). The results, including sample mean $(\bar{X})$, standard deviation (SD), skewness (SV), and kurtosis (KV), when μ = 2, σ = 2, and ξ = 0, 0.5-0.5, are shown in Table 1. Skewness can be used as a measure of the symmetry of distribution, where SV=0 for a symmetrical distribution. Meanwhile, the KV is often compared to the kurtosis of the normal distribution, which has KV=3. If KV>3, then the dataset has heavier tails than a normal distribution, and the data have lighter tails than a normal distribution when KV<3.

The results in Table 1 show that the parameter λ provides flexibility to the TL-GEV distribution. The parameter λ affects skewness and kurtosis of the TL-GEV distribution. For $\xi \geq 0$, the random variables for TL-GEV and GEV distributions are right-skewed (SV>0) and heavier tailed (KV>3) than a normal distribution. But the TL-GEV distribution provides both SV and KV below those of the GEV distribution. The proposed distribution has the SV and KV increase as λ increases. For ξ<0, the random variables for TL-GEV and GEV distributions are left-skewed (SV<0) and heavy-tailed (KV>3) for λ<3. But the TL-GEV distribution provides KV less than 3 for $\lambda \geq 3$, which indicates that it has lighter tails than the normal distribution.

Table 1. The simulation results for the GEV and TL-GEV distributions (Dist.) with μ=2, σ=2, and ξ=0, 0.5, -0.5

Dist.	λ	$\xi=0$				$\xi=0.5$				$\xi=-0.5$			
		$\bar{X}$	SD	SV	KV	$\bar{X}$	SD	SV	KV	$\bar{X}$	SD	SV	KV
GEV	-	3.15	2.56	1.13	5.35	5.07	10.32	9.95	174.05	2.46	1.85	-0.63	3.23
TL-GEV	0.5	0.78	1.78	0.65	3.79	1.29	1.87	3.54	31.96	0.07	2.42	-0.61	3.36
	1	1.77	1.65	0.71	4.04	2.15	2.20	3.57	32.19	1.41	1.73	-0.50	3.23
	1.5	2.31	1.59	0.76	4.15	2.72	2.41	3.58	32.30	2.02	1.44	-0.43	3.14
	3	2.68	1.56	0.79	4.26	3.86	2.84	3.61	32.94	2.83	1.08	-0.31	2.99
	5	3.81	1.46	0.88	4.53	4.78	3.20	3.66	33.88	3.30	0.89	-0.24	2.92
	10	4.61	1.41	0.94	4.72	6.24	3.76	3.70	34.26	3.80	0.69	-0.16	2.86
	15	5.06	1.39	0.98	4.84	7.21	4.16	3.73	34.69	4.04	0.61	-0.12	2.82
	30	5.82	1.36	1.01	4.94	9.11	4.93	3.79	35.33	4.38	0.49	-0.06	2.79
	50	6.37	1.34	1.03	4.97	10.71	5.55	3.81	35.89	4.59	0.42	-0.03	2.77

3.3 The parameter estimation of the TL-GEV distribution

Let $\tilde{x} = (x_1, \ldots, x_n)$ be observations of a random sample (X_1,…, X_n) such that X_i are independent and identically distributed random variables, of size n when X_i~TL-GEV (μ, σ, ξ, λ). According to the pdf in Equation (13), its corresponding likelihood functions are

Case 1: $\xi \neq 0$,

$$L_{\xi\neq 0} = 2^n \lambda^n \sigma^{-n} \prod_{i=1}^{n} \left\{ (1+\xi z_i)^{-1/\xi-1} \left[1-\exp\left\{-(1+\xi z_i)^{-1/\xi}\right\} \right] \exp\left\{-(1+\xi z_i)^{-1/\xi}\right\} \times \left[\exp\left\{-(1+\xi z_i)^{-1/\xi}\right\} \right]^{\lambda-1} \left[2-\exp\left\{-(1+\xi z_i)^{-1/\xi}\right\} \right] \right\},$$

where $z_i = (x_i - \mu)/\sigma$. Its log-likelihood function is

$$\ell(\mu,\sigma,\xi,\lambda) = n\log(2) + \sum_{i=1}^{n} \log\left[1-\exp\left\{-(1+\xi z_i)^{-1/\xi}\right\}\right] - \sum_{i=1}^{n} (1+\xi z_i)^{-1/\xi} - n\log\sigma - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\log(1+\xi z_i) + (\lambda-1)\sum_{i=1}^{n}\exp\left\{-(1+\xi z_i)^{-1/\xi}\right\} + n\log\lambda + \sum_{i=1}^{n}\log\left[2-\exp\left\{-(1+\xi z_i)^{-1/\xi}\right\}\right].$$

To estimate the unknown parameters, we take the partial derivatives with respect to each parameter, and equate them to zero, i.e.,

$$\frac{\partial \ell(\mu,\sigma,\xi,\lambda)}{\partial \mu} = 0, \quad \frac{\partial \ell(\mu,\sigma,\xi,\lambda)}{\partial \sigma} = 0, \quad \frac{\partial \ell(\mu,\sigma,\xi,\lambda)}{\partial \xi} = 0, \text{ and } \frac{\partial \ell(\mu,\sigma,\xi,\lambda)}{\partial \lambda} = 0.$$

The solutions of the ML estimators μ, σ, ξ, and λ, are obtained by using the **nlm** function in the R **stats** package (R Core Team, 2021).

Case 2: $\xi = 0,$

$$L_{\xi=0} = 2^n \lambda^n \sigma^{-n} \prod_{i=1}^{n} \left\{ \exp\left\{ -z_i(\xi+1) - e^{-z_i} \right\} \left[1 - \exp\left\{ -e^{-z_i} \right\} \right] \times \left[\exp\left\{ -e^{-z_i} \right\} \right]^{\lambda-1} \left[2 - \exp\left\{ -e^{-z_i} \right\} \right]^{\lambda-1} \right\}.$$

Its log-likelihood function is

$$\ell(\mu,\sigma,\lambda) = \log L_{\xi=0} = n\log(2) + n\log\lambda - n\log\sigma - \sum_{i=1}^{n} \left[-z_i(\xi+1) - e^{-z_i} \right]$$

$$+ \sum_{i=1}^{n} \log\left[1 - \exp\left\{ -e^{-z_i} \right\} \right] + (\lambda-1)\sum_{i=1}^{n} \log\left[\exp\left\{ -e^{-z_i} \right\} \right] + (\lambda-1)\sum_{i=1}^{n} \log\left[2 - \exp\left\{ -e^{-z_i} \right\} \right].$$

To estimate the unknown parameters, we take the partial derivatives with respect to each parameter, and equate them to zero, i.e.,

$$\frac{\partial \ell(\mu,\sigma,\lambda)}{\partial \mu} = 0, \quad \frac{\partial \ell(\mu,\sigma,\lambda)}{\partial \sigma} = 0, \text{ and } \frac{\partial \ell(\mu,\sigma,\lambda)}{\partial \lambda} = 0.$$

Since above equation cannot be derived in closed form, in the same manner as in Case 1 , the solutions of the ML estimators of μ, σ, and λ, are obtained by using the **nlm** function in the R **stats** package (R Core Team, 2021).

3.4 Simulation study for parameter estimation of distributions

Simulations are performed for different configurations of the GEV and TL-GEV distributions; (i) the GEV distribution, $Y_i = Q_{GEV}(u_i; \mu, \sigma, \xi)$ for fixed values of μ = 2 and σ = 2, including Gum(μ, σ, ξ = 0), Fr(μ, σ, ξ = 0.5), and RW (μ, σ, ξ = -0.5) distributions, and (ii) the TL-GEV distribution, $X_i = Q_{TL\text{-}GEV}(u_i; \lambda, \mu, \sigma, \xi)$, for fixed values of λ =1.5, μ = 2 and σ = 2, including the TL-Gum (λ, μ, σ, ξ=0), TL-Fr (λ, μ, σ, ξ=0.5), and TL-RW(λ, μ, σ, ξ=-0.5) distributions. Each case is considered by experiments at sizes of n set at 25, 50, 100, 200, 500, and 1,000 replications. The mean and mean square error (MSE) of ML estimators are shown in Table 2.

Table 2. The simulation results for parameter estimation of the GEV and TL-GEV distributions

ξ	n	Values	GEV			TL-GEV			
			μ = 2	σ = 1	ξ	λ = 1.5	μ = 2	σ = 1	ξ
0	25	Estimates	2.021	0.963	-	1.218	2.820	0.826	-
		MSE	0.044	0.025	-	8.608	1.400	0.061	-
	50	Estimates	2.011	0.984	-	1.544	2.598	0.880	-
		MSE	0.023	0.012	-	7.586	1.095	0.042	-
	100	Estimates	2.005	0.992	-	1.592	2.476	0.909	-
		MSE	0.011	0.006	-	5.156	0.940	0.031	-
	200	Estimates	2.001	0.995	-	1.634	2.263	0.950	-
		MSE	0.006	0.003	-	2.872	0.548	0.017	-
	500	Estimates	1.999	0.998	-	1.596	2.107	0.981	-
		MSE	0.002	0.001	-	1.153	0.235	0.007	-
0.5	25	Estimates	2.007	0.958	0.524	1.263	3.844	1.533	0.409
		MSE	0.063	0.057	0.064	10.332	5.679	1.129	0.077
	50	Estimates	2.008	0.976	0.514	1.410	3.688	1.449	0.402
		MSE	0.027	0.024	0.025	8.584	5.152	0.663	0.045
	100	Estimates	2.002	0.989	0.506	1.477	2.633	1.257	0.466
		MSE	0.013	0.013	0.012	4.562	1.315	0.218	0.015
	200	Estimates	1.999	0.995	0.505	1.573	2.242	1.084	0.488
		MSE	0.007	0.006	0.005	2.474	0.385	0.055	0.008
	500	Estimates	1.998	0.995	0.501	1.518	2.135	1.053	0.493
		MSE	0.003	0.002	0.002	1.312	0.205	0.029	0.003
-0.5	25	Estimates	2.052	1.042	-0.625	1.734	2.228	0.954	-0.435
		MSE	0.068	0.080	0.092	9.015	1.085	1.045	0.090
	50	Estimates	2.028	0.996	-0.537	1.709	2.138	0.934	-0.480
		MSE	0.024	0.017	0.015	5.467	0.587	0.341	0.027
	100	Estimates	2.046	1.010	-0.551	1.659	2.129	0.936	-0.485
		MSE	0.483	0.259	0.015	3.886	0.483	0.259	0.015
	200	Estimates	2.010	1.004	-0.516	1.558	2.146	0.925	-0.480
		MSE	0.009	0.006	0.007	2.749	0.404	0.188	0.009
	500	Estimates	2.007	1.001	-0.509	1.592	2.088	0.958	-0.485
		MSE	0.004	0.002	0.003	1.849	0.304	0.136	0.006

The results in Table 2, in all cases, show that each parameter's mean of ML estimates has a value close to the true parameter when the sample size is increased. The MSE of ML estimators decreases when the sample size is increased. The MSE of the ML estimators of GEV's parameters is less than the MSE of the ML estimators of TL-GEV's parameters.

3.5 Extreme value analysis and return level estimation of PM2.5 in Chiang Mai, Thailand

In this study, we analyze a Particulate Matter 2.5 (PM2.5) data set as an application of the TL-GEV distribution in extreme value analysis, and compare it to the GEV distribution. The two example data sets are on PM2.5 (ug/m^3) from air quality monitoring station at Mueang Chiang Mai, Chiang Mai Province in Thailand. Observations of daily PM2.5 are the 24-hour averages of PM2.5 (see http://air4thai.pcd.go.th/webV2/ history/#). In this study, the Kolmogorov-Smirnov (KS) test, Akaike Information Criterion (AIC), and Bayes Information Criterion (BIC) are used as criteria for the goodness of fit, where the model that gives the smaller values of AIC, BIC, and KS is the better fit to the data.

Data set I: We consider the PM2.5 data in 2012-2021 (the past 10 years) at Tambon Sri Phum. Let $X_{i(n)}$ is the maximum PM2.5 in each month (120 months), $X_{i(n)}=\max\{X_{i1}, X_{i2},...X_{i(ni)}\}$ where n_i is the number of days in the month i*th* as i=1, 2, . . ., 120. The data set is shown in Figure 4 (a) – (b). The parameter estimates and the goodness of fit test for these data are summarized in Table 3. The TL-Fr distribution gives lower AIC, BIC, and KS values than other distributions, namely Fr, Gum, and TL-Gum. We conclude that the TL-Fr distribution is appropriate to fit these data (KS=0.0588, p-value = 0.8008), Figure 4. The ML estimates of the TL-Fr distribution are $\hat{\mu}=18.250$, $\hat{\sigma}=14.171$, $\hat{\xi}=1.2737$, and $\hat{\lambda}=4.3529$. The expression of return levels of the PM2.5 value at Tambon Sri Phum is

$$x_p = 18.250+\frac{14.171}{1.2737}\left\{\left[-\log\left(\sqrt{1-p^{1/4.3529}}\right)\right]^{-1.2737}-1\right\},$$

where p=1/m. The return levels of the PM2.5 in Chiang Mai are presented in Table 4, for m=2, 3, 4, . . ., 13.

Data set II: We consider the PM2.5 data in 2017-2021 (the data are collected in the past 5 years) at Tambon Chang Phueak. Let $X_{i(n)}$ be the maximum PM2.5 in each month (60 months), $X_{i(n)}=\max\{X_{i1}, X_{i2},...X_{i(ni)}\}$ where in the month i*th* as i=1, 2, . . ., 60. The data are presented in Figure 4 (c) – (d). The parameter estimates and the goodness of fit test for these data are summarized in Table 3. The TL-Fr distribution gives lower AIC, BIC, and KS than the other distributions, namely Fr, Gum, and TL-Gum. We conclude that the TL-Fr distribution is appropriate to fit these data (KS=0.0924, p-value = 0.6854), Figure 4. The ML estimates of the TL-Fr distribution are $\hat{\mu}=25.189$, $\hat{\sigma}=16.538$, $\hat{\xi}=2.5449$, and $\hat{\lambda}=1.5546$. The expression of return levels of PM2.5 at Tambon Chang Phueak is

$$x_p = 25.189+\frac{16.538}{2.5449}\left\{\left[-\log\left(\sqrt{1-p^{1/1.5546}}\right)\right]^{-2.5449}-1\right\},$$

where p=1/m. The return levels of the PM2.5 in Chiang Mai are presented in Table 4, for m=2, 3, 4, . . ., 13.

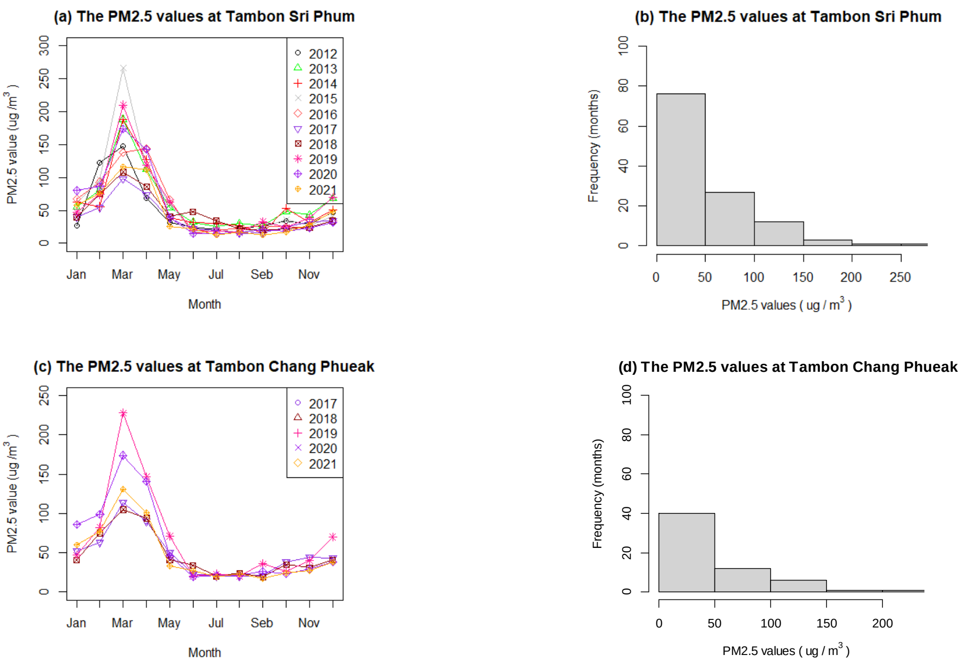

Figure 4. The maximum PM2.5 values in Chiang Mai, Thailand for each month

Table 3. Results of an extreme value analysis of the maximum PM2.5 values in each month in Chiang Mai, Thailand

Dist.	ML estimates (Standard error)				-log L	AIC	BIC	KS test (p-value)
	$\hat{\xi}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$				
Data I: The PM2.5 at Tambon Sri Phum								
Gum	-	35.920	26.553	-	595.83	1195.66	1201.24	0.1670
		(0.1702)	(0.2382)					(0.0025)
Fr	0.6732	26.531	16.452	-	589.89	1185.78	1194.14	0.0755
	(112.55)	(1.3995)	(1.1126)					(0.5014)
TL-Gum	-	2.1561	47.870	6.2724	601.24	1208.48	1216.84	0.1773
		(0.1420)	(0.1198)	(3.0495)				(0.0011)
TL-Fr	1.2737	18.250	14.171	4.3529	569.41	1146.82	1157.97	0.0588
	(42.981)	(1.5576)	(0.4338)	(6.3319)				(0.8008)
Data II: The PM2.5 at Tambon Chang Phueak								
Gum	-	36.900	25.219	-	294.66	593.32	597.51	0.1559
		(0.3072)	(0.3631)					(0.1082)
Fr	0.8892	28.360	13.503	-	279.98	565.96	572.24	0.0958
	(7.4387)	(1.3016)	(1.0245)					(0.6400)
TL-Gum	-	2.3799	45.333	6.8761	297.21	600.42	606.70	0.1701
		(0.2833)	(0.2696)	(1.1264)				(0.0622)
TL-Fr	1.5546	25.189	16.538	2.5449	277.74	563.48	571.86	0.0924
	(4.1912)	(1.3854)	(0.6631)	(3.0434)				(0.6854)

Table 4. The return levels of the maximum PM2.5 levels (ug/m^3) in Chiang Mai

m-month	2	3	4	5	6	7	8	9	10	11	12	13
Data I	18.9	23.2	26.4	29.0	31.3	33.3	35.2	36.8	38.3	39.8	41.1	42.4
Data II	32.4	43.6	53.5	62.6	71.2	79.2	86.9	94.2	101.3	108.1	114.7	121.2

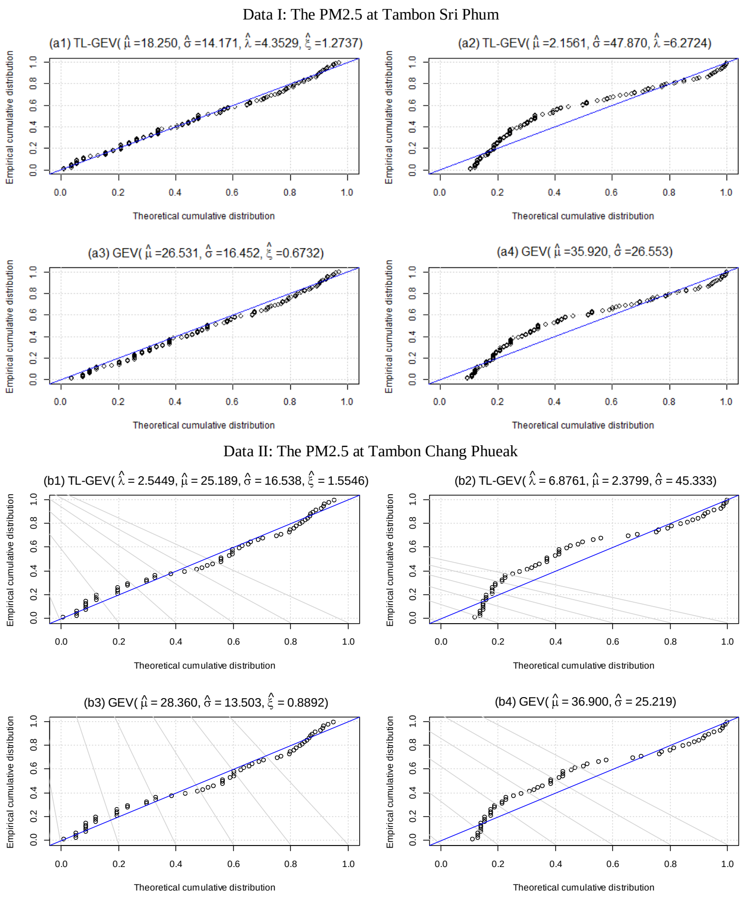

Figure 5. Probability plots of the fits of distributions to two real data sets

4. Conclusions

In this study, we introduced an extension of the generalized extreme value (GEV) distribution, generated based on the Topp Leone-generated (TL-G) family of distributions, which was proposed by Sangsanit and Bodhisuwan (2016). The new extension is called the Topp Leone-generalized extreme value (TL-GEV) distribution. The TL-GEV distribution has the four parameters (λ, μ, σ, ξ), and it has the three named sub-models TL-Gumbel (for ξ=0), TL-Fr**é**chet (for ξ>0), and TL-reversed Weibull (for ξ<0). Some properties of the proposed distribution, including quantile function, moments, moment generating function, and order statistics, were provided. The model parameters were estimated based on the maximum likelihood (ML) method. The simulation results show that each parameter's mean of ML estimates has a value close to the true parameter when the sample size is increased. The MSE of ML estimators decreases when the sample size is increased. Two real data sets of PM2.5 in Chiang Mai, Thailand, were used to demonstrate the efficiency of the proposed distribution. Based on the results, the TL-Fr**é**chet distribution gives the best fit to the studied data, on PM2.5 in Chaing Mai, among the considered distributions (Fr**é**chet, Gumbel, and TL-Gumbel distributions). The usefulness of the TL-GEV of heavy-tailed distributions has been demonstrated with two data sets, on PM2.5 data for Chiang Mai Province from the stations at Tambon Sri Phum and Tambon Chang Pheak, and the model performed reasonably well relative to the well-known competing heavy-tailed distributions, namely Gumbel, Fr**é**chet and revised Weibull. The developed distribution is promising for modeling data distributions, and may be helpful for researchers who deal with such datasets. The additional parameter λ provides flexibility to the distribution, affecting its skewness and kurtosis. Thus, the new model can provide good competition as an alternative to prior existing models. Future work includes (i) bivariate extension of actuarial measures and a Monte Carlo simulation study of these measures, (ii) modeling heavy-tailed data with bivariate extension, (iii) regression problems with covariates, and (iv) parameter reduction.

References

Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron, 71*(1), 63-79.

Bali, T. G. (2003). The generalized extreme value distribution. *Economics Letters, 79*(3), 423-427.

Castillo E., Hadi A. S., Balakrishnan N., Sarabia J. M. (2004). *Extreme value and related models with applications in engineering and science*. New York, NY: John Wiley.

Coles S. (2001). *An introduction to statistical modelling of extreme values*. London, England: Springer-Verlag.

Eugene, N., Lee, C., & Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods*, *31*(4), 497-512.

Greenwood, J. A., Landwehr, J., Matalas, N., & Wallis, J. (1979). Probability weighted moments definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research, 15*(5), 1049-1054.

Guloksuz, C. T., & Celik, N. (2020). An extension of generalized extreme value distribution: Uniform-GEV distribution and its application to earthquake data. *Thailand Statistician*, *18*(4), 491-506.

Hamed, D., & Alzaghal, A. (2021). New class of Lindley distributions: Properties and applications. *Journal of Statistical Distributions and Applications*, *8*(1), 1-22.

Kotx, S., & Nadaraja, S. (2000). *Extreme value distributions: Theory and applications*. Singapore: Imperial College Press.

Lee, C., Famoye, F., & Alzaatreh, A. Y. (2013). Methods for generating families of univariate continuous distributions in the recent decades. *Wiley Interdisciplinary Reviews: Computational Statistics, 5*(3), 219-238.

Nadarajah, S., & Pogány, T. K. (2013). On the characteristic functions for extreme value distributions. *Extremes*, *16*(1), 27-38.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing (2021). Retrieved from https:// www.R-project.org/

Sangsanit, Y., & Bodhisuwan, W. (2016). The Topp-Leone generator of distributions: properties and inferences. *Songklanakarin Journal of Science and Technology, 38*(5), 537-548.

Shukla, R. K., Trivedi, M., & Kumar, M. (2012). On the proficient use of GEV distribution: a case study of subtropical monsoon region in India. *arXiv preprint arXiv:1203.0642*.