Songklanakarin J. Sci. Technol.
44 (4), 1131–1144, Jul. – Aug. 2022

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

Original Article

A mixture Weibull-Rayleigh distribution and its application

Tanachot Chaito[1], Nawapon Nakharutai[2], Sirima Suwan[2],
Lampang Saenchan[2], and Manad Khamkong[2*]

[1] *Applied Statistics Program, Department of Statistics, Faculty of Science, Chiang Mai University, Mueang, Chiang Mai, 50200 Thailand*

[2] *Data Science Research Center, Department of Statistics, Faculty of Science, Chiang Mai University, Mueang, Chiang Mai, 50200 Thailand*

Received: 10 March 2021; Revised: 10 May 2022; Accepted: 20 June 2022

Abstract

In this paper, we introduced a mixture Weibull-Rayleigh (MWR) distribution, which was generated by the two-component mixture distribution, i.e., Weibull-Rayleigh and length-biased Weibull-Rayleigh distributions. We studied its properties such as the r^{th} moment, the survival function and the sub-model of the MWR distribution. We used the maximum likelihood estimation, the maximum product of spacing estimators, the Anderson-Darling minimum distance estimators and the Cramer-von Mises minimum distance estimators to estimate the parameters of the MWR distribution. Comparing with the lognormal, Weibull-Rayleigh, length-biased Weibull-Rayleigh, mixture generalized gamma and mixture exponentiated inverted Weibull distributions, we present an application of the MWR distribution on fitting hydrological datasets. We found that the MWR distribution provided a better fitting among these distributions. Therefore, we applied the MWR distribution to predict the return periods of such data.

Keywords: Weibull-Rayleigh distribution, maximum product of spacing estimators, Anderson-Darling minimum distance estimators

1. Introduction

Hydrological data, such as rain, runoff, streamflow rate, flow velocity and water surface elevation, are important to monitor the water situation. They are also used to analyze or predict the water situation in order to make decisions in planning and managing water in each area (Chow, Maidment, & Mays, 1988; Rittima, 2018). Hydrological data are usually random variables (Rittima, 2018) and could be right-skewed or left-skewed or symmetric, which depend on the spatial and temporal dimensions of the study. In addition, extreme values of hydrological data can be found when natural phenomena such as droughts or floods happen (Chow *et al.*, 1988; Rittima, 2018). Thus, various researchers have attempted to fit several distributions or models for hydrological data (Boonma & Tasaduak, 2021; Chaito, Khamkong, & Murnta, 2019; Cordeiro, Mansoor, & Provost, 2019; Yue & Hashino, 2007). Various statistical distributions have been used to model hydrological data such as lognormal (Boonma & Tasaduak, 2021; Chaito *et al.*, 2019), Weibull (Chaito *et al.*, 2019), gamma (Chaito *et al.*, 2019), Pareto (Cordeiro *et al.*, 2019), Pearson type III (Boonma & Tasaduak, 2021; Yue & Hashino, 2007), log-Pearson Type III (Yue & Hashino, 2007), and generalized extreme value (Cordeiro *et al.*, 2019) distributions.

Among these studies, the Weibull distribution, which is positively skewed and a heavy-tailed distribution, is often applied to such data. Therefore, various researchers have developed new distributions by mixing the Weibull distribution with other distributions to provide better fitting of the data. For example, the Weibull-Pareto (Alzaatreh, Famoye, & Lee, 2013), the Weibull-exponential (Oguntunde, Balogun, Okagbue, & Bishop, 2015), and the Weibull-Fréchet

*Corresponding author
Email address: manad.k@cmu.ac.th

(Afify, Yousof, Cordeiro, Ortega, & Nofal, 2016) distributions that have been used to fit reliability and lifetime data. However, as far as we know, only the Weibull-Rayleigh distribution (Ganji, Bevrani, Hami Golzar, & Zabihi, 2016) has been used to fit flood data.

Weibull-Rayleigh (WR) distribution was introduced by Ganji *et al.* (2016). Let X be a WR random variable, denoted by X~WR (α, β, δ), then the probability density function (pdf) is given by

$$f(x)=\frac{\alpha x}{\beta\delta^2}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha-1}\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right], \tag{1}$$

where $x>0$, $\alpha>0$ is a shape parameter, and $\beta>0$ and $\delta>0$ are scale parameters. The cumulative distribution function (cdf) can be written as

$$F(x)=1-\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]. \tag{2}$$

Ganji *et al.,* (2016) showed that, among beta-Pareto, generalized exponential, Weibull, three-parameter Weibull, and Pareto distributions, the WR distribution can provide a better fit for the exceedance of flood peaks data of the Wheaton River near Carcross in Yukon Territory, Canada.

Later, Chaito and Khamkong (2021) presented the length-biased Weibull-Rayleigh (LBWR) distribution, which was modified from the WR distribution using the length–biased distribution. Let X be a LBWR random variable, denoted by X~LBWR (α, β, δ), then the pdf can be written as

$$f(x)=\frac{\alpha x^2}{\beta\delta^2\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha-1}\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right], \tag{3}$$

where $x > 0$, $\alpha > 0$is a shape parameter, and $\beta > 0$ and $\delta > 0$ are scale parameters. The cdf is given by

$$F(x)=\frac{\gamma\left(1+\frac{1}{2\alpha},\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}, \tag{4}$$

where $\Gamma(\alpha)=\int_0^{\infty}u^{\alpha-1}e^{-u}du$ is a gamma function, $\gamma(\alpha,x)=\int_0^{x}u^{\alpha-1}e^{-u}du$ is a lower incomplete gamma function. In their study, the LBWR distribution performed better fitting of the flood data than the Rayleigh, Weibull, Pareto, and WR distributions. Nevertheless, there is still room for improvement.

There are various methods for developing statistical distributions. Mixture distribution method is commonly and widely used to create statistical distributions for economic, environmental and reliable data (Nanuwong, 2015; Seenoi, 2014). The mixture distribution can be formed of a mixture of two or more probability distributions. Afterwards, Newcomb (1886) introduced the finite mixture distribution. This method is a mixture of a finite number of distributions in different proportions of distribution via a mixture weight. An advantage of the finite mixture distribution is that it can be easily modified to create an appropriate distribution for the data (Ghosh, Hamedani, Bansal, & Maadooliat, 2018). Various researchers have applied the finite mixture distribution to improve distributions for fitting data, such as the mixture exponentiated inverted Weibull (MEIW) distribution (Seenoi, 2014), the mixture Pareto distribution (Nanuwong, Bodhisuwan, & Pudprommarat, 2015) and the mixture Weibull and Pareto (IV) distribution (Ghosh *et al.*, 2018), all of which were used to fit lifetime data. Meanwhile, the mixture generalized gamma (MGG) distribution (Suksaengrakcharoen & Bodhisuwan, 2014), the mixture beta-Pareto distribution (Nanuwong, 2015) and the mixture gamma Weibull distribution (Chen, 2020) were applied to hydrological data for fitting hydrological data.

The contribution of this study is to propose a new distribution called the mixture Weibull-Rayleigh (MWR) distribution, which is a mixture of WR and LBWR distributions. The MWR distribution will be compared the efficiency for fitting hydrological data with the following distributions, i.e., lognormal, WR, LBWR, MGG and MEIW distributions. Furthermore, the MWR distribution will be used to predict the return period to determine the risk of flooding. To do so, the article is organized as follows. Section 2 gives the pdf, the cdf and properties of the MWR distribution. The parameter estimation methods of the MWR distribution are described in Section 3. Section 4 presents simulation studies of comparing the efficiency for estimating parameters of the MWR distribution. Section 5 shows the results of applying the MWR distribution to hydrological data and the outcomes of the return period of hydrological data. Section 6 concludes the study.

2. The MWR Distribution

In this section, we introduce the pdf, the cdf and properties of the MWR distribution. Before that, we will state the pdf and the cdf of the two-component mixture distribution.

Definition 1. Let $f_1(x)$ and $f_2(x)$ be the pdfs of random variables X_1 and X_2, respectively. If $p \in [0, 1]$, is a mixture weight (or a mixture parameter), then the pdf of two-component mixture distribution for a random variable X can be written as

$$f(x) = pf_1(x) + (1-p) f_2(x), \quad x > 0. \tag{5}$$

Definition 2. Let $F_1(x)$ and $F_2(x)$ be the cdfs of random variables X_1 and X_2, respectively. If $p \in [0, 1]$, then the cdf of two-component mixture distribution for a random variable X is given by

$$F(x) = pF_1(x) + (1-p) F_2(x), \quad x > 0. \tag{6}$$

2.1 The pdf and the cdf of the MWR distribution

The two-component mixture distribution will be used to construct the MWR distribution as follows.

Theorem 1. Let X be a random variable of the MWR distribution with parameters p, α, β, and δ, denoted by X~MWR (p, α, β, δ), then the pdf of the MWR distribution is given by

$$f(x) = \left[p + \frac{(1-p)x}{\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)} \right] \left[\frac{\alpha x}{\beta\delta^2}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha-1} \exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\right],$$

where $x > 0$, $0 \le p \le 1$ is a mixture parameter, $\alpha > 0$ is a shape parameter, and $\beta > 0$ and $\delta > 0$ are scale parameters.

Proof. By substituting the pdf of the WR distribution in Equation (1) as $f_1(x)$ and the pdf of the LBWR distribution in Equation (3) as $f_2(x)$ into Equation (5), then the pdf of the MWR can be written as

$$\begin{aligned} f(x) &= p\left\{ \frac{\alpha x}{\beta\delta^2}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha-1} \exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\right\} \\ &\quad + (1-p)\left\{ \frac{\alpha x^2}{\beta\delta^2\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha-1} \exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\right\} \\ &= \left[p + \frac{(1-p)x}{\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)} \right] \left[\frac{\alpha x}{\beta\delta^2}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha-1} \exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\right]. \end{aligned}$$

Theorem 2. Let X be a random variable of the MWR distribution, then the cdf of the MWR distribution can be given as

$$F(x) = p\left[1 - \exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\right] + (1-p)\left[\frac{\gamma\left(1+\frac{1}{2\alpha}, \left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right]. \tag{7}$$

Proof. The cdf of the MWR distribution can be directly obtained by substituting the cdf of the WR and LBWR distributions in Equations (2) and (4) into Equation (6).

Figure 1 (a) illustrates the pdf behaviors of the MWR distribution for several values of p, α, β, and δ. The MWR pdf has various shapes such as the right-skewed shape and close to symmetric shape. Figure 1 (b) displays the MWR cdf curves for several values of p, α, β, and δ as a non-decreasing function. Figure 2 shows the pdf behaviors of the WR, LBWR and MWR distributions for several values of parameters.

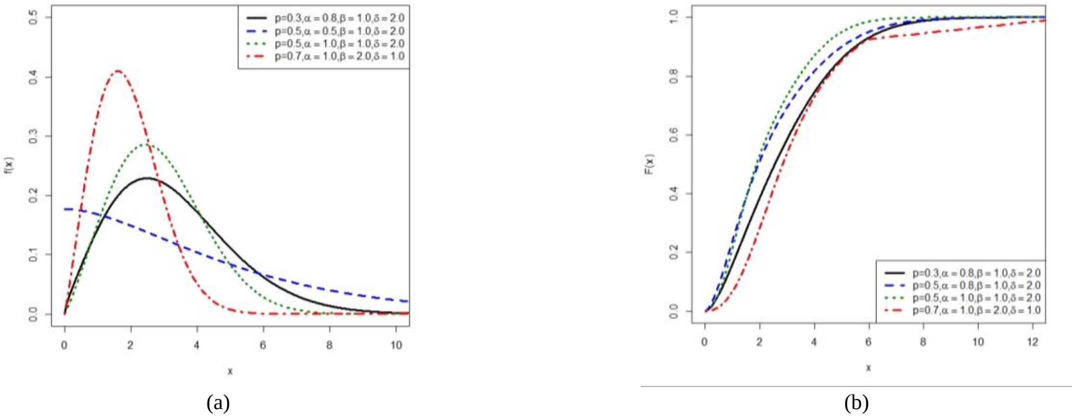

Figure 1. (a) MWR pdfs with vary parameter values and (b) MWR cdfs with vary parameter values

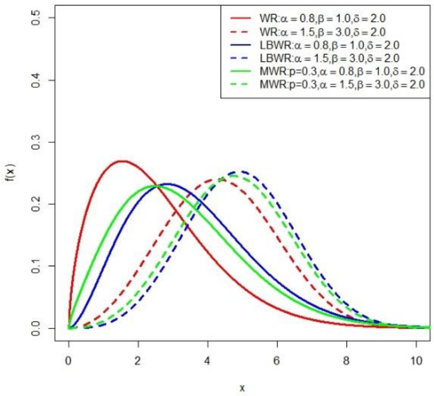

Figure 2. WR, LBWR and MWR pdfs with vary parameter values

2.2 The survival function

The survival function of random variable X with cdf $F(x)$ is given as follows:

$$S(x) = 1 - F(x) \tag{8}$$

By substituting Equation (7) into Equation (8), the survival function of the MWR distribution is

$$S(x) = 1 - \left\{ p\left[1-\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\right] + (1-p)\left[\frac{\gamma\left(1+\frac{1}{2\alpha}, \left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right]\right\}. \tag{8}$$

2.3 The rth moment

Let X be a random variable of the MWR distribution, then the rth moment of X is given by

$$E\left(X^r\right) = \left(2\beta\delta^2\right)^{\frac{r}{2}}\left[p\Gamma\left(1+\frac{r}{2\alpha}\right) + (1-p)\frac{\Gamma\left(1+\frac{r+1}{2\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right], \quad r = 1,2,3,\dots \,. \tag{9}$$

Proof. If $f_1(x)$ and $f_2(x)$ be the pdfs of random variables X_1 and X_2, respectively, then the rth moment of two-component mixture distribution for a random variable X is

$$E\left(X^r\right)=p\int_{-\infty}^{\infty}x^r f_1(x)\,dx+(1-p)\int_{-\infty}^{\infty}x^r f_2(x)\,dx.$$

If X~MWR (p, α, β, δ), then the r$^{\text{th}}$ moment of X becomes

$$\begin{aligned}
E\left(X^r\right)&=p\int_{-\infty}^{\infty}x^r\left\{\frac{\alpha x}{\beta\delta^2}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha-1}\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\right\}dx\\
&\quad+(1-p)\int_{-\infty}^{\infty}x^r\left\{\frac{\alpha x^2}{\beta\delta^2\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha-1}\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\right\}dx\\
&=p\left[\left(2\beta\delta^2\right)^{\frac{r}{2}}\Gamma\left(1+\frac{r}{2\alpha}\right)\right]+(1-p)\left[\frac{\left(2\beta\delta^2\right)^{\frac{r}{2}}\Gamma\left(1+\frac{r+1}{2\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right]\\
&=\left(2\beta\delta^2\right)^{\frac{r}{2}}\left[p\Gamma\left(1+\frac{r}{2\alpha}\right)+(1-p)\frac{\Gamma\left(1+\frac{r+1}{2\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right].
\end{aligned}$$

From the r$^{\text{th}}$ moment of the MWR distribution in Equation (9), we can obtain the mean and the variance of the MWR distribution, respectively, as

$$\mathrm{E}(X)=\sqrt{2\beta\delta^2}\left[p\Gamma\left(1+\frac{1}{2\alpha}\right)+(1-p)\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right],$$

and

$$\mathrm{Var}(X)=2\beta\delta^2\left\{\left[p\Gamma\left(1+\frac{1}{\alpha}\right)+(1-p)\frac{\Gamma\left(1+\frac{3}{2\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right]-\left[p\Gamma\left(1+\frac{1}{2\alpha}\right)+(1-p)\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right]^2\right\}.$$

Moreover, the skewness and the kurtosis of the MWR distribution are respectively expressed as

$$\mathrm{Skewness}(X)=\frac{\kappa_3(p,\alpha)-3\kappa_1(p,\alpha)\kappa_2(p,\alpha)+2\left[\kappa_1(p,\alpha)\right]^3}{\left[\kappa_2(p,\alpha)-\left[\kappa_1(p,\alpha)\right]^2\right]^{\frac{3}{2}}},$$

and

$$\mathrm{Kurtosis}(X)=\frac{\kappa_4(p,\alpha)-4\kappa_1(p,\alpha)\kappa_3(p,\alpha)+6\kappa_2(p,\alpha)\left[\kappa_1(p,\alpha)\right]^2-3\left[\kappa_1(p,\alpha)\right]^4}{\left[\kappa_2(p,\alpha)-\left[\kappa_1(p,\alpha)\right]^2\right]^2},$$

where

$$\kappa_i(p,\alpha)=p\Gamma\left(1+\frac{i}{2\alpha}\right)+(1-p)\frac{\Gamma\left(1+\frac{i+1}{2\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}.$$

2.4 The sub-model of the MWR distribution

Note that the MWR distribution can be transformed to the following distributions with specific parameters.

1. If X~MWR (p = 0, α, β, δ), then the MWR distribution becomes the LBWR distribution with the pdf in Equation (3).

2. If X~MWR ($p = 0, \alpha = 1, \beta = 1, \delta$), then the MWR distribution becomes the length-biased Rayleigh distribution.
3. If X ~MWR ($p = 1, \alpha, \beta, \delta$), then the MWR distribution becomes the WR distribution with the pdf in Equation (1).
4. If X ~MWR ($p = 1, \alpha = 1, \beta = 1, \delta$), then the MWR distribution becomes the Rayleigh distribution.
5. Let X~MWR (p, α, β, δ). If $p = 1$, then a random variable $Y = \frac{X^2}{2\delta^2}$ is Weibull distributed, that is Y~ Weibull (α, β).

3. Parameter Estimation Methods for the MWR Distribution

To estimate the parameters of the MWR distribution, we will use the maximum likelihood estimation (MLE), the maximum product of spacing estimators (MPSE), the Anderson-Darling minimum distance estimators (ADE) and the Cramer-von Mises minimum distance estimators (CMVE). Therefore, in this section, we will explain these methods.

3.1 Maximum likelihood estimation

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from the MWR distribution with a parameter vector $\theta = (p, \alpha, \beta, \delta)$, and let $x_1, x_2, \ldots, x_n$ be the observed values, then the likelihood function is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left\{\left[p + \frac{(1-p)x_i}{\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)}\right]\frac{\alpha x_i}{\beta\delta^2}\left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha-1}\exp\left[-\left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha}\right]\right\}.$$

The log-likelihood function can be obtained as follows:

$$\begin{aligned}\log L(\boldsymbol{\theta}) = {} & \sum_{i=1}^{n}\log\left[p\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)+(1-p)x_i\right] - n\log\left(\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)\right) + n\log\alpha \\ & + \sum_{i=1}^{n}\log x_i - n\log\beta - 2n\log\delta + (\alpha-1)\sum_{i=1}^{n}\log\left(\frac{x_i^2}{2\beta\delta^2}\right) - \sum_{i=1}^{n}\left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha}.\end{aligned} \tag{10}$$

The maximum likelihood estimates of p, α, β and δ are obtained by differentiating the log-likelihood function in Equation (10) with respect to p, α, β and δ and setting the results equal to zero:

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial p} = \sum_{i=1}^{n}\left[\frac{\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right) - x_i}{p\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)+(1-p)x_i}\right], \tag{11}$$

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\theta})}{\partial \alpha} = {} & \sum_{i=1}^{n}\left[\frac{p\sqrt{2\beta\delta^2}\,\varphi\left(\Gamma\left(1+\frac{1}{2\alpha}\right)\right)}{p\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)+(1-p)x_i}\right] - \frac{n\varphi\left(\Gamma\left(1+\frac{1}{2\alpha}\right)\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)} \\ & + \frac{n}{\alpha} + \sum_{i=1}^{n}\log\left(\frac{x_i^2}{2\beta\delta^2}\right) - \sum_{i=1}^{n}\left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha}\log\left(\frac{x_i^2}{2\beta\delta^2}\right),\end{aligned} \tag{12}$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \beta} = \sum_{i=1}^{n}\left[\frac{p\delta^3\sqrt{2\beta}}{\sqrt{2\beta\delta^2}\left(p\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)+(1-p)x_i\right)}\right] - \frac{n\alpha}{\beta} + \frac{2\alpha\delta^2}{\left(2\beta\delta^2\right)^{\alpha+1}}\sum_{i=1}^{n}x_i^{2\alpha}, \tag{13}$$

and

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \delta} = \sum_{i=1}^{n}\left[\frac{2\sqrt{2}\,p\beta^{\frac{3}{2}}\delta^2}{\sqrt{2\beta\delta^2}\left(p\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)+(1-p)x_i\right)}\right] - \frac{2n\alpha}{\delta} + \frac{4\alpha\beta\delta}{\left(2\beta\delta^2\right)^{\alpha+1}}\sum_{i=1}^{n}x_i^{2\alpha}, \tag{14}$$

where $\varphi(y)=\frac{d}{dy}\Gamma(y)=\frac{\Gamma'(y)}{\Gamma(y)}$ is a logarithmic derivative of the gamma function. The maximum likelihood estimates of the parameters p, α, β and δ can be obtained numerically from the non-linear equations Equation (11) - Equation (14). We solve the system of equations in order to calculate MLEs with the Newton-Raphson method using the *mle* function in stats4 package in R program (R Core Team, 2020).

3.2 The maximum product spacing estimators

The maximum product spacing estimators (MPSE) method was proposed by Cheng and Amin (1983). Later, Ranneby (1984) developed the MPSE method to be an approximation to the Kullback-Leibler measure of information. The MPSE method can be a good choice for MLE to estimate the unknown parameters of continuous univariate distributions (Al-Mofleh, Afify, & Ibrahim, 2020). Let $X_1, X_2, \ldots, X_n$ be random samples from MWR distribution having the cdf $F(p, \alpha, \beta, \delta)$ and $X_{(1)} < X_{(2)} < \ldots < X_n$ represent the corresponding ordered samples, then the geometric mean of the spacings can be defined as

$$G(p,\alpha,\beta,\delta)=\left[\prod_{i=1}^{n+1}D_i(p,\alpha,\beta,\delta)\right]^{\frac{1}{n+1}},\quad i=1,2,\ldots,n+1,$$

where $D_i(p,\alpha,\beta,\delta)=F(x_i \mid p,\alpha,\beta,\delta)-F(x_{i-1} \mid p,\alpha,\beta,\delta)$. When $F(x_0 \mid p,\alpha,\beta,\delta)=0$, $F(x_{n+1} \mid p,\alpha,\beta,\delta)=1$ and $\sum_{i=1}^{n+1}D_i(p,\alpha,\beta,\delta)=1$. The logarithm of the geometric mean of sample spacings is

$$H(p,\alpha,\beta,\delta)=\frac{1}{n+1}\sum_{i=1}^{n+1}\log D_i(p,\alpha,\beta,\delta). \tag{15}$$

The MPSE of parameters p, α, β and δ can be obtained by differentiating the logarithm of the geometric mean of sample spacings in Equation (15) with respect to p, α, β and δ and setting them equal to zero:

$$\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(p,\alpha,\beta,\delta)}\times\left[\Delta_s(x_i \mid p,\alpha,\beta,\delta)-\Delta_s(x_{i-1} \mid p,\alpha,\beta,\delta)\right]=0,\quad s=1,2,3,4 \tag{16}$$

When

$$\Delta_1(\cdot \mid p,\alpha,\beta,\delta)=\left[1-\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\right]-\left[\frac{\gamma\left(1+\frac{1}{2\alpha},\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right], \tag{17}$$

$$\begin{aligned}\Delta_2(\cdot \mid p,\alpha,\beta,\delta)={}&p\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\log\left(\frac{x^2}{2\beta\delta^2}\right)\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]\\&+(1-p)\left[\frac{\psi\left(\gamma\left(1+\frac{1}{2\alpha},\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)\right)}{\Psi\left(\Gamma\left(1+\frac{1}{2\alpha}\right)\right)}\right],\end{aligned} \tag{18}$$

$$\Delta_3(\cdot \mid p,\alpha,\beta,\delta)=\frac{-\alpha p}{\beta}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right]+(1-p)\left[\frac{\psi\left(\gamma\left(1+\frac{1}{2\alpha},\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right], \tag{19}$$

and

$$\Delta_4(\cdot|\ p,\alpha,\beta,\delta) = \frac{-2\alpha p}{\delta}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right] + (1-p)\left[\frac{\psi\left(\gamma\left(1+\frac{1}{2\alpha},\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}\right]. \tag{20}$$

where $\Psi(w) = \frac{d}{dw}\Gamma(w)$ is a derivative of the gamma function and $\psi(z) = \frac{d}{dz}\gamma(z)$ is a derivative of the lower incomplete gamma function. The MPSE of parameters p, α, β and δ can be obtained numerically from the non-linear Equation (16). We solve the system of equations in order to calculate MPSEs with the Newton-Raphson method using the *optim* function in stats package in R program (R Core Team, 2020).

3.3 The Anderson-Darling minimum distance estimators

The Anderson-Darling estimator (ADE), a type of minimum distance estimator, is based on an Anderson-Darling (AD) statistic (Anderson & Darling, 1952) and classified as quadratic empirical distribution function (EDF) statistics. Let $X_1, X_2, \ldots, X_n$ be random samples from MWR distribution having the cdf $F(p, \alpha, \beta, \delta)$ and $X_{(1)} < X_{(2)} < \ldots < X_n$ represent the corresponding ordered samples, then the ADE of the MWR parameters are obtained by minimizing

$$A(p,\alpha,\beta,\delta) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log F(x_i|\ p,\alpha,\beta,\delta) - \log S(x_{n+1-i}|\ p,\alpha,\beta,\delta)\right]. \tag{21}$$

The ADE can be obtained by taking derivative Equation (21) with respect to p, α, β and δ and setting them equal to zero:

$$\sum_{i=1}^{n}(2i-1)\left[\frac{\Delta_s(x_i|\ p,\alpha,\beta,\delta)}{F(x_i|\ p,\alpha,\beta,\delta)} - \frac{\Delta_s(x_{n+1-i}|\ p,\alpha,\beta,\delta)}{S(x_{n+1-i}|\ p,\alpha,\beta,\delta)}\right] = 0,\ \ s = 1,2,3,4, \tag{22}$$

where $\Delta_1(\cdot|\ p,\alpha,\beta,\delta)$, $\Delta_2(\cdot|\ p,\alpha,\beta,\delta)$, $\Delta_3(\cdot|\ p,\alpha,\beta,\delta)$ and $\Delta_4(\cdot|\ p,\alpha,\beta,\delta)$ are defined in Equation (17) - Equation (20). The ADE of the parameters p, α, β and δ are obtained by solving Equation (22). We solve the system of equations in order to calculate ADEs with the Newton-Raphson method using the *optim* function in stats package in R program (R Core Team, 2020).

3.4 The Cramer-von Mises minimum distance estimators

The Cramer-von Mises minimum distance estimators (CMVE) is another type of minimum distance estimator and is based on a Cramer-von Mises (CVM) statistic. The CMVE method has less bias of the estimator than the other minimum distance estimators (Al-Mofleh *et al.*, 2020). Let $X_1, X_2, \ldots, X_n$ be random samples from MWR distribution having the cdf $F(p, \alpha, \beta, \delta)$ and $X_{(1)} < X_{(2)} < \ldots < X_n$ represent the corresponding ordered samples, then the CMVE of the MWR parameters are obtained by minimizing

$$C(p,\alpha,\beta,\delta) = \frac{1}{12n} + \sum_{i=1}^{n}\left[F(x_i|\ p,\alpha,\beta,\delta) - \frac{2i-1}{2n}\right]^2. \tag{23}$$

The CMVE can be obtained by taking derivative Equation (23) with respect to p, α, β, and δ and setting them equal to zero:

$$\sum_{i=1}^{n}\left[F(x_i|\ p,\alpha,\beta,\delta) - \frac{2i-1}{2n}\right]\Delta_s(x_i|\ p,\alpha,\beta,\delta) = 0,\ \ s = 1,2,3,4, \tag{24}$$

where $\Delta_1(\cdot|\ p,\alpha,\beta,\delta)$, $\Delta_2(\cdot|\ p,\alpha,\beta,\delta)$, $\Delta_3(\cdot|\ p,\alpha,\beta,\delta)$ and $\Delta_4(\cdot|\ p,\alpha,\beta,\delta)$ are defined in Equation (17) - Equation (20). The CMVE of the parameters p, α, β, and δ can be derived by solving Equation (24). We solve the system of equations in order to calculate CMVEs with the Newton-Raphson method using the *optim* function in stats package in R program (R Core Team, 2020).

4. Simulation Studies

In this section, we will generate the MWR random variables and comparing the efficiencies of the parameter estimation methods for MWR distribution.

4.1 The generating MWR random variables

To generate random data x_i, i = 1, 2, . . ., n that are MWR distributed, we use the inverse transformation method as in the following steps:

1. Generate uniform variable $U_i \sim U(0,1),\ i = 1,2,\ldots,n$.

2. Generate a random variable V_i~WR (α, β, δ), i = 1, 2, . . ., n by taking $v_i = F_{\mathrm{WR}}^{-1}(U_i) = \delta\sqrt{2\beta(-\log(1-U_i))^{\frac{1}{\alpha}}}$, where F_{WR}^{-1} is the inverse function of the cdf in Equation (2).

3. Generate a random variable W_i ~LBWR (α, β, δ), i = 1, 2, . . ., n by taking $w_i = F_{\mathrm{LBWR}}^{-1}(U_i), = \delta\sqrt{2\beta A^{\frac{1}{\alpha}}}$ where F_{LBWR}^{-1} is inverse function of the cdf in Equation (4) and $A = \gamma^{-1}\left[1+\frac{1}{2\alpha}, \Gamma\left(1+\frac{1}{2\alpha}\right)U_i\right]$, when γ^{-1} is inverted of the lower incomplete gamma function and Γ is the gamma function. To compute γ^{-1}, we use *Igamma.Inv* function in *zipfR* package in R program (R Core Team, 2020)."

4. For each i = 1, 2, . . ., n, if $U_i \geq p$, then set $x_i = w_i$, otherwise, set $x_i = v_i$.

4.2 Comparing the efficiencies of the parameter estimation methods for MWR distribution

We will conduct a simulation to compare the performance of the four parameter estimation methods for estimating parameters of the MWR distribution. We perform a simulation study as follows:

1. Set the sample size n and the parameter vector $\theta = (p, \alpha, \beta, \delta)$
2. Generate a random sample from the MWR distribution using various parameters and sample size with *R* program (R Core Team, 2020).
3. Estimate the parameters of the MWR distributions using the MLE, MPSE, ADE and CMVE methods. The estimated parameters will be collected in a parameter vector only if they are inside the parameter space.
4. Repeat steps (1) to (3) until we have 1000 estimated parameter vectors.
5. Calculate the MSE and AvRB values of these methods and select the best method with the lowest values of the MSE and AvRB.

Throughout this simulation, we will consider the cases of p = 0.3, 0.5 and 0.7 α = 0.8, β = 5, and δ = 10. For each case, we varied the sample sizes of x_i for i = 1, 2, . . ., n where n =15, 30, 50, 100, and 200. Next, for each case, we estimate parameters of the MWR distribution. To perform MLE method, we use built-in functions called *mle* in stats4 package in R program to estimate parameters of the MWR distribution. On the other hand, to perform MPSE, ADE and CMVE methods, we use another built-in function called *optim* to estimate parameters of the MWR distribution. Then, we use the mean square error (MSE) and the average relative bias (AvRB) to evaluate the performance of the estimators. The formular of these criteria are provided as fllows:

$$\mathrm{MSE}(\boldsymbol{\theta}) = \frac{1}{M}\sum_{j=1}^{M}\left(\boldsymbol{\theta}_j - \boldsymbol{\theta}\right)^2,$$

$$\mathrm{AvRB}(\boldsymbol{\theta}) = \frac{1}{M}\sum_{j=1}^{M}\left(\boldsymbol{\theta}_j - \boldsymbol{\theta}\right).$$

The MSE and AvRB values for each method are results and are shown in Table 1 - Table 3. According to the results, if the number of sample sizes n is increasing, then most of the values of MSE and AvRB are decreasing. When we fixed α, β, δ and changed the value of p to be p = 0.3 and p = 0.5, we found that in the case of the sample size is smaller than 50, the ADE method provided the lowest values of AvRB, but the MLE method provided the lowest values of AvRB when p = 0.7 and the sample is smaller than 100. Nevertheless, there is no obvious method that provide the lowest value of MSE in such case. In the case that the sample size is either 100 or 200, we found that the CMVE method gave the lowest values of both MSE and AvRB except the case that p = 0.5 and sample size is 200 where there is no obvious method. Overall, we can conclude that the ADE method is recommended to estimate the parameters from the given data in the case that the number of observations is smaller. For the case that the number of observations is large, the CMVE method could be a recommended choice.

In Section 5.2 we will consider the annual maximum runoff data which contain less than 100 record observations, so we will use the ADE method to estimate the parameters.

5. Applications of the MWR Distribution

In this section, the MWR distribution will be applied to fit the hydrological dataset and to predict the return periods of the hydrological dataset.

Table 1. Simulation results of parameter estimates for the MWR distribution with parameters: p= 0.3, α= 0.8, β= 5 and δ= 10.

n	Parameter	MSE				AvRB			
		MLE	MPSE	ADE	CMVE	MLE	MPSE	ADE	CMVE
15	$\hat{p}$	0.2381	0.2599	**0.1437**	0.1787	0.2337	0.1973	**0.1201**	0.1701
	$\hat{\alpha}$	0.0739	0.1738	**0.0499**	0.1024	0.1299	0.2948	**0.0399**	0.1190
	$\hat{\beta}$	1.3252	**1.2215**	1.2286	1.2860	0.4664	0.6079	**0.1693**	0.3043
	$\hat{\delta}$	1.3256	**1.2079**	1.2233	1.2843	0.4698	0.6063	**0.1722**	0.3068
30	$\hat{p}$	0.1984	0.2482	**0.1364**	0.1545	0.1776	0.1982	**0.1217**	0.1230
	$\hat{\alpha}$	0.0334	0.0576	**0.0278**	0.0392	0.0738	0.1574	**0.0236**	0.0621
	$\hat{\beta}$	1.0544	**0.9951**	1.0853	1.0850	0.3415	0.4640	**0.2020**	0.2635
	$\hat{\delta}$	1.0550	**0.9961**	1.0822	1.0843	0.3443	0.4657	**0.2039**	0.2652
50	$\hat{p}$	0.1854	0.2387	**0.1143**	0.1472	0.1669	0.2060	**0.0937**	0.1279
	$\hat{\alpha}$	0.0240	0.0312	**0.0205**	0.0254	0.0535	0.1004	**0.0155**	0.0429
	$\hat{\beta}$	1.0542	1.0149	**0.9165**	0.9599	0.2873	0.4057	**0.1729**	0.2405
	$\hat{\delta}$	1.0539	1.0157	**0.9138**	0.9583	0.2894	0.4072	**0.1740**	0.2419
100	$\hat{p}$	0.1567	0.2189	**0.1144**	0.1271	0.1379	0.2008	0.0863	**0.0858**
	$\hat{\alpha}$	**0.0144**	0.0235	0.0150	0.0171	0.0291	0.0793	**0.0075**	0.0262
	$\hat{\beta}$	0.8636	1.0283	0.8680	**0.8629**	0.2373	0.4012	0.1930	**0.1350**
	$\hat{\delta}$	0.8627	1.0294	0.8656	**0.8608**	0.2384	0.4028	0.1937	**0.1362**
200	$\hat{p}$	0.1419	0.1946	0.1154	**0.1070**	0.1249	0.1713	0.1228	**0.0623**
	$\hat{\alpha}$	0.0116	0.0158	**0.0100**	0.0119	0.0223	0.0491	0.0119	**0.0072**
	$\hat{\beta}$	0.7606	0.9596	0.7715	**0.7367**	0.1888	0.3174	0.2998	**0.0612**
	$\hat{\delta}$	0.7603	0.9606	0.7704	**0.7344**	0.1893	0.3184	0.2999	**0.0621**

Note: The number in bold presents the lowest of MSE and AvRB values.

Table 2. Simulation results of parameter estimates for the MWR distribution with parameters: p= 0.5, α= 0.8, β= 5 and δ= 10.

n	Parameter	MSE				AvRB			
		MLE	MPSE	ADE	CMVE	MLE	MPSE	ADE	CMVE
15	$\hat{p}$	0.1717	0.2126	**0.1356**	0.1490	0.1039	0.1310	**-0.0075**	0.0572
	$\hat{\alpha}$	0.0624	0.1351	**0.0457**	0.1125	0.1156	0.2441	**0.0109**	0.1048
	$\hat{\beta}$	1.0209	**0.9322**	1.0558	1.0529	0.2414	0.3350	**-0.1015**	0.1302
	$\hat{\delta}$	1.0252	**0.9335**	1.0508	1.0504	0.2469	0.3393	**-0.0976**	0.1334
30	$\hat{p}$	0.1579	0.1964	0.1308	**0.1295**	0.0797	0.1333	**0.0101**	0.0385
	$\hat{\alpha}$	0.0227	0.0443	**0.0218**	0.0285	0.0449	0.1355	**-0.0037**	0.0282
	$\hat{\beta}$	0.8910	**0.7127**	0.7998	0.7675	0.1536	0.2938	**-0.0290**	0.0664
	$\hat{\delta}$	0.8921	**0.7158**	0.7995	0.7677	0.1567	0.2971	**-0.0277**	0.0681
50	$\hat{p}$	0.1508	0.1802	**0.1235**	0.1294	0.1032	0.1678	**0.0039**	0.0403
	$\hat{\alpha}$	0.0165	0.0244	**0.0143**	0.0171	0.0366	0.0893	-0.0193	**0.0095**
	$\hat{\beta}$	0.7786	0.7508	0.7425	**0.7171**	0.1791	0.3272	-0.0596	**0.0471**
	$\hat{\delta}$	0.7786	0.7531	0.7411	**0.7163**	0.1810	0.3296	-0.0589	**0.0484**
100	$\hat{p}$	0.1376	0.1696	**0.1106**	0.1159	0.0824	0.1648	**0.0046**	0.0202
	$\hat{\alpha}$	0.0107	0.0156	0.0104	**0.0103**	0.0146	0.0637	-0.0269	**-0.0037**
	$\hat{\beta}$	0.7773	0.7895	0.7132	**0.6668**	0.1353	0.3497	-0.0384	**0.0137**
	$\hat{\delta}$	0.7766	0.7917	0.7114	**0.6659**	0.1364	0.3516	-0.0379	**0.0145**
200	$\hat{p}$	0.1218	0.1537	**0.1023**	0.1083	0.0759	0.1637	0.0245	**-0.0203**
	$\hat{\alpha}$	**0.0078**	0.0104	0.0079	0.0083	**0.0066**	0.0406	-0.0256	-0.0198
	$\hat{\beta}$	0.6694	0.7448	**0.6484**	0.6788	0.0943	0.3103	**0.0187**	-0.1049
	$\hat{\delta}$	0.6690	0.7465	**0.6476**	0.6762	0.0948	0.3114	**0.0189**	-0.1041

Note: The number in bold presents the lowest of MSE and AvRB values.

5.1 Applications of the MWR distribution to hydrological data

We will consider two hydrological datasets and then apply MWR distribution to the data. The first data that we will consider is October rainfall data measured at Mueang Phrae station located in Phrae, Thailand, collected during 1957 to 2018 by the Upper Northern Region Irrigation Hydrology Center, Thailand (Upper Northern Region Irrigation Hydrology Center, 2020) (Table 4). Another hydrological data that we will consider is the annual maximum runoff data collected at Muang, Phayao,

Table 3. Simulation results of parameter estimates for the MWR distribution with parameters: p= 0.7, α= 0.8, β= 5 and δ= 10.

n	Parameter	MSE				AvRB			
		MLE	MPSE	ADE	CMVE	MLE	MPSE	ADE	CMVE
15	$\hat{p}$	0.1268	0.1641	0.1407	**0.1233**	**0.0128**	0.0154	-0.0955	-0.0397
	$\hat{\alpha}$	0.0467	0.1171	**0.0329**	0.0711	0.0950	0.2141	**-0.0197**	0.0635
	$\hat{\beta}$	0.8964	**0.6920**	1.0292	0.8628	**0.0414**	0.0717	-0.2675	-0.1528
	$\hat{\delta}$	0.8977	**0.6900**	1.0261	0.8574	**0.0476**	0.0766	-0.2638	-0.1484
30	$\hat{p}$	0.1140	0.1212	0.1271	**0.1134**	**0.0004**	0.0691	-0.0977	-0.0488
	$\hat{\alpha}$	0.0187	0.0349	**0.0171**	0.0280	**0.0304**	0.1129	-0.0253	0.0128
	$\hat{\beta}$	0.6959	**0.5054**	0.7225	0.5863	**-0.0181**	0.1129	-0.2777	-0.0878
	$\hat{\delta}$	0.6945	**0.5062**	0.7209	0.5862	**-0.0150**	0.1159	-0.2766	-0.0857
50	$\hat{p}$	0.1146	0.1049	0.1165	**0.0983**	**0.0018**	0.0940	-0.0962	-0.0264
	$\hat{\alpha}$	0.0133	0.0186	**0.0119**	0.0132	0.0194	0.0744	-0.0323	**-0.0079**
	$\hat{\beta}$	0.6622	**0.4622**	0.6350	0.5257	**-0.0068**	0.1866	-0.2300	-0.1131
	$\hat{\delta}$	0.6621	**0.4635**	0.6341	0.5247	**-0.0049**	0.1885	-0.2293	-0.1114
100	$\hat{p}$	0.0904	**0.0833**	0.0928	0.0964	**0.0174**	0.1199	-0.0334	-0.0300
	$\hat{\alpha}$	0.0076	0.0095	**0.0068**	0.0074	**0.0073**	0.0484	-0.0278	-0.0137
	$\hat{\beta}$	0.5200	**0.4063**	0.5068	0.4538	**0.0220**	0.2331	-0.1043	-0.0891
	$\hat{\delta}$	0.5196	**0.4070**	0.5065	0.4533	**0.0230**	0.2341	-0.1040	-0.0886
200	$\hat{p}$	0.0837	**0.0763**	0.0814	0.0762	0.0104	0.1006	-0.0121	**0.0039**
	$\hat{\alpha}$	0.0052	0.0057	0.0050	**0.0044**	**-0.0015**	0.0313	-0.0261	-0.0064
	$\hat{\beta}$	0.4780	0.3768	0.4501	**0.3603**	-0.0102	0.1922	-0.0364	**-0.0085**
	$\hat{\delta}$	0.4778	0.3770	0.4504	**0.3602**	-0.0097	0.1928	-0.0365	**-0.0084**

Note: The number in bold presents the lowest of MSE and AvRB values.

Table 4. October rainfall data of the Mueang Phrae station from 1957 to 2018 and annual maximum runoff data of I.17 station from 1993 to 2020 arranged from left to right, and top to bottom.

October rainfall data (mm.)									
23.9	16.3	50.4	70.4	144.6	317.1	253.2	144.2	79.5	24.1
53.2	86.7	95.0	125.4	52.8	122.3	124.9	71.6	111.2	47.6
57.9	92.8	49.1	98.8	157.8	266.5	59.8	94.4	85.4	100.0
72.9	96.2	34.6	48.5	120.6	121.9	84.7	70.2	137.9	71.6
118.2	3.2	1.1	53.0	182.4	81.1	26.7	70.9	79.6	68.9
98.1	35.3	112.5	80.5						
Annual maximum runoff data (m^3/s)									
66.4	77.2	54.3	42.0	26.2	18.8	47.0	49.2	50.3	33.4
0.8	76.0	106.3	64.1	30.8	0.9				

Thailand. The data were collected during 1993 to 2020 by the Upper Northern Region Irrigation Hydrology Center, Thailand (Upper Northern Region Irrigation Hydrology Center, 2020) (Table 4).

To determine whether the MWR distribution can be a better fitting than other distributions in the literature or not, we use the Kolmogorov-Smirnov (KS) test, the Akaike information criterion (AIC) (Akaike, 1973) and Bayesian information criterion (BIC) (Schwarz, 1978) to be criteria for testing their performance and the results were shown in Table 5. It summarizes the parameter estimates based on the MLE method for the October rainfall data for each distribution. As we can see, the KS test and AIC values of the MWR distribution were lower than the lognormal, WR, LBWR, MGG and MEIW distributions. Meanwhile, Figure 3 presents a comparison of fitting the data on histograms and theoretical densities and the empirical data and theoretical cdfs based on the MLE method of the data. The results in both Table 5 and Figure 3 show that the MWR distribution was the best fitting to October rainfall data among our comparison.

Table 6 presents the parameter estimates based on the MLE, MPSE, ADE and CMVE methods for October rainfall data. The results show that the CMVE gave smaller values of the KS test than other methods.

The parameter estimates based on the MLE for the annual maximum runoff data are reported in Table 5. The results show that the AIC, and BIC values for the MWR distribution were lower than the lognormal, WR, LBWR, MGG and MEIW distributions. The comparison of fitting of the histogram and theoretical densities and empirical data and theoretical cdfs of the data is presented in Figure 4.

Table 6 presents the parameter estimates based on the MLE, MPSE, ADE and CMVE methods for the annual maximum runoff data. The results show that the ADE provided a smaller KS test than other methods.

Table 5. The parameter estimates under MLE and goodness of fit statistics for the October rainfall data and the annual maximum runoff data.

	Distribution					
	Lognormal	WR	LBWR	MGG	MEIW	MWR
October rainfall data						
Estimates	$\hat{\mu}$ = 4.2335 $\hat{\sigma}$ = 0.8940	$\hat{\alpha}$ = 0.7507 $\hat{\beta}$ = 14.8200 $\hat{\delta}$ = 18.3871	$\hat{\alpha}$ = 0.5193 $\hat{\beta}$ = 10.7230 $\hat{\delta}$ = 10.2858	$\hat{p}$ = 0.0569 $\hat{\alpha}$ = 1.0272 $\hat{\beta}$ = 0.8602 $\hat{\delta}$ = 0.0274	$\hat{p}$ = 0.1745 $\hat{\alpha}$ = 1.3730 $\hat{\beta}$ = 16.3154	$\hat{p}$ = 0.1286 $\hat{\alpha}$ = 0.5756 $\hat{\beta}$ = 11.5522 $\hat{\delta}$ = 11.9017
KS test	0.1728	0.1136	0.1082	0.1592	0.3682	0.1057
p-value	0.0494	0.4009	0.4623	0.0865	<0.001	0.4923
AIC	691.0005	673.9509	674.2292	677.4453	793.3728	673.8835
BIC	695.2547	680.3323	680.6106	685.9538	799.7542	682.3920
Annual maximum runoff data						
Estimates	$\hat{\mu}$ = 3.4798 $\hat{\sigma}$ = 1.3599	$\hat{\alpha}$ = 0.5746 $\hat{\beta}$ = 14.0714 $\hat{\delta}$ = 18.6015	$\hat{\alpha}$ = 0.4225 $\hat{\beta}$ = 2.4957 $\hat{\delta}$ = 8.6434	$\hat{p}$ = 0.1907 $\hat{\alpha}$ = 3.5578 $\hat{\beta}$ = 0.4347 $\hat{\delta}$ = 1.0804	$\hat{p}$ = 0.0694 $\hat{\alpha}$ = 1.3541 $\hat{\beta}$ = 5.5391	$\hat{p}$ = 0.3396 $\hat{\alpha}$ = 0.5935 $\hat{\beta}$ = 5.6589 $\hat{\delta}$ = 10.6613
KS test	0.2708	0.3593	0.1865	0.1942	0.3290	0.2368
p-value	0.1175	0.0138	0.5006	0.4498	0.0309	0.2260
AIC	191.4239	189.4127	184.6287	187.0308	221.1213	181.5431
BIC	193.2046	192.0838	187.2998	190.5923	223.7925	185.1046

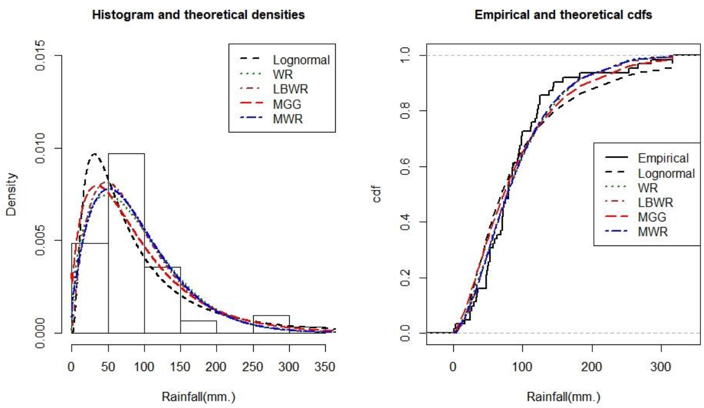

Figure 3. Plots of the observed histogram and estimated pdfs of the fitted distributions (the left column) and the empirical data and estimated cdfs of the fitted distributions (the right column) for the October rainfall data. The horizontal axis represents the October rainfall data (mm). The vertical axis in the left column represents densities while the vertical axis in the right column represents cdfs. The label indicates different distributions.

Table 6. The parameter estimates under MLE, MPSE, ADE and CMVE, and goodness of fit statistics for the October rainfall data and the annual maximum runoff data.

Method	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	KS test	p–value
October rainfall data						
MLE	0.1286	0.5756	11.5522	11.9017	0.1057	0.4923
MPSE	0.0032	0.4941	9.9820	9.9639	0.1210	0.3246
ADE	0.1749	0.5290	8.7651	12.8235	0.1559	0.0982
CMVE	0.0908	0.7670	8.8832	15.8535	0.0855	0.7551
Annual maximum runoff data						
MLE	0.3396	0.5935	5.6589	10.6613	0.2368	0.2260
MPSE	0.3625	0.7476	6.6119	11.6203	0.2271	0.2679
ADE	0.2828	0.6139	6.0441	11.0480	0.1833	0.5218
CMVE	0.6228	0.9267	8.1082	13.1096	0.2708	0.1176

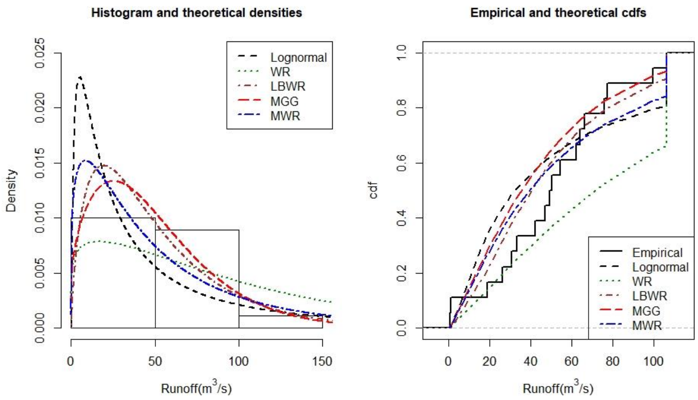

Figure 4. Plots of the observed histogram and the estimated pdfs of the fitted distributions (the left column) and the empirical data and the estimated cdfs of the fitted distributions (the right column) for the annual maximum runoff data. The horizontal axis represents the annual maximum runoff data (m^3/s). The vertical axis in the left column represents densities while the vertical axis in the right column represents cdfs. The label indicates different distributions.

5.2 The return periods of the MWR distribution

A return period (T) is an average time or an estimated average time between events such as floods, or earthquakes to occur (Chow *et al.*, 1988). The return period of events means the period that events will occur once in T years. The return period is

$$T = \frac{1}{P\left(X \geq c\right)},$$

where P is the probability of the occurrence of an extreme event which is defined as having occurred when random variable X is greater than or equal to level C. T is return period, C is threshold value or flood risk value, $F(x)$ is the cdf of the distribution. The return periods of the MWR distribution can be calculated as follows:

$$T = \frac{1}{1 - F\left(x \leq c; \hat{p}, \hat{\alpha}, \hat{\beta}, \hat{\delta}\right)},$$

where T is return period, c is threshold value or flood risk value, $F(x)$ is the cdf of the MWR distributions in Equation (7) with the estimated parameters $\hat{p}$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\delta}$ from the previous subsection.

Note that we have only the threshold value of the annual maximum runoff data, so we use only the annual maximum runoff data to predict the return periods. Moreover, the results of the simulation studies show that ADE gave much lower MSE and AvRB values than other methods and the goodness of fit test in Table 6, the ADE method provided the small KS test values compared to other methods. Therefore, the return period can be predicted via estimated parameters using the ADE in Table 6.

The threshold value or capacity value of I.17 station is 118 m^3/s, which this reported by the Upper Northern Region Irrigation Hydrology Center, Thailand (Upper Northern Region Irrigation Hydrology Center, 2020). If the annual maximum runoff at this station exceeds the capacity value, then this station has a chance of flooding. The return periods based on the MWR distribution of the annual maximum runoff data of I.17 station is

$$T = \frac{1}{1 - F\left(x \leq 118; \hat{p} = 0.2828, \hat{\alpha} = 0.6139, \hat{\beta} = 6.0441, \hat{\delta} = 11.0480\right)} = \frac{1}{1 - 0.9465} = 18.69.$$

The return period of the annual maximum runoff data of I.17 station is 18.69, which means that the period that flood will occur once in 18.69 years.

6. Conclusions

In this study, we proposed the MWR distribution and investigated the survival function and the r^{th} moment and the sub-model of the MWR distribution. The MLE, MPSE, ADE and CMVE methods were used to estimate parameters of the MWR distribution. The simulation study showed that the ADE outperforms other estimation methods. Application of the MWR distribution was conducted on two hydrological datasets, that is, the October rainfall data and annual maximum runoff data.

Based on the KS test, AIC, and BIC values, we found that the MWR distribution gave the best fit for two hydrological datasets. Furthermore, the CMVE was shown to be the best estimation method for the October rainfall dataset, while the ADE was found to be the best estimation method for the annual maximum runoff dataset. We also calculated the return period of the annual maximum runoff data of the station in Muang Phayao based on the MWR distribution. The return period showed that there could be a flood in around 18 years. Finally, we are hopeful and certain that the MWR distribution can be used for the modeling and analysis of hydrological data.

Acknowledgements

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

References

Afify, A. Z., Yousof, H. M., Cordeiro, G. M., Ortega, E. M., & Nofal, Z. M. (2016). The Weibull Fréchet distribution and its applications. *Journal of Applied Statistics, 43*(14), 2608-2626. doi:10.1080/02664763.2016.1142945

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle.* Paper presented at the the 2nd International Symposium on Information Theory. Budapest, Hungary: Akademiai Kiado.

Al-Mofleh, H., Afify, A. Z., & Ibrahim, N. A. (2020). A new extended two-parameter distribution: properties, estimation methods, and applications in medicine and geology. *Mathematics, 8*(9), 1578-1571-1520. doi:10.3390/math8091578

Alzaatreh, A., Famoye, F., & Lee, C. (2013). Weibull-Pareto distribution and its applications. *Communications in Statistics - Theory and Methods, 42*(9), 1673-1691. doi:10.1080/03610926.2011.599002

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics, 23*(2), 193-212.

Boonma, K., & Tasaduak, S. (2021). *Test of frequency distribution analysis of low-flow for Yom Basin.* Paper presented at the the 26th National Convention on Civil Engineering, Rayong, Thailand.

Chaito, T., & Khamkong, M. (2021). The length-biased Weibull-Rayleigh distribution for application to hydrological data. *Lobachevskii Journal of Mathematics, 42*(13), 3253–3265. doi:10.1134/S1995080222010218

Chaito, T., Khamkong, M., & Murnta, P. (2019). Appropriate transformation techniques to determine a modified standardized precipitation index for the Ping River in northern Thailand. *EnvironmentAsia, 12,* 32 - 42. doi:10.14456/ea.2019.43

Chen, H.-W. (2020). Modeling of daily precipitation amounts using the mixed gamma Weibull distribution. Retrieved from https://arxiv.org/abs/2008.13068

Cheng, R. C. H., & Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological), 45*(3), 394-403.

Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied yydrology*. New York, NY: McGraw-Hill.

Cordeiro, G., Mansoor, M., & Provost, S. (2019). The Harris extended Lindley distribution for modeling hydrological data. *Chilean Journal of Statistics, 10*(1), 77-94.

Ganji, M., Bevrani, H., Hami Golzar, N., & Zabihi, S. (2016). The Weibull–Rayleigh distribution, some properties, and applications. *Journal of Mathematical Sciences, 218*(3), 269-277. doi:10.1007/s10958-016-3028-2

Ghosh, I., Hamedani, G. G., Bansal, N., & Maadooliat, M. (2018). On the mixtures of Weibull and Pareto (IV) distribution: An alternative to Pareto distribution. *Communications in Statistics - Theory and Methods, 47*(9), 2073-2084. doi:10.1080/03610926.2016.1171353

Nanuwong, N. (2015). *The generalized mixture Pareto distributions* (Doctoral thesis, Kasetsart University, Bangkok, Thailand),

Nanuwong, N., Bodhisuwan, W., & Pudprommarat, C. (2015). A new mixture Pareto distribution and its application *Thailand Statistician, 13*(2), 191-207.

Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics, 8*(4), 343-366. doi:10.2307/2369392

Oguntunde, P. E., Balogun, O. S., Okagbue, H. I., & Bishop, S. A. (2015). The Weibull-exponential distribution: its properties and applications. *Journal of Applied Sciences, 15*(11), 1305-1311. doi:10.3923/jas.2015.1305.1311

R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ranneby, B. (1984). The maximum spacing method. an estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics, 11*(2), 93-112.

Rittima, A. (2018). Hydrology. Retrieved from http://irre.ku.ac.th/books/pdf/48.pdf

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

Seenoi, P. (2014). *A mixture of exponentiated inverted Weibull distribution* (Doctoral thesis, Kasetsart University, Bangkok, Thailand).

Suksaengrakcharoen, S., & Bodhisuwan, W. (2014). A new family of generalized gamma distribution and its application. *Journal of Mathematics and Statistics, 10*(2), 211-220. doi:10.3844/jmssp.2014.211.220

Upper Northern Region Irrigation Hydrology Center. (2020). Rainfall and runoff data. Retrieved from https://www.hydro-1.net/

Yue, S., & Hashino, M. (2007). Probability distribution of annual, seasonal and monthly precipitation in Japan. *Hydrological Sciences Journal, 52*(5), 863-877. doi:10.1623/hysj.52.5.863