



Thailand Statistician
April 2022; 20(2): 293-307
<http://statassoc.or.th>
Contributed paper

The Length-Biased Gamma-Rayleigh Distribution with Applications

Tanachot Chaito [a][b], Pimwarat Nanthaprut [b], Nawapon Nakharutai [b] and Manad Khamkong*[b]

[a] Ph.D. Program in Applied Statistics, Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand.

[b] Data Science Research Centre, Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand.

*Corresponding author; e-mail: manad.k@cmu.ac.th

Received: 8 July 2020

Revised: 30 September 2020

Accepted: 2 February 2021

Abstract

In this study, we introduce a new class of length-biased distribution called the length-biased Gamma-Rayleigh (LBGR) distribution. Various mathematical properties, including its limit behavior, the r^{th} moment, the moment generating function, the survival function, and the hazard ratio function, are derived. Maximum-likelihood estimation is used to estimate the parameters of the LBGR distribution. Three environmental datasets were used in a comparison of distributions to determine which one could best fit the data. The results indicate that the LBGR distribution provided a better fit than the Rayleigh, lognormal, gamma, and gamma-Rayleigh distributions.

Keywords: Average wind speed, fit distribution, maximum likelihood estimation, model selection criteria, positively skewness.

1. Introduction

The study of distribution functions and properties is important for the modeling of random phenomena (Akarawak et al. 2017). Moreover, knowing the appropriate distribution of a particular phenomenon will increase the power and efficiency of any statistical tests involving it (Nasiru and Luguterah 2015). When analyzing environmental events, various distributions or models can be applied to environmental data, including the Weibull, gamma, lognormal, Pearson type III, Fréchet, Gumbel, Rayleigh and generalized extreme value distributions (Bidaoui et al. 2019, Blain 2011, Chang 2011, Khamkong and Bookkamana 2015, Chaito and Khamkong 2018, Chaito et al. 2019, and Cordeiro et al. 2019)

The gamma distribution is another distribution that is frequently used because it is positively skewed, which is appropriate for environmental data such as runoff, rainfall, runoff hydrograph, and wind speed modeling (Bhunya et al. 2003, Husak et al. 2007, Sharma and Singh 2010, Voytenko 2011, Abdulkarim et al. 2017, Akyuz and Gamgam 2017, Ghorbani et al. 2017, and Chaito et al. 2019). Thus, many researchers have developed new distributions by mixed the gamma distribution

with other distributions to provide better fitting of the data. Alzaatreh et al. (2012) introduced the gamma-Pareto distribution applied to annual flood rates data for the Floyd River in James, Iowa, USA. Later, Alzaatreh et al. (2016) presented a gamma half-Cauchy distribution for application to hydrology data. However, mixed gamma distributions have been applied to not only environmental and hydrology data but also data from other subject areas, such as survival, reliability, and lifetime data with the gamma-Weibull distribution (Provost et al. 2011), the gamma-exponentiated exponential distribution (Ristić and Balakrishnan 2012), the gamma-normal distribution (Alzaatreh et al. 2014), the Gamma-Pareto (IV) distribution (Alzaatreh and Ghosh 2016), and the gamma generalized pareto distribution (De Andrade et al. 2017)

Another interesting distribution that involves mixing with the gamma distribution is the gamma-Rayleigh distribution developed by Akarawak et al. (2017). An interesting phenomenon about this distribution is that it is right-skewed, which makes it more suitable for application to environmental data. Another point of interest is that it is more effective and fits data better than the Weibull, Rayleigh, gamma, and lognormal distributions, all of which are frequently applied to environmental data, especially hydrological data such as rainfall and runoff. Let X is the gamma-Rayleigh random variable, the probability density function (pdf) is given by

$$f(x) = \frac{2\theta^\alpha}{\Gamma(\alpha)} x^{2\alpha-1} e^{-\theta x^2}; \quad x > 0, \alpha > 0, \theta > 0, \quad (1)$$

where α is a shape parameter, θ is a scale parameter and $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$ is a gamma function.

The cumulative distribution function (cdf) can be written as

$$F(x) = \frac{\gamma(\alpha, \theta x^2)}{\Gamma(\alpha)}, \quad (2)$$

where $\gamma(\alpha, x) = \int_0^x u^{\alpha-1} e^{-u} du$ is the lower incomplete gamma function. The mean and variance of the gamma-Rayleigh distribution are, respectively,

$$E(X) = \frac{\Gamma\left(\frac{1}{2} + \alpha\right)}{\theta^{\frac{1}{2}} \Gamma(\alpha)}, \quad (3)$$

and

$$Var(X) = \frac{\alpha(\Gamma(\alpha))^2 - \left(\Gamma\left(\frac{1}{2} + \alpha\right)\right)^2}{\theta(\Gamma(\alpha))^2}. \quad (4)$$

The idea of applying weights to a distribution, which was introduced by Fisher (1934), is another popular method of improving the fit. Weighted distribution is used when the probabilities of observations recorded from a random process are not equal. Although it is incorrect to apply the original distribution to explain these observations, a weighted distribution can be used instead (Atikankul et al. 2020). Moreover, the weighted distribution can be applied to environment because, in traditional environmental statistical methods, the obtained observations or samples are often non-random, non-replicated, and/or non-experimental, which means that the probabilities of the observations recorded from a random process are not equal. Thus, the weighted distribution provides a unifying approach to this problem (Nanuwong 2015, and Ahmad et al. 2016). For the weighted

distribution, it is assumed that X is a random variable with pdf $f(x)$ then the weighted distribution can be written as

$$f(x) = \frac{w(x)f(x)}{E[w(x)]}, \quad (5)$$

where $w(x)$ is a non-negative weight function and $E[w(x)]$ exists. Afterwards, Rao (1965) adjusted and offered general terms for weighted distributions where the weighted function is the form $w(x) = x^c$ which is called a size-biased distributions of order c . In this case, the pdf of X becomes

$$f_c(x) = \frac{x^c f(x)}{E[X^c]}. \quad (6)$$

When $c = 1$, we obtain a length-biased distribution, and so its pdf can be defined as

$$f_L(x) = \frac{xf(x)}{E(X)}, \quad (7)$$

where $E(X) = \int_{-\infty}^{\infty} xf(x)dx$. Adding length-biased distribution has been used to develop distributions such as the length-biased weighted Weibull distribution (Das and Roy 2011a), the length-biased weighted generalized Rayleigh distribution (Das and Roy 2011b), the length-biased exponentiated inverted Weibull distribution (Seenoi et al. 2014), the length-biased weighted Lomax distribution (Ahmad et al. 2016), the length-biased Sushila distribution (Rather and Subramanian 2018) and the length-biased weighted quasi gamma distribution (Ganaie and Rajagopalan 2020).

The contribution of this study is to introduce a new distribution called the length-biased gamma-Rayleigh (LBGR) distribution, which is an improvement of gamma-Rayleigh distribution, and compare its efficiency with other distributions (gamma-Rayleigh, Rayleigh, lognormal, and gamma). Moreover, the LBGR distribution is used to fit data from environmental datasets. This article is structured as follows. Section 2 provides the proposed distribution and mathematical properties of the proposed distribution, respectively. Moreover, generating random data of the LBGR distribution is presented in Section 3. Parameter estimation of the proposed distribution is illustrated in Section 4. In Section 5, the outcomes of applying the proposed distribution to three environmental datasets are provided. Finally, concluding remarks on the study are presented in Section 6.

2. The LBGR Distribution

In this section, we provide the pdf, the cdf, and the mathematical properties of the LBGR distribution.

2.1. The pdf and the cdf of the LBGR Distribution

Definition 1. If X is a non-negative random variable with pdf $f(x)$ and the expected value or mean $E(X) < \infty$, then the length-biased distribution of X is given by

$$g(x) = \frac{xf(x)}{E(X)}. \quad (8)$$

Theorem 1 Let $X \sim LBGR(\alpha, \theta)$, then the pdf and the cdf of a random variable X are, respectively,

$$g(x) = \frac{2\theta^{\alpha+\frac{1}{2}}}{\Gamma\left(\frac{1}{2} + \alpha\right)} x^{2\alpha} e^{-\theta x^2}; \quad x > 0, \alpha > 0, \theta > 0, \quad (9)$$

and

$$G(x) = \frac{\gamma\left(\frac{1}{2} + \alpha, \theta x^2\right)}{\Gamma\left(\frac{1}{2} + \alpha\right)}, \quad (10)$$

where α is a shape parameter, θ is a scale parameter, $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$ is a gamma function and

$\gamma(\alpha, x) = \int_0^x u^{\alpha-1} e^{-u} du$ is the lower incomplete gamma function.

Proof: Substitute the pdf in (1) and the mean in (3) of the gamma-Rayleigh distribution into (8) in Definition 1, then the pdf of the LBGR distribution is obtained by

$$g(x) = \frac{xf(x)}{E(X)},$$

$$g(x) = \frac{x \left(\frac{2\theta^\alpha}{\Gamma(\alpha)} x^{2\alpha-1} e^{-\theta x^2} \right)}{\frac{1}{\theta^2 \Gamma(\alpha)}} = \frac{2\theta^{\alpha+\frac{1}{2}} x^{2\alpha} e^{-\theta x^2}}{\Gamma\left(\frac{1}{2} + \alpha\right)}.$$

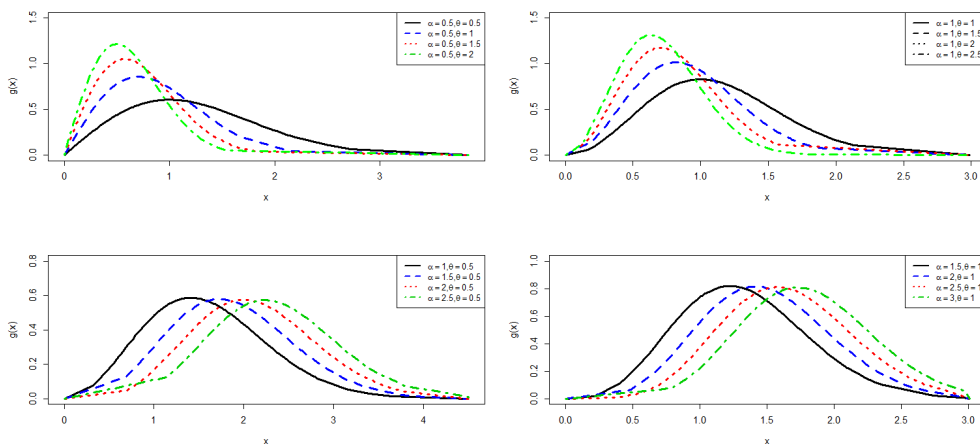


Figure 1 The pdf plots of the LBGR distribution for different parameter values

If X is a random variable with pdf $g(x)$, then its cdf can be written as $G(x) = \int_0^x g(t)dt$ and

the cdf of the LBGR distribution becomes $G(x) = \int_0^x \frac{2\theta^{\alpha+\frac{1}{2}}}{\Gamma\left(\frac{1}{2}+\alpha\right)} t^{2\alpha} e^{-\theta t^2} dt$.

Let $u = \theta t^2$ and $0 < u < \theta x^2$, then by applying an integration transform, we obtain

$$G(x) = \frac{2\theta^{\alpha+\frac{1}{2}}}{2\theta^{\frac{1}{2}}\Gamma\left(\frac{1}{2}+\alpha\right)} \int_0^{\theta x^2} \left(\frac{u}{\theta}\right)^{\alpha} u^{-\frac{1}{2}} e^{-u} du = \frac{1}{\Gamma\left(\frac{1}{2}+\alpha\right)} \int_0^{\theta x^2} u^{\alpha-\frac{1}{2}} e^{-u} du = \frac{\gamma\left(\frac{1}{2}+\alpha, \theta x^2\right)}{\Gamma\left(\frac{1}{2}+\alpha\right)}.$$

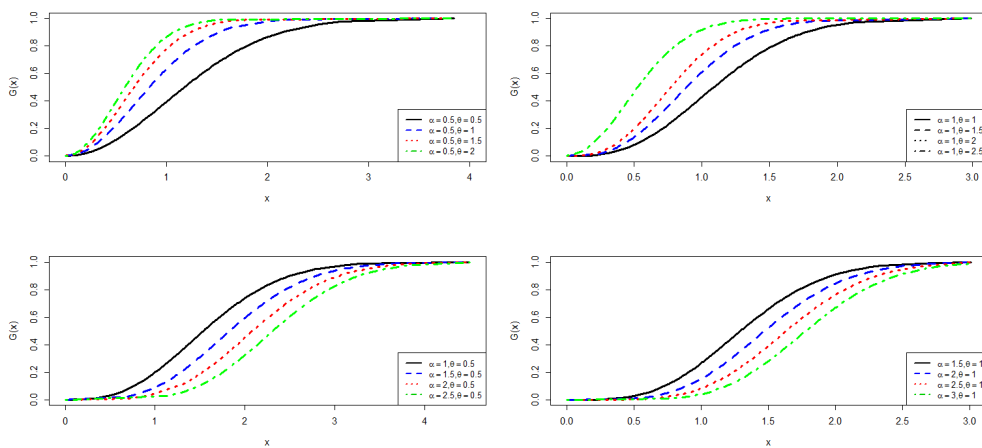


Figure 2 The cdf plots of the LBGR distribution for different parameter values

Pdf plots of the LBGR distribution for different parameter values are presented in Figure 1, showing that the distribution is right-skewed or approximately symmetric. Likewise, plots of the LBGR cdf with different parameter values are presented in Figure 2.

2.2. Mathematical properties of the LBGR distribution

In this section, we discuss the mathematical properties of the LBGR distribution, such as limit behavior, the r^{th} moment, a moment generating function, a survival function, and a hazard rate function.

2.2.1. Limit behavior

The limits of the pdf for the LBGR distribution are 0 as $x \rightarrow \infty$ and as $x \rightarrow 0$. By placing these limits into (9), the limit of the pdf for the LBGR distribution when $x \rightarrow \infty$ becomes

$$\lim_{x \rightarrow \infty} g(x) = \frac{2\theta^{\alpha+\frac{1}{2}}}{\Gamma\left(\frac{1}{2}+\alpha\right)} \lim_{x \rightarrow \infty} x^{2\alpha} e^{-\theta x^2} = \frac{2\theta^{\alpha+\frac{1}{2}}}{\Gamma\left(\frac{1}{2}+\alpha\right)} \times 0 = 0,$$

and when $x \rightarrow 0$, it becomes

$$\lim_{x \rightarrow 0} g(x) = \frac{2\theta^{\alpha+\frac{1}{2}}}{\Gamma\left(\frac{1}{2}+\alpha\right)} \lim_{x \rightarrow 0} x^{2\alpha} e^{-\theta x^2} = \frac{2\theta^{\alpha+\frac{1}{2}}}{\Gamma\left(\frac{1}{2}+\alpha\right)} \times 0 = 0.$$

2.2.2. The r^{th} moment

Theorem 2 Let $X \sim \text{LBGR}(\alpha, \theta)$, then the r^{th} moment of X is given by

$$E(X^r) = \frac{\Gamma\left(1+\alpha+\frac{r-1}{2}\right)}{\theta^{\frac{r}{2}}\Gamma\left(\frac{1}{2}+\alpha\right)}, \quad r=1,2,3,\dots \quad (11)$$

Proof: If X is a random variable with the pdf $g(x)$, then the r^{th} moment can be derived as

$$E(X^r) = \int_{-\infty}^{\infty} x^r g(x) dx,$$

and if $X \sim \text{LBGR}(\alpha, \theta)$, then the r^{th} moment of X becomes

$$E(X^r) = \int_0^{\infty} x^r \frac{2\theta^{\alpha+\frac{1}{2}}}{\Gamma\left(\frac{1}{2}+\alpha\right)} x^{2\alpha} e^{-\theta x^2} dx.$$

Let $u = \theta x^2$, then by applying an integration transform, we obtain

$$E(X^r) = \frac{\theta^{\alpha+\frac{1}{2}} 2^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}+\alpha\right)} \int_0^{\infty} \left(\frac{u}{\theta}\right)^{\frac{2\alpha+r}{2}} u^{-\frac{1}{2}} e^{-u} du = \frac{\theta^{\alpha}}{\theta^{\frac{2\alpha+r}{2}} \Gamma\left(\frac{1}{2}+\alpha\right)} \int_0^{\infty} u^{\frac{2\alpha+r-1}{2}} e^{-u} du = \frac{\Gamma\left(1+\alpha+\frac{r-1}{2}\right)}{\theta^{\frac{r}{2}} \Gamma\left(\frac{1}{2}+\alpha\right)}.$$

From (11), we can derive the mean and variance of the LBGR distribution, respectively, as

$$E(X) = \frac{\Gamma(1+\alpha)}{\sqrt{\theta}\Gamma\left(\frac{1}{2}+\alpha\right)}, \quad (12)$$

and

$$\text{Var}(X) = \frac{\Gamma\left(\frac{3}{2}+\alpha\right)}{\theta\Gamma\left(\frac{1}{2}+\alpha\right)} - \left[\frac{\Gamma(1+\alpha)}{\sqrt{\theta}\Gamma\left(\frac{1}{2}+\alpha\right)} \right]^2 = \frac{1}{\theta} \left[\frac{\Gamma\left(\frac{3}{2}+\alpha\right)}{\Gamma\left(\frac{1}{2}+\alpha\right)} - \left[\frac{\Gamma(1+\alpha)}{\Gamma\left(\frac{1}{2}+\alpha\right)} \right]^2 \right]. \quad (13)$$

Furthermore, we can find the skewness and kurtosis of the LBGR distribution, respectively, as follows:

$$\text{Skewness}(X) = \varphi^{-3} \left\{ \frac{\Gamma(2+\alpha)}{\theta^{\frac{3}{2}}\Gamma\left(\frac{1}{2}+\alpha\right)} - 3 \left[\frac{\Gamma(1+\alpha)}{\theta^{\frac{1}{2}}\Gamma\left(\frac{1}{2}+\alpha\right)} \right] \left[\frac{1}{\theta} \left[\frac{\Gamma\left(\frac{3}{2}+\alpha\right)}{\Gamma\left(\frac{1}{2}+\alpha\right)} - \left[\frac{\Gamma(1+\alpha)}{\Gamma\left(\frac{1}{2}+\alpha\right)} \right]^2 \right] \right] \right\} - \left[\frac{\Gamma(1+\alpha)}{\theta^{\frac{1}{2}}\Gamma\left(\frac{1}{2}+\alpha\right)} \right]^3,$$

(14)

and

$$\begin{aligned}
 \text{Kurtosis}(X) = \varphi^{-4} & \left\{ \frac{\Gamma\left(\frac{5}{2} + \alpha\right)}{\theta^2 \Gamma\left(\frac{1}{2} + \alpha\right)} - 4 \left[\frac{\Gamma(2 + \alpha)}{\theta^{\frac{3}{2}} \Gamma\left(\frac{1}{2} + \alpha\right)} \right] \left[\frac{\Gamma(1 + \alpha)}{\theta^{\frac{1}{2}} \Gamma\left(\frac{1}{2} + \alpha\right)} \right] \right\} + 6 \left[\frac{\Gamma\left(\frac{3}{2} + \alpha\right)}{\theta \Gamma\left(\frac{1}{2} + \alpha\right)} \right] \left[\frac{\Gamma(1 + \alpha)}{\theta^{\frac{1}{2}} \Gamma\left(\frac{1}{2} + \alpha\right)} \right]^2 \\
 & - 3 \left[\frac{\Gamma(1 + \alpha)}{\theta^{\frac{1}{2}} \Gamma\left(\frac{1}{2} + \alpha\right)} \right]^4,
 \end{aligned} \tag{15}$$

$$\text{where } \varphi = \sqrt{\frac{1}{\theta} \left[\frac{\Gamma\left(\frac{3}{2} + \alpha\right)}{\Gamma\left(\frac{1}{2} + \alpha\right)} - \left[\frac{\Gamma(1 + \alpha)}{\Gamma\left(\frac{1}{2} + \alpha\right)} \right]^2 \right]}.$$

2.2.3. Moment generating function

Theorem 3 Let $X \sim \text{LBGR}(\alpha, \theta)$, then the moment generating function of X is given as

$$M_x(t) = \sum_{r=0}^{\infty} \frac{t^r \Gamma\left(1 + \alpha + \frac{r-1}{2}\right)}{r! \theta^{\frac{r}{2}} \Gamma\left(\frac{1}{2} + \alpha\right)}. \tag{16}$$

Proof: If X is a random variable with the pdf $g(x)$, then by definition, the moment generating function is

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} g(x) dx,$$

and so, the moment generating function of X becomes

$$M_x(t) = \int_0^{\infty} e^{tx} \frac{2\theta^{\alpha + \frac{1}{2}}}{\Gamma\left(\frac{1}{2} + \alpha\right)} x^{2\alpha} e^{-\theta x^2} dx.$$

Using Taylor series, we can obtain

$$M_x(t) = \int_{-\infty}^{\infty} \left[1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \dots + \frac{(tx)^r}{r!} \right] g(x) dx = \sum_{r=0}^{\infty} \frac{t^r E(X^r)}{r!} = \sum_{r=0}^{\infty} \frac{t^r \Gamma\left(1 + \alpha + \frac{r-1}{2}\right)}{r! \theta^{\frac{r}{2}} \Gamma\left(\frac{1}{2} + \alpha\right)}.$$

2.2.4. The survival and hazard rate functions

We show the derivations of the survival and hazard rate functions of the LBGR distribution in this section

1) Survival function

By the definition, the survival function of a random variable X with the cdf $G(x)$ is derived as

$$S(x) = 1 - G(x). \quad (17)$$

By substituting the cdf in (10) into (17), we can obtain the survival function of the LBGR distribution as

$$S(x) = 1 - \frac{\gamma\left(\frac{1}{2} + \alpha, \theta x^2\right)}{\Gamma\left(\frac{1}{2} + \alpha\right)} = \frac{\Gamma\left(\frac{1}{2} + \alpha\right) - \gamma\left(\frac{1}{2} + \alpha, \theta x^2\right)}{\Gamma\left(\frac{1}{2} + \alpha\right)}. \quad (18)$$

2) Hazard rate function

By the definition, the hazard rate function of a random variable X with the pdf $g(x)$ and cdf $G(x)$ is defined by

$$h(x) = \frac{g(x)}{1 - G(x)}. \quad (19)$$

By substituting the pdf in (9) and the cdf in (10) into (19), the hazard rate function of the LBGR distribution can be rewritten as

$$h(x) = \frac{\frac{2\theta^{\alpha+\frac{1}{2}} x^{2\alpha} e^{-\theta x^2}}{\Gamma\left(\frac{1}{2} + \alpha\right)}}{1 - \frac{\gamma\left(\frac{1}{2} + \alpha, \theta x^2\right)}{\Gamma\left(\frac{1}{2} + \alpha\right)}} = \frac{2\theta^{\alpha+\frac{1}{2}} x^{2\alpha} e^{-\theta x^2}}{\Gamma\left(\frac{1}{2} + \alpha\right) - \gamma\left(\frac{1}{2} + \alpha, \theta x^2\right)}. \quad (20)$$

The survival and hazard rate functions are displayed as plots of the LBGR distribution for different parameter values in Figures 3 and 4, respectively.

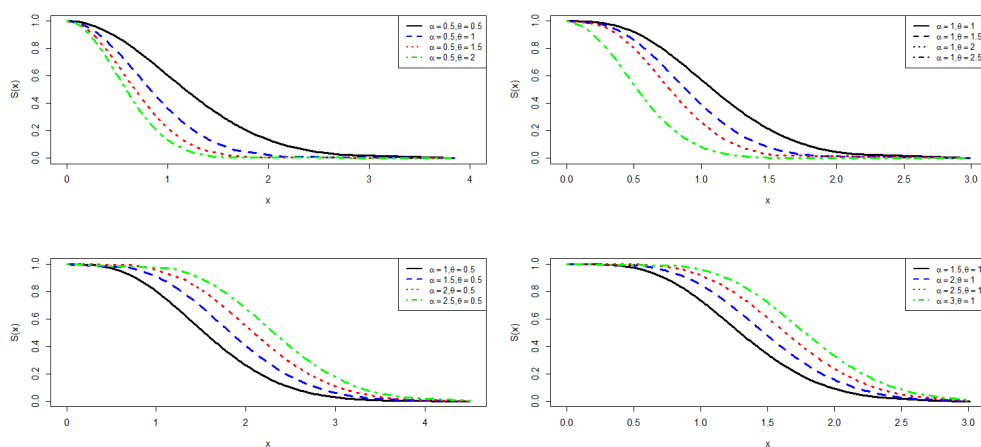


Figure 3 The survival plots of the LBGR distribution for different parameter values

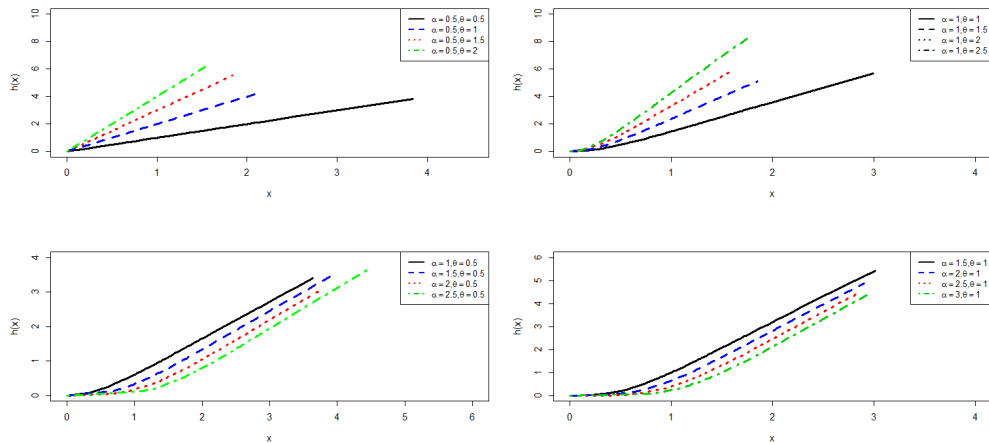


Figure 4 The hazard rate plots of the LBGR distribution for different parameter values

3. Random Variate Generation

The inverse transformation technique is used to generate random data for the LBGR distribution by setting $U = G^{-1}(x)$, where U is uniform distribution on $(0,1)$, and $G^{-1}(x)$ is inverse of the cdf in (10).

For generating random data $X_i, i = 1, 2, \dots, n$, from $LBGR(\alpha, \theta)$, one can use the following algorithm:

1) Generate $U_i, i = 1, 2, \dots, n$, from $U(0,1)$.

2) Set $U_i = G(x_i)$, then $U_i = \frac{\gamma\left(\frac{1}{2} + \alpha, \theta x^2\right)}{\Gamma\left(\frac{1}{2} + \alpha\right)}$ and $U_i \Gamma\left(\frac{1}{2} + \alpha\right) = \gamma\left(\frac{1}{2} + \alpha, \theta x^2\right)$.

3) Set $W_i = U_i \Gamma\left(\frac{1}{2} + \alpha\right)$ and $t_i = \gamma^{-1}\left(\frac{1}{2} + \alpha, W_i\right)$, where γ^{-1} is the inversion of the lower incomplete gamma function.

4) Set $t_i = \theta x^2$.

5) Subsequently, $x_i = \left(\frac{t_i}{\theta}\right)^{\frac{1}{2}}$.

4. Maximum Likelihood Estimation

In this section, we present the maximum likelihood estimation of the LBGR distribution. Let $X_1, X_2, \dots, X_n$ be distributed random variables from the LBGR and let $\Theta = (\alpha, \theta)^T$ be a vector of the parameters, then the likelihood function is given by

$$L(\Theta) = \prod_{i=1}^n \left\{ \frac{2\theta^{\alpha+\frac{1}{2}}}{\Gamma\left(\frac{1}{2} + \alpha\right)} x^{2\alpha} e^{-\theta x^2} \right\}. \quad (21)$$

Thus, the log-likelihood function can be written as

$$\log L(\Theta) = n \log 2 + n \left(\frac{1}{2} + \alpha \right) \log \theta - n \log \Gamma \left(\frac{1}{2} + \alpha \right) + \sum_{i=1}^n \log x_i^{2\alpha} - \sum_{i=1}^n \theta x_i^2. \tag{22}$$

The derivatives of (22) with respect to α and θ are, respectively, given by

$$\frac{\partial \log L(\Theta)}{\partial \alpha} = n \log \theta - n \psi \Gamma \left(\frac{1}{2} + \alpha \right) + 2 \sum_{i=1}^n \log x_i, \tag{23}$$

$$\frac{\partial \log L(\Theta)}{\partial \theta} = \frac{n \left(\alpha + \frac{1}{2} \right)}{\theta} - \sum_{i=1}^n x_i^2, \tag{24}$$

where $\psi(y) = \frac{d}{dy} \log \Gamma(y) = \frac{\Gamma'(y)}{\Gamma(y)}$ is a digamma function. The maximum likelihood estimates of parameters α and θ ($\hat{\alpha}$ and $\hat{\theta}$, respectively) can be estimated by setting (23) and (24) equal to zero and solving them simultaneously.

5. Application of the LBGR Distribution

We applied the LBGR distribution to fit the data from three environmental datasets comprising data on average annual runoff, average monthly wind speed, and annual runoff. We compare the LBGR distribution with the Rayleigh, lognormal, gamma, and gamma-Rayleigh distributions using the Kolmogorov-Smirnov (K-S) statistic, the Akaike information criterion (AIC) (Akaike 1973), and the Bayesian information criterion (BIC) (Schwarz 1978) to determine the best-fitting distribution.

We used averaged data in datasets I and II because averaging is commonly conducted on environmental data that are available online and this is how they have been presented. However, the annual runoff data in dataset III was as-recorded.

Dataset I: This comprises average annual runoff data recorded at the P.24A station located in Chom Thong, Chiang Mai, Thailand, collected by the Upper Northern Region Irrigation Hydrology Center, Thailand (Upper Northern Region Irrigation Hydrology Center 2020). The data from 1974 to 2019 were used for selecting the best-fitting distribution.

Table 1 Descriptive statistics for the average annual runoff data (million cubic meter) recorded at the P.24A station, Chom Thong, Chiang Mai, Thailand.

Sample Size	Q ₁	Q ₂	Q ₃	Mean	SD	Skewness	Kurtosis
46	9.43	12.70	15.88	12.5700	4.6840	0.2702	2.9876

Note: Q_i is the i^{th} quartile of data; SD is the standard deviation.

Table 2 Maximum-likelihood estimates and distribution-fitting statistics for the average annual runoff (million cubic meters) of P.24A station, Chom Thong, Chiang Mai, Thailand.

	Distribution				
	Rayleigh	Lognormal	Gamma	Gamma-Rayleigh	LBGR
MLE	$\hat{\delta} = 9.4730$	$\hat{\mu} = 2.4530$ $\hat{\sigma} = 0.4171$	$\hat{\alpha} = 6.5520$ $\hat{\beta} = 0.5213$	$\hat{\alpha} = 1.0285$ $\hat{\theta} = 0.0057$	$\hat{\alpha} = 1.4105$ $\hat{\theta} = 0.0106$
K-S test	0.1612	0.1070	0.0941	0.1564	0.0834
p-value	0.1831	0.6680	0.8099	0.2103	0.9064
AIC	282.0239	279.7928	276.0778	283.2793	275.8255
BIC	283.8526	283.4501	279.7350	286.9366	279.4828

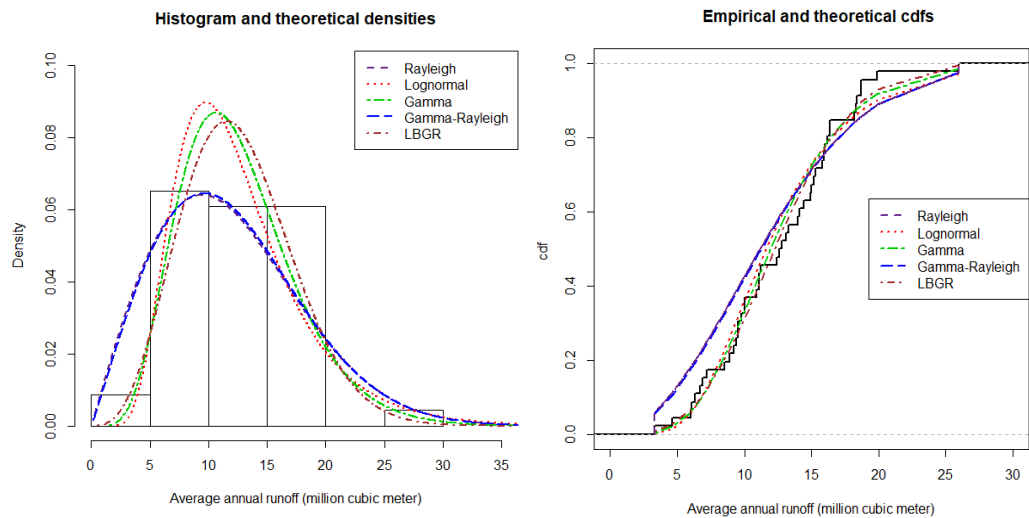


Figure 5 Histogram and theoretical densities and empirical and theoretical CDFs for the average annual runoff of P.24A station, Chom Thong, Chiang Mai, Thailand

The results in Table 1 show that the average annual runoff data consisting of 46 observations were positively skewness. Table 2 reports the parameter estimates of the average annual runoff data for each distribution, while Figure 5 illustrate the comparison of fitting for histogram theoretical densities and empirical and theoretical CDFs of the data. From Table 2, it can be seen that the K-S test, AIC, and BIC values of the LBGR distribution were lower than the others.

Dataset II: This comprises the average monthly wind speed data reported by the Chiang Rai airport weather observation station, Chiang Rai, Thailand, collected by the Thai Meteorological Department, Thailand (Thai Meteorological Department Automatic Weather System 2008). The data from January 2017 to June 2020 (missing October 2019) were used in this study.

Table 3 Descriptive statistics for the average monthly wind speed data (km/h) from the Chiang Rai airport weather observation station, Chiang Rai, Thailand.

Sample Size	Q_1	Q_2	Q_3	Mean	SD	Skewness	Kurtosis
41	1.60	2.10	2.60	2.1460	0.7291	0.2480	2.1544

Note: km/h is Kilometer per hour; Q_i is the i^{th} quartile of data; SD is standard deviation.

Table 4 Maximum-likelihood estimates and distribution-fitting statistics for the average monthly wind speed (km/h) from the Chiang Rai airport weather observation station, Chiang Rai, Thailand.

	Distribution				
	Rayleigh	Lognormal	Gamma	Gamma-Rayleigh	LBGR
MLE	$\hat{\delta} = 1.6008$	$\hat{\mu} = 0.7036$ $\hat{\sigma} = 0.3548$	$\hat{\alpha} = 8.4773$ $\hat{\beta} = 3.9496$	$\hat{\alpha} = 2.2561$ $\hat{\theta} = 0.4696$	$\hat{\alpha} = 1.8557$ $\hat{\theta} = 0.4596$
K-S test	0.2003	0.1037	0.0855	0.1043	0.0813
p-value	0.0746	0.7696	0.9255	0.7638	0.9492
AIC	103.4702	93.0770	92.0356	92.2506	91.8093
BIC	105.1838	96.5041	95.4627	95.6778	95.2364

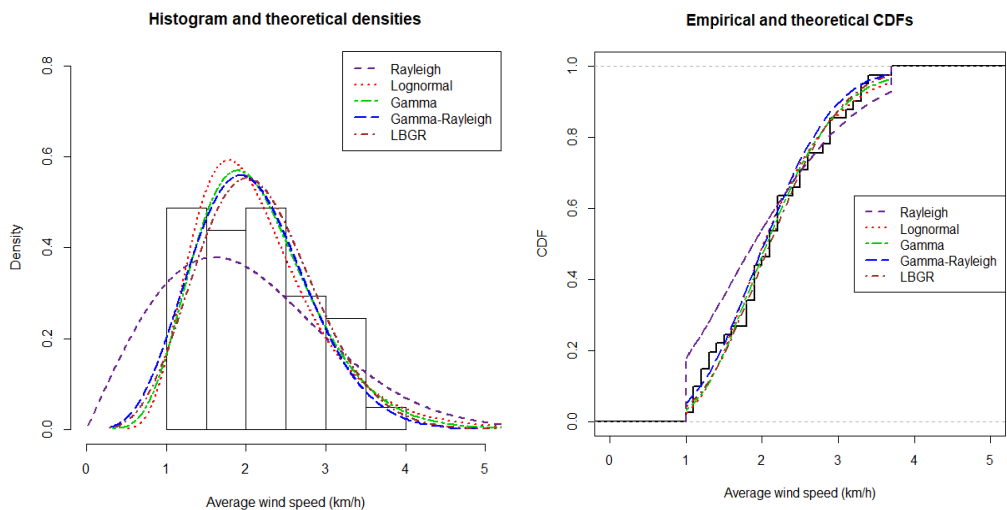


Figure 6 Histogram and theoretical densities and empirical and theoretical CDFs for the average monthly wind speed of Chiang Rai airport weather observing station, Chiang Rai, Thailand

Table 3 reports the average monthly wind speed at Chiang Rai airport weather observing station, Chiang Rai, Thailand, consisting of 41 observations. The results show that the data are positive skewed. A summary the parameter estimation for average monthly wind speed data is presented in Table 4, while Figure 6 display a comparison of the fitting histogram theoretical densities and empirical and theoretical CDFs of the data. The results in Table 4 show that the K-S test, AIC, and BIC values of the LBGR distribution were the lowest.

Dataset III: This comprises annual runoff data from G.8 station located in Mae Lao, Chiang Rai, Thailand, collected by the Upper Northern Region Irrigation Hydrology Center, Thailand (Upper Northern Region Irrigation Hydrology Center 2020). The data from 1995 to 2019 were used for selecting the best-fitting distribution.

Table 5 Descriptive statistics for the annual runoff data (one hundred million cubic meter) from the G.8 station, Mae Lao, Chiang Rai, Thailand.

Sample Size	Q_1	Q_2	Q_3	Mean	SD	Skewness	Kurtosis
25	6.0700	7.1800	9.7500	7.8620	3.1401	0.3361	0.2063

Note: Q_i is the i^{th} quartile of data; SD is the standard deviation.

Table 6 Maximum-likelihood estimates and selected distribution-fitting statistics for the annual runoff (one hundred million cubic meters) from the G.8 station, Mae Lao, Chiang Rai, Thailand.

	Distribution				
	Rayleigh	Lognormal	Gamma	Gamma-Rayleigh	LBGR
MLE	$\hat{\delta} = 5.5700$	$\hat{\mu} = 1.9647$	$\hat{\alpha} = 5.2975$	$\hat{\alpha} = 1.0705$	$\hat{\alpha} = 1.0509$
		$\hat{\sigma} = 0.4906$	$\hat{\beta} = 1.4840$	$\hat{\theta} = 0.0151$	$\hat{\theta} = 0.0218$
K-S test	0.1855	0.1810	0.1341	0.1713	0.1262
p-value	0.3156	0.3016	0.7103	0.4092	0.7751
AIC	132.9387	137.5800	133.0664	133.6641	131.3505
BIC	134.1576	140.0178	135.5041	136.1018	133.7883

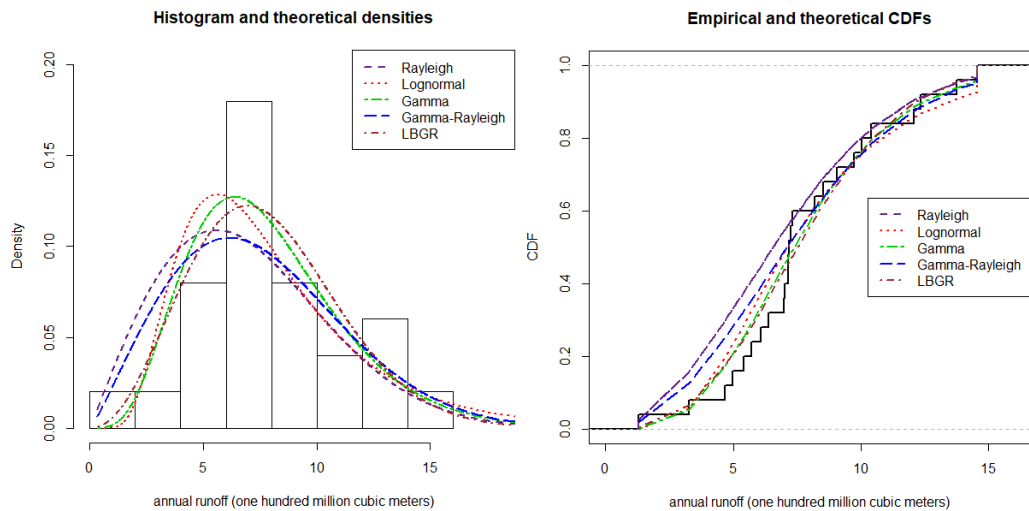


Figure 7 Histogram and theoretical densities and empirical and theoretical CDFs for the annual runoff (one hundred million cubic meter) of G.8 station, Mae Lao, Chiang Rai, Thailand

The results in Table 5 show that the annual runoff data consisting of 25 observations is right-skewed. Table 6 summarizes the parameter estimates of the annual runoff data for each distribution, while Figure 7 show the comparison of fitting for histogram theoretical densities and empirical and theoretical CDFs of the data. The results in Table 6 show that the K-S test, AIC, and BIC values for the LBGR distribution were lower than the others.

6. Conclusions

In this work, we introduced a new distribution, the LBGR distribution. We studied and investigated various mathematical properties of the LBGR distribution, such as its limit behavior, the r^{th} moment, the moment generating function, the survival function, the hazard ratio function and the maximum-likelihood estimates of parameters of the LBGR distribution. Application of the LBGR distribution was conducted on environmental datasets, including the average annual runoff, average monthly wind speed, and annual runoff data. Judging from the K-S test, AIC, and BIC values, the LBGR distribution provided the best fit for all three environmental datasets. Thus, the LBGR distribution can be applied to environmental datasets with right-skewed or approximately symmetric characteristics. As rainfall and temperature data also have these characteristics, we are certain that the LBGR distribution can be applied for the modeling and analysis of such data.

Acknowledgements

This study was supported via a teaching assistant and research assistant (TA/RA) scholarship from the Graduate School, Chiang Mai University, Chiang Mai, Thailand.

References

- Abdulkarim A, Abdelkader SM, Morrow DJ, Falade AJ, Adediran YA. Statistical analysis of wind speed for electrical power generation in some selected sites in northern Nigeria. *Niger J Technol.* 2017; 36(4): 1249-1257.
- Ahmad A, Ahmad SP, Ahmed A. Length-biased Weighted Lomax distribution: statistical properties and application. *Pak J Stat Oper Res.* 2016; 12(2): 245-255.

- Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. *Proceedings of the 2nd International Symposium on Information Theory*; 1971 Sep 2-8; Tsahkadsor, Armenia, USSR. Budapest: Akademiai Kiado; 1973. pp. 267-281.
- Akarawak EEE, Adeleke IA, Okafor RO. The Gamma-Rayleigh distribution and applications to survival data. *Nig J Basic Appl Sci*. 2017; 25(2): 130-142.
- Akyuz HE, Gamgam H. Statistical analysis of wind speed data with Weibull, lognormal and gamma Distributions. *Cumhuriyet Science J*. 2017; 38(4): 68-76.
- Alzaatreh A, Famoye F, Lee C. Gamma-Pareto distribution and its applications. *J Mod Appl Stat Meth*. 2012; 11(1): 78-94.
- Alzaatreh A, Famoye F, Lee C. The gamma-normal distribution: properties and applications. *Comput Stat Data Anal*. 2014; 69(C): 67-80.
- Alzaatreh A, Ghosh I. A study of the Gamma-Pareto (IV) distribution and its applications. *Commun Stat Theory Methods*. 2016; 45(3): 636-654.
- Alzaatreh A, Mansoor M, Tahir MH, Zubair M, Ghazali SA. The gamma half-Cauchy distribution: properties and applications. *Hacet J Math Stat*. 2016; 45(4): 1143-1159.
- Atikankul Y, Thongteeraparp A, Bodhisuwan W, Volodin A. The Length-Biased Weighted Lindley Distribution with Applications. *Lobachevskii J Math*. 2020; 41(3): 308-319.
- Bhunya PK, Mishra SK, Berndtsson R. Simplified Two-Parameter Gamma Distribution for Derivation of Synthetic Unit Hydrograph. *J Hydrolog Eng*. 2003; 8(4): 226-230.
- Bidaoui H, Abbassi IE, Bouardi AE, Darcherif A. Wind Speed Data Analysis Using Weibull and Rayleigh Distribution Functions, Case Study: Five Cities Northern Morocco. *Procedia Manuf*. 2019; 32: 786-793.
- Blain GC. Standardized precipitation index based on pearson type III distribution. *Rev Bras Meteorol*. 2011; 26(2): 167-180.
- Chaito T, Khamkong M. A modified box and cox power transformation to determine the standardized precipitation index. *Songkla J Sci Technol*. 2018; 40(4): 867-877.
- Chaito T, Khamkong M, Murnta P. Appropriate transformation techniques to determine a modified standardized precipitation index for the ping river in Northern Thailand. *EnvironmentAsia*. 2019; 12(3): 32-42.
- Cordeiro GM, Mansoor M, Provost SB. The Harris extended Lindley distribution for modeling hydrological data. *Chil J Stat*. 2019; 10(1): 77-94.
- Chang TP. Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. *Appl Energ*. 2011; 88(1): 272-282.
- Das KK, Roy TD. On some length-biased weighted Weibull distribution. *Adv Appl Sci Res*. 2011a; 2(5): 465-475
- Das KK, Roy TD. Applicability of length biased weighted generalized Rayleigh distribution. *Adv Appl Sci Res*. 2011b; 2(4): 320-327.
- De Andrade TA, Fernandez LMZ, Gomes-Silva F, Coreiro GM. The gamma generalized pareto distribution with applications in survival analysis. *Int J Stat Probab*. 2017; 6(3): 141-156.
- Fisher RA. The effect of methods of ascertainment upon the estimation of frequencies. *Ann Eugen*. 1934; 6: 13-25.
- Ganaie RA, Rajagopalan V. Length biased weighted Quasi gamma distribution with characterizations and applications. *Int J Innovat Res Comput Comm Eng*. 2020; 9(5): 1110-1117.
- Ghorbani MA, Singh VP, Sivakumar B, Kashani MH, Atre AA, Asadi H. Probability distribution functions for unit hydrographs with optimization using genetic algorithm. *Appl Water Sci*. 2017; 7: 663-676.

- Husak GJ, Michaelsen J, Funk C. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *Int J Climatol*. 2007; 27(7): 935-944.
- Khamkong M, Bookkamana P. Development of statistical models for maximum daily rainfall in upper northern region of Thailand. *Chiang Mai J Sci*. 2015; 42(4): 1044-1053.
- Nanuwong N. The generalized of mixture Pareto distributions. PhD [dissertation]. Bangkok: Kasetsart University; 2015.
- Nasiru S, Luguterah A. The new Weibull-Pareto distribution. *Pak J Stat Oper Res*. 2015; 11(1): 103-114.
- Provost SB, Saboor A, Ahmad M. The gamma-weibull distribution. *Pak J Stat*. 2011; 27(2): 111-131.
- Rao C. On discrete distributions arising out of methods of ascertainment. *Sankhya Ser A*. 1965; 27(2/4): 331-324.
- Rather AA, Subramanian C. length biased Sushila distribution *Uni Review J*. 2018; 7(12): 1010-1023.
- Ristić MM, Balakrishnan N. The gamma-exponentiated exponential distribution. *J Stat Comput Simul*. 2012; 82(8): 1191-1206.
- Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978; 6(2): 461-464.
- Seenoi P, Supapakorn T, Bodhisuwan W. The length-biased exponentiated inverted Weibull distribution. *Int J Pure Appl Math*. 2014; 92(2): 191-206.
- Sharma MA, Singh JB. Use of probability distribution in rainfall analysis. *New York Sci J*. 2010; 3(9): 40-49.
- Thai Meteorological Department Automatic Weather System. Monthly statistics data [Internet]. 2008 [cited 2020 June 15]. Available from: http://www.aws-observation.tmd.go.th/web/reports/weather_months.asp.
- Upper Northern Region Irrigation Hydrology Center. Runoff data [Internet]. 2020 [cited 2020 June 15]. Available from: <http://www.hydro-1.net/>.
- Voytenko D. Modeling direct runoff hydrographs with the surge function: Master [dissertation]. Florida: University of South Florida; 2011.