Mathematics

AIMS Mathematics, 10(6): 14539–14559.
DOI: 10.3934/math.2025655
Received: 19 May 2025
Revised: 15 June 2025
Accepted: 23 June 2025
Published: 26 June 2025

https://www.aimspress.com/journal/Math

Research article

Analyzing runoff variability index in northern Thailand using length-biased Weibull-Rayleigh distribution

Tanachot Chaito and Manad Khamkong*

Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand

* **Correspondence:** Email: manad.k@cmu.ac.th.

Abstract: The runoff variability index evaluates the fluctuations in runoff levels. In this article, we outline a method for quantifying the runoff variability index using the length-biased Weibull-Rayleigh (LBWR) distribution and the selecting a suitable parameter estimation technique (the maximum likelihood estimators (MLE), method of moment (MOM), maximum product of spacings estimators (MPSE), Anderson-Darling minimum distance estimators (ADE), and Cramér-von Mises minimum distance estimators (CMVE) methods). Our simulations results indicated that most ADE methods showed enhanced efficiency compared to other estimation methods in terms of mean square error (MSE) and average relative bias (AvRB). This study represents the first investigation into the runoff variability index that integrates the LBWR distribution with ADE parameter estimation. Four stations were studied: Two in Nan province and two in Phrae province. The results indicated that Nan province experiences events more frequently than once every ten years, in contrast to Phrae province. Furthermore, the runoff variability index values are useful for classifying the runoff at the four study locations, which corresponds with the particular geographic conditions at each site. Local and regional authorities can use this runoff variability index values to formulate evidence-based water management strategies, improve flood preparedness, and support long-term water security. This directly contributes to the development of more sustainable and resilient infrastructure in the face of an increasingly variable climate.

Keywords: climate change adaptation; length-biased Weibull-Rayleigh distribution; model selection; parameter estimation; return periods; runoff data
Mathematics Subject Classification: 62E10, 62F10, 62P12

1. Introduction

The upper northern region of Thailand is the origin of four essential rivers: The Ping, Wang, Yom, and Nan. The principal rivers that merge to create the Chao Phraya River, which traverses central Thailand. Thus, the water discharge volume in the primary rivers of upper northern Thailand naturally affects the central area, the economic core of the nation. Runoff analysis in northern Thailand demonstrates intricate hydrological processes within hilly watersheds. Consequently, numerous researchers have sought to create models for predicting runoff volume in northern Thailand's upper regions. The efficiency of runoff is comparatively poor, averaging 20%–25% of precipitation [1]. Groundwater accounts for the predominant portion of stormflow, comprising 62%–80%, followed by shallow subsurface flow at 17%–36%, and surface runoff at 2%–13% [2]. Unpaved roads produce considerable Horton's overland flow, with runoff coefficients of 80% and elevated sediment yield, whereas agricultural fields demonstrate lower runoff coefficients ranging from 0% to 20% [3]. The prediction of rainfall-runoff in ungauged basins can be enhanced by utilizing weather radar for rainfall estimation and employing hydrological modeling tools such as HEC-HMS [4]. The Meteorological Research Institute–Simple Biosphere Model (MRI-SiB) and the Simple Biosphere, including Urban Canopy (SiBUC), are two Land Surface Models (LSMs) that produce runoff characteristics, prompting concerns regarding their influence on river discharge when integrated with the 1-km distributed flow routing model in the upper Ping River basin, Thailand. We examined variations in the simulated river discharge from both LSMs, with special emphasis on discharge volume and the timing of peak discharge [5]. The primary findings concentrate on elucidating the reasons for the disparate runoff estimations between the two models, examining factors such as soil conditions, variations in water budgets, and the fundamental architecture of the models. Climate change has caused heavy rainfall and flooding in northern Thailand to occur more frequently and occasionally with greater severity. This issue affects water resource allocation, agricultural planning, and flood management, particularly in central Thailand. In order to facilitate long-term water warning and management planning, statistical models with skew and heavy tails are more appropriate for hydrological data. However, most research has been on drought indices, and there are no well-defined standards for a runoff variability index in northern Thailand. Several researchers have examined statistical methods for quantifying maximum daily rainfall [6], converting rainfall data into standardized drought indices [7], and developing the LBWR distribution [8], which has been applied to water data in the region. A key feature of the LBWR distribution is its flexibility in modeling skewed and heavy-tailed data, along with its suitability for length-biased sampling, an important property for environmental and hydrological data such as runoff, where longer or extreme events are more likely to be observed. This approach addresses unequal observation probabilities and more accurately represents the characteristics of hydrological extremes. Figure 1 displays the probability density function (pdf) of the LBWR distribution under different parameter settings, demonstrating its adaptability in capturing runoff variability. Considering these factors, the LBWR distribution is a suitable choice for modeling runoff data, which are often skewed and heavy-tailed. Nevertheless, this study has limitations. This research is based on historical runoff data, which could be impacted by biases in observations or problems with data quality. Furthermore, estimating complex parameters for the LBWR distribution can be sensitive to sample size and data variability. The pdf of the LBWR distribution is defined as follows:

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

Where $x>0$, $\alpha>0$ is a shape parameter, and $\beta>0$ and $\delta>0$ are scale parameters. Cumulative distribution function (cdf) of the LBWR distribution is given by

$$F(x)=\frac{\gamma\left(1+\frac{1}{2\alpha},\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}, \tag{1.1}$$

where $\Gamma(\alpha)=\int_0^{\infty}u^{\alpha-1}e^{-u}du$ is a gamma function and $\gamma(\alpha,x)=\int_0^{x}u^{\alpha-1}e^{-u}du$ is a lower incomplete gamma function. The r^{th} moment of the LBWR distribution can be written as

$$E(X^r)=\frac{\left(2\beta\delta^2\right)^{\frac{r}{2}}\Gamma\left(1+\frac{r+1}{2\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)},\quad r=1,2,3,\ldots\ .$$

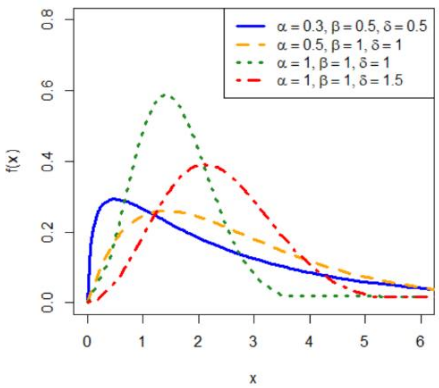

Figure 1. The pdf plots of the LBWR distribution for different parameter values.

This paper is structured as follows: In Section 2, we delineate the research methodology utilized, encompassing a comparison of parameter estimators appropriate for the LBWR distribution, the strategy for formulating the runoff variability index, and the extent of the study. In Section 3, we present obtained from data simulation and the application of the developed framework to runoff data from the Wang and Nan Rivers. In Section 4, we provide a summary, conclusions, and recommendations for practical application and future research.

2. Methodology

In this section, we focus on parameter estimation for the LBWR distribution. This outlines a method for analyzing the runoff variability index, specifies the extent of the data simulation performed for a comparative assessment of appropriate parameter estimations, and delineates the research region employed for the runoff data analysis.

2.1. Parameter estimation for the LBWR distribution

2.1.1. Maximum likelihood estimation

Let $X_1, X_2, \ldots, X_n$ be a random sample from the LBWR distribution with parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \delta)$, and $x_1, x_2, \ldots, x_n$ be the observed values, then the likelihood function can be defined as

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

The log-likelihood function is given by

$$\log L(\boldsymbol{\theta}) = n\log\alpha + \sum_{i=1}^{n} \log x_i^2 - n\log\beta - 2n\log\delta - \frac{1}{2} n\log\left(2\beta\delta^2\right) - n\log\Gamma\left(1+\frac{1}{2\alpha}\right) + (\alpha-1)\sum_{i=1}^{n}\log\left(\frac{x_i^2}{2\beta\delta^2}\right) - \sum_{i=1}^{n}\left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha}. \quad (2.1)$$

By differentiating Eq (2.1) with respect to α, β, and δ, we obtain

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial\alpha} = \frac{n}{\alpha} - n\varphi\left(\Gamma\left(1+\frac{1}{2\alpha}\right)\right) + \sum_{i=1}^{n}\log\left(\frac{x_i^2}{2\beta\delta^2}\right) - \sum_{i=1}^{n}\left[\left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha}\log\left(\frac{x_i^2}{2\beta\delta^2}\right)\right], \quad (2.2)$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial\beta} = -\frac{n\alpha}{\beta} - \frac{n}{2\beta} + \frac{2\alpha\delta^2}{\left(2\beta\delta^2\right)^{\alpha+1}}\sum_{i=1}^{n}\left(x_i\right)^{2\alpha}, \quad (2.3)$$

and

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial\delta} = -\frac{(2\alpha+1)n}{\delta} + \frac{4\alpha\beta\delta}{\left(2\beta\delta^2\right)^{\alpha+1}}\sum_{i=1}^{n}\left(x_i\right)^{2\alpha}, \quad (2.4)$$

where $\varphi(y) = \frac{d}{dy}\Gamma(y) = \frac{\Gamma'(y)}{\Gamma(y)}$ is a logarithmic derivative of the gamma function. The maximum likelihood estimators (MLE) for the parameters α, β, and δ are derived by setting the Eqs (2.2)–(2.4) to zero. This involves solving the resulting system of equations to compute the MLEs using the Newton-Raphson method, implemented through the *mle* function in the stats4 package of the R programming language [9].

2.1.2. Method of moment

Let $x_1, x_2, \ldots, x_n$ be observations from a population with the LBWR distribution which have parameters α, β, and δ, then the first three population moments for estimating α, β, and δ are given by

$$E(X)=\frac{\left(2\beta\delta^2\right)^{\frac{1}{2}}\Gamma\left(1+\frac{1}{\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}, \tag{2.5}$$

$$E(X^2)=\frac{\left(2\beta\delta^2\right)\Gamma\left(1+\frac{3}{2\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}, \tag{2.6}$$

and

$$E(X^3)=\frac{\left(2\beta\delta^2\right)^{\frac{3}{2}}\Gamma\left(1+\frac{2}{\alpha}\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}. \tag{2.7}$$

Correspondingly, the first three sample moments are $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$, $m_2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$, and $m_3=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^3$, respectively. The method of moments (MOM) estimators for parameters α, β, and δ are determined by equating the population moment Eqs (2.5)–(2.7) to the sample moments and solving for the parameters. The MOM method is executed using the built-in function *gmm* from the gmm package in the R program [9].

2.1.3. Maximum product of spacings

To estimate the parameters of a statistical model, the maximum product of spacings estimators (MPSE), is an alternate method to the more popular MLE. Ranneby [10] and Cheng and Amin [11] both separately proposed it. When the likelihood function is unbounded or MLE presents challenges, the MPSE approach is frequently seen as a reliable substitute that can be more effective for small sample numbers. The goal of MPSE is to make the observed data as uniformly spaced as possible on the probability scale. Let $x_{(1)},x_{(2)},\ldots,x_{(n)}$ be the ordered observed values of a sample of size n from $F\left(x_i|\,\alpha,\beta,\delta\right)$, then the geometric mean of the spacings can be defined as

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

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

$$H(\alpha,\beta,\delta)=\frac{1}{n+1}\sum_{i=1}^{n+1}\log D_i(\alpha,\beta,\delta). \tag{2.8}$$

The MPSE of parameters α, β, and δ for the LBWR distribution can be obtained by differentiating the logarithm of the geometric mean of sample spacings in Eq (2.8) with respect to α, β, and δ and setting them equal to zero:

$$\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(\alpha,\beta,\delta)}\left[\Delta_s(x_i\mid\alpha,\beta,\delta)-\Delta_s(x_{i-1}\mid\alpha,\beta,\delta)\right]=0,\quad s=1,2,3, \tag{2.9}$$

where

$$\Delta_1(\cdot\mid\alpha,\beta,\delta)=\frac{\psi\left(\gamma\left(1+\frac{1}{2\alpha},\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)\right)}{\Psi\left(\Gamma\left(1+\frac{1}{2\alpha}\right)\right)}, \tag{2.10}$$

$$\Delta_2(\cdot\mid\alpha,\beta,\delta)=\frac{\psi\left(\gamma\left(1+\frac{1}{2\alpha},\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}, \tag{2.11}$$

and

$$\Delta_3(\cdot\mid\alpha,\beta,\delta)=\frac{\psi\left(\gamma\left(1+\frac{1}{2\alpha},\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right)\right)}{\Gamma\left(1+\frac{1}{2\alpha}\right)}. \tag{2.12}$$

Here, $\Psi(w)=\frac{d}{dw}\Gamma(w)$ 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 α, β, and δ are obtained by solving Eq (2.9). Using the optimized function from the stats package in R program [9], we solve a system of equations to determine the MPSEs with the Newton-Raphson technique.

2.1.4. The Anderson-Darling minimum distance estimators

The method of Anderson-Darling minimum distance estimators (ADE) aims to determine the parameter values of a distribution that minimize a specific goodness-of-fit measure called the Anderson-Darling statistic. This approach falls within a broader category of techniques known as minimum distance estimation. Let $X_1,X_2,\ldots,X_n$ be random samples from the LBWR distribution with parameters α, β, and δ, and $x_{(1)},x_{(2)},\ldots,x_{(n)}$ be the ordered observed values of a sample of size n from $F(x_i\mid\alpha,\beta,\delta)$, then the ADE of the LBWR parameters is obtained by minimizing

$$A(\alpha,\beta,\delta) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\times\left[\log F(x_i \mid \alpha,\beta,\delta) - \log S(x_{n+1-i} \mid \alpha,\beta,\delta)\right], \tag{2.13}$$

where $S(\cdot \mid \alpha,\beta,\delta)$ is the survival function of the LBWR distribution. The ADE can be derived by differentiating Eq (2.13) with respect to α, β, and δ, and equating the results to zero.

$$\sum_{i=1}^{n}(2i-1)\left[\frac{\Delta_s(x_i \mid \alpha,\beta,\delta)}{F(x_i \mid \alpha,\beta,\delta)} - \frac{\Delta_s(x_{n+1-i} \mid \alpha,\beta,\delta)}{S(x_{n+1-i} \mid \alpha,\beta,\delta)}\right] = 0, \; s = 1,2,3, \tag{2.14}$$

where $\Delta_1(\cdot \mid \alpha,\beta,\delta)$, $\Delta_2(\cdot \mid \alpha,\beta,\delta)$, and $\Delta_3(\cdot \mid \alpha,\beta,\delta)$ are defined in Eqs (2.10)–(2.12). The ADE of the parameters α, β, and δ are obtained by solving Eq (2.14). We resolve the system of equations to compute ADEs utilizing the Newton-Raphson approach with the *optim* function in the stats package of the R program [9].

2.1.5. The Cramér-von-Mises minimum distance estimators

The Cramér-von Mises minimum distance estimators (CMVE) method is utilized to estimate the parameters of a probability distribution. It is classified as a minimal distance estimation approach. Let $X_1, X_2, \ldots, X_n$ be random samples from the LBWR distribution with parameters α, β, and δ, and $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ be the ordered observed values of a sample of size n from $F(x_i \mid \alpha,\beta,\delta)$, then the CMVE of the LBWR parameters are obtained by minimizing

$$C(\alpha,\beta,\delta) = \frac{1}{12n} + \sum_{i=1}^{n}\left[F(x_i \mid \alpha,\beta,\delta) - \frac{2i-1}{2n}\right]^2. \tag{2.15}$$

The CMVE can be derived by differentiating Eq (2.15) with regard to α, β, and δ, and equating them to zero:

$$\sum_{i=1}^{n}\left[F(x_i \mid \alpha,\beta,\delta) - \frac{2i-1}{2n}\right]\Delta_s(x_i \mid \alpha,\beta,\delta) = 0, \; s = 1,2,3, \tag{2.16}$$

where $\Delta_1(\cdot \mid \alpha,\beta,\delta)$, $\Delta_2(\cdot \mid \alpha,\beta,\delta)$, and $\Delta_3(\cdot \mid \alpha,\beta,\delta)$ are defined in Eqs (2.10)–(2.12). The CMVE of the parameters α, β, and δ is derived by resolving Eq (2.16). We utilize the Newton-Raphson approach to solve the system of equations for calculating CMVEs, employing the *optim* function from the stats package in R program [9].

2.2. Analysis of runoff variability and extreme events

In this section, we present a methodological approach to assess runoff variability and the occurrence of extreme events. The approach is applied to runoff data from Thailand's upper northern region based on the LBWR distribution. The first step is to determine the best-fit probability distribution for the runoff data by utilizing the Kolmogorov-Smirnov (KS) statistic, coupled with the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The particulars are as follows:

$$\text{KS test} = \sup_x \left[F_n(x_i) - F(x_i) \right],$$

where $F_n(x)$ is the empirical distribution function of the observed data and $F(x)$ is the cdf of the hypothesized distribution. The AIC and BIC are respectively defined as

$$\text{AIC} = 2k - 2\log L(\hat{\boldsymbol{\theta}}),$$

$$\text{BIC} = k\log(n) - 2\log L(\hat{\boldsymbol{\theta}}),$$

where k is the number of parameters, n is the sample size, and $L(\hat{\boldsymbol{\theta}})$ is the maximized value of the likelihood function. The lower values for the KS statistic, AIC, and BIC indicate that the distribution used for modeling runoff is appropriate [9]. In statistical analysis, after identifying the best-fit distribution for the runoff data, the runoff variability index is developed to classify runoff volumes. This index serves as a criterion to monitor runoff within the study area. In addition, the return levels and recurrence periods for specific runoff magnitudes are examined to determine the probability that they will occur again in the future.

2.2.1. Return levels

The return levels of runoff data, where x_T corresponds to a specific return period T, indicate that the level of x_T is expected to be exceeded on average once every T years. The probability of an event exceeding the return level x_T is therefore $1/T$. This study utilized the inverse of the cumulative distribution function of the LBWR distribution to forecast the return levels of the runoff data. The LBWR distribution's T-year return level can be calculated by solving F(x_T) = 1- (1/T) as follows:

(1) Set $\left(1-\frac{1}{T}\right)=F(x_i)$, then $\left(1-\frac{1}{T}\right)=\frac{\gamma\left(1+\frac{1}{2\hat{\alpha}},\left(\frac{x_T^2}{2\hat{\beta}\hat{\delta}^2}\right)^{\hat{\alpha}}\right)}{\Gamma\left(1+\frac{1}{2\hat{\alpha}}\right)}$ and $\Gamma\left(1+\frac{1}{2\hat{\alpha}}\right)\left(1-\frac{1}{T}\right)=\gamma\left(1+\frac{1}{2\hat{\alpha}},\left(\frac{x_T^2}{2\hat{\beta}\hat{\delta}^2}\right)^{\hat{\alpha}}\right)$.

(2) Set $W=\Gamma\left(1+\frac{1}{2\hat{\alpha}}\right)\left(1-\frac{1}{T}\right)$ and $Z=\Gamma^{-1}\left(1+\frac{1}{2\hat{\alpha}},W\right)$ where Γ^{-1} is the inversion of the incomplete gamma function.

(3) Set $Z=\left(\frac{x_T^2}{2\hat{\beta}\hat{\delta}^2}\right)^{\hat{\alpha}}$.

(4) Subsequently, $x_T=\hat{\delta}\sqrt{2\hat{\beta}Z^{\frac{1}{\hat{\alpha}}}}$ where T is return period and $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\delta}$ are the estimated parameters.

2.2.2. Return periods

In general, the return period (T) refers to the average time span over which an event of a specific magnitude or greater is expected to occur once. The probability of an event exceeding or equaling a threshold value (c) within any given time unit is defined as p, where $p = \mathrm{P}(X \geqslant c)$. In other words, T equals $1/p$, which is the reciprocal of this probability. Then, the return period (T) of the LBWR distribution can be calculated as follows:

$$T = \frac{1}{1 - F(c; \hat{\alpha}, \hat{\beta}, \hat{\delta})},$$

where T is the return period and $F(c; \hat{\alpha}, \hat{\beta}, \hat{\delta})$ is the cdf of the LBWR distribution in Eq (1.1) with the estimated parameters $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\delta}$. Salvadori et al. [12] proposed a classification system for return periods, categorizing hazards into four classes based on the frequency of events, as detailed in Table 1.

Table 1. Hazard class based on return period.

Description	Hazard class
Events occur more frequently than once every 10 years	High
Events occur once every 10–100 years	Moderate
Events occur once every 100–1000 years	Low
Events occur less frequently than every 1,000 years	Very Low

2.2.3. Runoff variability index

The development of the runoff variability index involves adapting the criteria from the rainfall variability index [13]. The index is defined as follows:

$$\mathrm{RV}_i = \frac{R_i - \mu}{\sigma}, \tag{2.17}$$

where RV_i represents the runoff variability index for year i, R_i is the annual runoff for that specific year, and μ and σ denote the mean annual runoff and standard deviation, respectively, for the designated study period. As a result of converting the runoff data to a standard normal (Z) distribution described in Eq (2.17), the runoff is classified as Extremely Dry for $R < \mu - 2\sigma$, Dry for $\mu - 2\sigma < R < \mu - \sigma$, Normal for $\mu - \sigma < R < \mu + \sigma$, and Wet for $R > \mu + \sigma$, respectively. An alternative method for calculating the runoff variability index has been proposed. These methods do not require the transformation of runoff data into a standard normal distribution beforehand. This approach is beneficial because runoff data typically exhibit a right skew, which makes normalization using Eq (2.17) problematic. Consequently, we utilize percentiles from the distribution that are considered the most suitable fit for the specific runoff characteristics, as demonstrated in Figure 2.

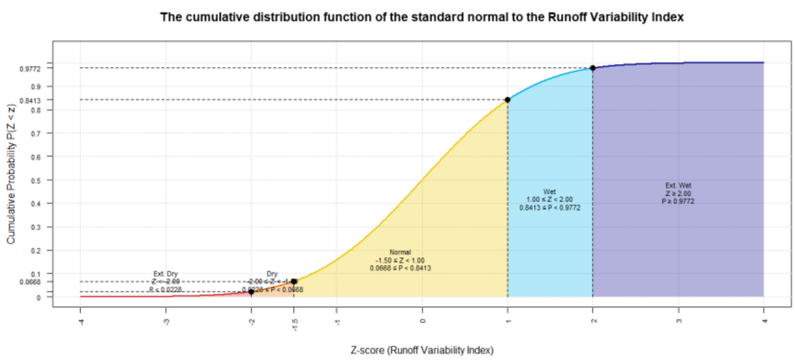

Figure 2. The transformation of the cumulative distribution function of the standard normal distribution into the Runoff Variability Index.

2.3. Scope of study

Our goal of this study is to determine appropriate parameter estimation methods for the LBWR distribution, which has never been investigated previously, by using data simulation. Additionally, we examine the goodness-of-fit of the LBWR distribution, since the true distribution of the data is unknown, using runoff data from northern Thailand's upper region. Moreover, the proposed runoff variability index is also applied to the real data mentioned earlier.

2.3.1. Simulation studies

- *The generation of a LBWR random variate*

The inverse transformation technique is applied to generate random data for the LBWR distribution by setting $U = F^{-1}(x)$, where U is a uniform distribution on (0,1), which denote $U = (0,1)$, and $F^{-1}(x)$ is the inverse of the cdf in Eq (1.1). For generating random data $x_i, i = 1,2,\ldots,n$ from the LBWR distribution, one can use the following steps:

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

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

(3) Set $W_i = U_i\Gamma\left(1+\frac{1}{2\alpha}\right)$ and $t_i = \text{Igamma.Inv}\left(1+\frac{1}{2\alpha}, W_i\right)$ where *Igamma.Inv* is the inversion of the incomplete gamma function in the *zipfR* package in the R program [9].

(4) Set $t_i = \left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha}$.

(5) Subsequently, $x_i = \delta\sqrt{2\beta\left(t_i\right)^{\frac{1}{\alpha}}}$.

- *The comparison of the efficiencies of the parameter estimation methods for the LBWR distribution*

To compare the performance of the five parameter estimation methods for estimating parameters of the LBWR distribution, we perform the simulation studies. To do so, we conduct the following simulation study:

(1) Set the sample size n = 15, 30, 50, 100 and 200, and the parameter vector $\boldsymbol{\theta} = (\alpha,\beta,\delta)$. In this simulation, we consider the cases of α = 0.5, 1.0 and 1.5, β = 5 and δ = 10 because we think these parameters will most likely cover the estimated parameters obtained from the annual maximum runoff data.

(2) Generate a random sample from the LBWR distribution using different parameters and sample sizes with R program [9].

(3) Estimate the parameters of the LBWR distribution using the MLE, MOM, MPSE, ADE, and CMVE methods.

(4) Repeat steps (1) to (2) M = 1,000 times for each experiment.

(5) Compute the $\text{MSE}(\theta) = \frac{1}{M}\sum_{j=1}^{M}(\theta_j - \theta)^2$, and $\text{AvRB}(\theta) = \frac{1}{M}\sum_{j=1}^{M}(\theta_j - \theta)$ values of these methods and choose the optimal method with the lowest MSE and AvRB values.

2.3.2. Study area and runoff data

The annual maximum runoff data were collected by the Hydrology and Water Management Center for the Upper Northern Region of Thailand [14]. For this study, runoff data were selected from two gauging stations on the Wang River in Phrae Province and from two gauging stations on the Nan River in Nan Province. The geographical locations of these stations are shown in Figure 3, and detailed information for each station is provided in Table 2.

Table 2. Station name, location, and study periods for four stations.

Station	Station name	Province	Latitude	Longitude	Study period
Y.1C	Mueang Phrae	Phrae	18.133894	100.124646	1982–2020
Y.20	Song	Phrae	18.584167	100.154722	1982–2020
N.49	Pua	Nan	18.612606	98.868161	1982–2020
N.65	Tha Wang Pha	Nan	18.753869	98.804220	1996–2020

Figure 3. Locations of the runoff stations in Upper Northern Region of Thailand.

3. Results

3.1. Simulation studies

The results of simulations comparing methods for estimating parameters (MLE, MOM, MPSE, ADE, and CMVE) for the LBWR distribution are presented in Tables 3–5, utilizing MSE and AvRB values. The MSE and AvRB values typically decrease as the sample size increases. Evaluations of different estimation methods indicate that ADE and CMVE consistently demonstrate strong performance. Specifically, ADE often yields the lowest MSE, while CMVE reliably produces the lowest AvRB. As a result, both methods are considered reliable options for estimating LBWR parameters. Since there is no closed-form expression for LBWR parameter estimators, the consistency in MSE serves as the key criterion for selecting LBWR estimators. Thus, the ADE method is preferred for its ability to provide relatively accurate parameter estimates for the LBWR distribution.

Table 3. Simulation results of parameter estimates for the LBWR distribution with parameters: $\alpha = 0.5$, $\beta = 5$, and $\delta = 10$.

n	Parameter	MSE					AvRB				
		MLE	**MOM**	**MPSE**	**ADE**	**CMVE**	**MLE**	**MOM**	**MPSE**	**ADE**	**CMVE**
15	$\hat{\alpha}$	0.0400 (3)	0.0169 (1)	0.0931 (5)	0.0242 (2)	0.0540 (4)	0.0845 (2)	0.1063 (4)	0.2179 (5)	0.0244 (1)	0.0868 (3)
	$\hat{\beta}$	7.3570 (5)	0.5019 (1)	4.6439 (4)	3.7370 (3)	3.4869 (2)	1.0171 (4)	0.6349 (3)	1.2326 (5)	-0.2821 (2)	0.0771 (1)
	$\hat{\delta}$	10.2568 (5)	0.4357 (1)	4.8667 (4)	3.6192 (3)	3.5285 (2)	0.4101 (3)	-0.4284 (4)	1.2679 (5)	-0.2399 (2)	0.1491 (1)
30	$\hat{\alpha}$	0.0120 (2)	0.0072 (1)	0.0265 (5)	0.0130 (3)	0.0184 (4)	0.0332 (2)	0.0793 (4)	0.1085 (5)	0.0152 (1)	0.0362 (3)
	$\hat{\beta}$	2.7322 (5)	0.4633 (1)	2.5212 (4)	2.4495 (3)	2.4010 (2)	0.6570 (4)	0.6326 (3)	0.7922 (5)	-0.1214 (2)	-0.0433 (1)
	$\hat{\delta}$	5.4567 (5)	0.3475 (1)	2.5746 (4)	2.2734 (2)	2.3546 (3)	0.0286 (2)	-0.4194 (4)	0.8026 (5)	-0.1028 (3)	-0.0023 (1)
50	$\hat{\alpha}$	0.0074 (3)	0.0051 (1)	0.0122 (5)	0.0067 (2)	0.0108 (4)	0.0213 (2)	0.0688 (4)	0.0695 (5)	0.0070 (1)	0.0219 (3)
	$\hat{\beta}$	2.0158 (5)	0.3568 (1)	1.4934 (3)	1.3959 (2)	1.6965 (4)	0.5031 (3)	0.5594 (5)	0.5400 (4)	-0.1289 (2)	-0.0088 (1)
	$\hat{\delta}$	3.7218 (5)	0.2090 (1)	1.5079 (3)	1.3316 (2)	1.6693 (4)	-0.0186 (2)	-0.3008 (4)	0.5452 (5)	-0.1118 (3)	0.0110 (1)
100	$\hat{\alpha}$	0.0028 (1)	0.0037 (3)	0.0044 (4)	0.0034 (2)	0.0045 (5)	0.0102 (2)	0.0600 (5)	0.0351 (4)	0.0057 (1)	0.0104 (3)
	$\hat{\beta}$	1.0483 (5)	0.2478 (1)	0.7551 (3)	0.7497 (2)	0.8662 (4)	0.4227 (4)	0.4762 (5)	0.3001 (3)	-0.0860 (2)	-0.0217 (1)
	$\hat{\delta}$	1.5472 (5)	0.0740 (1)	0.7602 (3)	0.6626 (2)	0.8436 (4)	-0.2118 (4)	-0.1671 (3)	0.3023 (5)	-0.0721 (2)	-0.0104 (1)
200	$\hat{\alpha}$	0.0014 (1)	0.0031 (5)	0.0019 (3)	0.0014 (1)	0.0020 (4)	0.0063 (3)	0.0552 (5)	0.0197 (4)	0.0013 (1)	0.0035 (2)
	$\hat{\beta}$	0.7245 (5)	0.1951 (1)	0.3721 (3)	0.3290 (2)	0.4003 (4)	0.3321 (4)	0.4288 (5)	0.1781 (3)	-0.0757 (2)	-0.0656 (1)
	$\hat{\delta}$	0.8454 (5)	0.0219 (1)	0.3737 (3)	0.3206 (2)	0.3908 (4)	-0.1348 (4)	-0.0991 (3)	0.1792 (5)	-0.0754 (2)	-0.0656 (1)

Note: The number in parentheses () denotes the rank within each row. A rank of 1 is assigned to the lowest MSE and the lowest AvRB.

Table 4. Simulation results of parameter estimates for the LBWR distribution with parameters: α = 1.0, β = 5, and δ = 10.

n	Parameter	MSE					AvRB				
		MLE	MOM	MPSE	ADE	CMVE	MLE	MOM	MPSE	ADE	CMVE
15	$\hat{\alpha}$	0.1490 (3)	0.1033 (2)	0.2810 (5)	0.0915 (1)	0.1996 (4)	0.1335 (3)	0.1659 (4)	0.3586 (5)	0.0482 (1)	0.1245 (2)
	$\hat{\beta}$	1.0935 (4)	1.4699 (5)	0.6829 (1)	0.7351 (3)	0.6872 (2)	0.5173 (5)	0.4100 (4)	0.3190 (3)	-0.1220 (2)	-0.0740 (1)
	$\hat{\delta}$	2.3576 (5)	0.9880 (4)	0.6742 (2)	0.6946 (3)	0.6689 (1)	-0.3490 (5)	0.0848 (2)	0.3194 (4)	-0.1208 (3)	-0.0641 (1)
30	$\hat{\alpha}$	0.0423 (3)	0.0372 (1)	0.0827 (5)	0.0419 (2)	0.0654 (4)	0.0656 (3)	0.0951 (4)	0.1822 (5)	0.0373 (1)	0.0645 (2)
	$\hat{\beta}$	0.6452 (4)	0.6806 (5)	0.3256 (1)	0.3675 (3)	0.3579 (2)	0.4503 (5)	0.2476 (4)	0.1968 (3)	-0.0405 (2)	-0.0366 (1)
	$\hat{\delta}$	1.1443 (5)	0.5999 (4)	0.3248 (1)	0.3720 (3)	0.3540 (2)	-0.2809 (5)	0.1114 (3)	0.1983 (4)	-0.0550 (2)	-0.0323 (1)
50	$\hat{\alpha}$	0.0215 (1)	0.0259 (3)	0.0366 (5)	0.0220 (2)	0.0318 (4)	0.0408 (2)	0.0860 (4)	0.1112 (5)	0.0181 (1)	0.0435 (3)
	$\hat{\beta}$	0.4229 (4)	0.4721 (5)	0.2186 (1)	0.2219 (2)	0.2310 (3)	0.3589 (5)	0.2759 (4)	0.1357 (3)	-0.0424 (2)	-0.0208 (1)
	$\hat{\delta}$	0.6194 (5)	0.2989 (4)	0.2185 (2)	0.2151 (1)	0.2263 (3)	-0.2560 (5)	0.0645 (3)	0.1371 (3)	-0.0425 (2)	-0.0179 (1)
100	$\hat{\alpha}$	0.0092 (1)	0.0138 (4)	0.0132 (3)	0.0108 (2)	0.0154 (5)	0.0099 (1)	0.0598 (5)	0.0593 (4)	0.0150 (2)	0.0289 (3)
	$\hat{\beta}$	0.2570 (4)	0.3350 (5)	0.0962 (1)	0.1025 (2)	0.1101 (3)	0.2659 (5)	0.1995 (4)	0.0797 (3)	-0.0119 (2)	-0.0057 (1)
	$\hat{\delta}$	0.3448 (5)	0.1693 (4)	0.0964 (1)	0.0989 (2)	0.1086 (3)	-0.2738 (5)	0.0765 (3)	0.0805 (3)	-0.0099 (2)	-0.0046 (1)
200	$\hat{\alpha}$	0.0047 (1)	0.0085 (5)	0.0057 (3)	0.0051 (2)	0.0071 (4)	0.0106 (3)	0.0505 (5)	0.0343 (4)	0.0027 (2)	0.0010 (1)
	$\hat{\beta}$	0.1583 (4)	0.1931 (5)	0.0514 (2)	0.0433 (1)	0.0596 (3)	0.2261 (5)	0.1888 (4)	0.0545 (3)	-0.0082 (1)	-0.0303 (2)
	$\hat{\delta}$	0.1758 (5)	0.0747 (4)	0.0514 (2)	0.0513 (1)	0.0592 (3)	-0.1945(5)	0.0751 (4)	0.0549 (3)	-0.0153 (1)	-0.0312 (2)

Note: The number in parentheses () denotes the rank within each row. A rank of 1 is assigned to the lowest MSE and the lowest AvRB.

Table 5. Simulation results of parameter estimates for the LBWR distribution with parameters: α = 1.5, β = 5, and δ = 10.

n	Parameter	MSE					AvRB				
		MLE	MOM	MPSE	ADE	CMVE	MLE	MOM	MPSE	ADE	CMVE
15	$\hat{\alpha}$	0.2109 (3)	0.2012 (2)	0.4999 (5)	0.1605 (1)	0.3502 (4)	0.1874 (3)	0.2167 (4)	0.4833 (5)	0.0679 (1)	0.1801 (2)
	$\hat{\beta}$	0.5894 (4)	0.9873 (5)	0.2444 (2)	0.2392 (1)	0.2782 (3)	0.4415 (5)	0.2841 (4)	0.1197 (3)	-0.0261 (1)	-0.0410 (2)
	$\hat{\delta}$	1.0965 (5)	0.8906 (4)	0.2433 (2)	0.2171 (1)	0.2734 (3)	-0.3461 (5)	-0.0328 (2)	0.1213 (4)	-0.0250 (1)	-0.0381 (3)
30	$\hat{\alpha}$	0.0822 (2)	0.1100 (3)	0.1582 (5)	0.0769 (1)	0.1382 (4)	0.0804 (2)	0.1613 (4)	0.2450 (5)	0.0414 (1)	0.0846 (3)
	$\hat{\beta}$	0.3468 (4)	0.5473 (5)	0.1261 (2)	0.1218 (1)	0.1374 (3)	0.3738 (5)	0.2288 (4)	0.0679 (3)	-0.0057 (2)	-0.0047 (1)
	$\hat{\delta}$	0.5279 (5)	0.5248 (4)	0.1260 (2)	0.1235 (1)	0.1359 (3)	-0.3276 (5)	-0.0370 (3)	0.0694 (4)	-0.0097 (2)	-0.0052 (1)
50	$\hat{\alpha}$	0.0392 (1)	0.0584 (3)	0.0779 (5)	0.0450 (2)	0.0740 (4)	0.0435 (2)	0.1081 (4)	0.1600 (5)	0.0320 (1)	0.0574 (3)
	$\hat{\beta}$	0.2197 (4)	0.4616 (5)	0.0691 (2)	0.0671 (1)	0.0840 (3)	0.3030 (5)	0.1831 (4)	0.0502 (3)	0.0014 (1)	-0.0152 (2)
	$\hat{\delta}$	0.2971 (4)	0.4015(5)	0.0693 (2)	0.0574 (1)	0.0827 (3)	-0.2607 (5)	-0.0041 (2)	0.0511 (4)	0.0019 (1)	-0.0146 (3)
100	$\hat{\alpha}$	0.0196 (1)	0.0345 (5)	0.0293 (3)	0.0236 (2)	0.0297 (4)	0.0290 (3)	0.0869 (4)	0.0899 (5)	0.0085 (1)	0.0232 (2)
	$\hat{\beta}$	0.1364 (4)	0.3361(5)	0.0372 (1)	0.0395 (2)	0.0413 (3)	0.2470 (5)	0.1475(4)	0.0268 (3)	-0.0060 (1)	-0.0100 (2)
	$\hat{\delta}$	0.1619 (4)	0.2851(5)	0.0373 (2)	0.0352 (1)	0.0394 (3)	-0.2156 (5)	0.0031 (1)	0.0273 (4)	-0.0152 (3)	-0.0109 (2)
200	$\hat{\alpha}$	0.0091 (1)	0.0217 (5)	0.0119 (3)	0.0116 (2)	0.0143 (4)	0.0123 (2)	0.0752 (5)	0.0467 (4)	0.0093 (1)	0.0174 (3)
	$\hat{\beta}$	0.0912 (4)	0.1843 (5)	0.0182 (1)	0.0222 (3)	0.0192 (2)	0.2022 (5)	0.0789 (4)	0.0181 (2)	-0.0190 (3)	-0.0004 (1)
	$\hat{\delta}$	0.0811 (4)	0.1421 (5)	0.0182 (1)	0.0199 (3)	0.0194 (2)	-0.1793 (5)	0.0541 (4)	0.0183 (3)	-0.0039 (2)	0.0010 (1)

Note: The number in parentheses () denotes the rank within each row. A rank of 1 is assigned to the lowest MSE and the lowest AvRB.

3.2. Application to runoff data

Table 6 presents the results of the descriptive statistics for the annual maximum runoff data from four locations. The Y.20 station exhibited the highest average annual maximum runoff at 856.80 m^3/s, whilst the N.65 station recorded the lowest average at 309.90 m^3/s. Figure 4 shows annual maximum runoff data do not follow a normal distribution. The data are right-skewed and display outliers at all four stations. Furthermore, the distributions exhibit high kurtosis, suggesting they have heavier tails than a normal distribution.

Table 6. Descriptive statistics of the annual maximum runoff data (m^3/s*) for four stations.

Station	Min	Max	Q_1	Q_2	Q_3	Mean	SD	Skewness	Kurtosis
Y.1C	161.80	3525.00	466.10	750.20	1054.00	850.20	583.9374	2.6004	12.4532
Y.20	161.10	3851.00	449.80	733.00	1022.00	856.80	637.5332	2.8243	13.7383
N.49	130.40	898.00	239.30	329.60	373.70	339.80	159.9153	1.8561	6.8175
N.65	109.60	1486.40	170.30	260.10	334.00	309.90	273.7456	3.3693	14.9774

Note: *Cubic meter per second ; Qi denotes the ith quartile of data; SD denotes standard deviation.

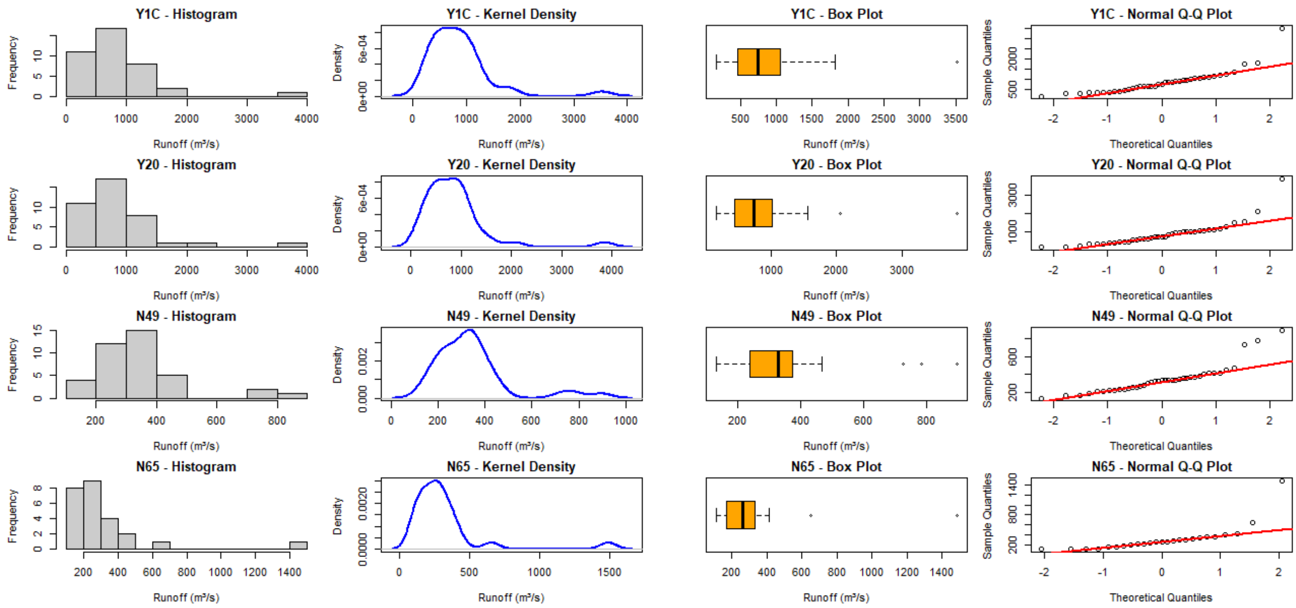

Figure 4. Nonparametric plots for the annual maximum runoff data at stations Y.1C, Y.20, N.49, and N.65.

To determine the optimal method for distributing runoff at the four gauging stations, we examined right-skewed distributions, including Weibull, Rayleigh, Weibull-Rayleigh (WR), Length-Biased Pareto (LP), Length-Biased Rayleigh (LR) and LBWR Distributions. Parameters were estimated using the ADE method. The results, presented in Table 7, show that the LBWR distribution is the most suitable for modeling runoff across all four stations. Similarly, Figure 5 illustrates the theoretical cumulative distribution functions for all four stations, confirming that the LBWR distribution provides the best overall fit.

Table 8 presents the LBWR distribution applied to the runoff data from all four stations, estimating the return levels of the highest annual runoff. The levels, measured in cubic meters per second (m^3/s), correspond to four gauging stations (Y.1C, Y.20, N.49, and N.65) over return periods of 2 to 100 years. A discernible trend is observed with increasing return periods, the runoff return levels elevate across all stations. Stations Y.1C and Y.20 demonstrated markedly higher return levels compared to stations N.49 and N.65 across all assessed periods. The 100-year return levels were 2,127.82 m^3/s for Y.1C and 2,224.33 m^3/s for Y.20, while stations N.49 and N.65 recorded significantly lower values of 668.18 m^3/s and 751.44 m^3/s, respectively. Additionally, a comparison reveals that station Y.20 generally yield higher return levels than Y.1C over extended durations (≥10 years), while station N.65 exceeded N.49 for return periods of 20 years or greater.

Table 7. Summary of selected distributions using the KS test, AIC and BIC for the annual maximum runoff data in four stations.

Station	Distribution	Estimates	KS test	p-value	AIC	BIC
Y.1C	Weibull	$\hat{\alpha} = 1.6388$, $\hat{\beta} = 957.7686$	0.1200	0.5868	593.6780	597.0051
	Rayleigh	$\hat{\delta} = 726.3254$	0.1456	0.3464	595.4306	597.0942
	WR	$\hat{\alpha} = 0.8192$, $\hat{\beta} = 69.1731$, $\hat{\delta} = 81.3543$	0.1195	0.5917	595.6781	600.6688
	LP	$\hat{\alpha} = 1.31229$, $\hat{\beta} = 26.5614$	0.4944	<0.001	687.1774	690.5046
	LR	$\hat{\delta} = 300.5064$	0.5716	<0.001	785.7665	787.4300
	LBWR	$\hat{\alpha} = 0.6046$, $\hat{\beta} = 41.7354$, $\hat{\delta} = 58.75025$	0.1044	0.7503	592.3731	597.3638
Y.20	Weibull	$\hat{\alpha} = 1.5370$, $\hat{\beta} = 959.6792$	0.1418	0.3778	597.3758	600.7030
	Rayleigh	$\hat{\delta} = 751.6945$	0.1946	0.0907	602.2633	603.9269
	WR	$\hat{\alpha} = 0.7685$, $\hat{\beta} = 69.0514$, $\hat{\delta} = 81.6651$	0.1418	0.3777	599.3758	604.3665
	LP	$\hat{\alpha} = 1.2913$, $\hat{\beta} = 21.9849$	0.4688	<0.001	689.6102	692.0479
	LR	$\hat{\delta} = 287.0613$	0.558	<0.001	857.1853	858.8489
	LBWR	$\hat{\alpha} = 0.5555$, $\hat{\beta} = 39.2940$, $\hat{\delta} = 55.4508$	0.1237	0.5484	596.0581	601.0488
N.49	Weibull	$\hat{\alpha} = 2.2499$, $\hat{\beta} = 384.4913$	0.1713	0.1802	503.4991	506.8262
	Rayleigh	$\hat{\delta} = 264.9285$	0.1602	0.2425	502.5579	504.2214
	WR	$\hat{\alpha} = 1.1255$, $\hat{\beta} = 38.6500$, $\hat{\delta} = 43.7387$	0.1714	0.1797	505.4991	510.4898
	LP	$\hat{\alpha} = 1.3758$, $\hat{\beta} = 21.0777$	0.5072	<0.001	607.1232	610.4503
	LR	$\hat{\delta} = 156.5664$	0.4620	<0.001	529.7843	531.4479
	LBWR	$\hat{\alpha} = 0.8996$, $\hat{\beta} = 27.7107$, $\hat{\delta} = 38.5703$	0.1571	0.2621	502.4223	507.4130
N.65	Weibull	$\hat{\alpha} = 1.4176$, $\hat{\beta} = 345.5709$	0.1966	0.2536	335.3171	337.7548
	Rayleigh	$\hat{\delta} = 289.7929$	0.2944	0.0205	342.2104	343.4293
	WR	$\hat{\alpha} = 0.7089$, $\hat{\beta} = 34.3603$, $\hat{\delta} = 41.6962$	0.1967	0.2530	337.3171	340.9737
	LP	$\hat{\alpha} = 1.4175$, $\hat{\beta} = 23.3162$	0.4760	<0.001	374.1661	376.6039
	LR	$\hat{\delta} = 171.0983$	0.1273	<0.001	374.6476	375.8664
	LBWR	$\hat{\alpha} = 0.5160$, $\hat{\beta} = 20.5507$, $\hat{\delta} = 25.2679$	0.1755	0.3799	333.5630	337.2196

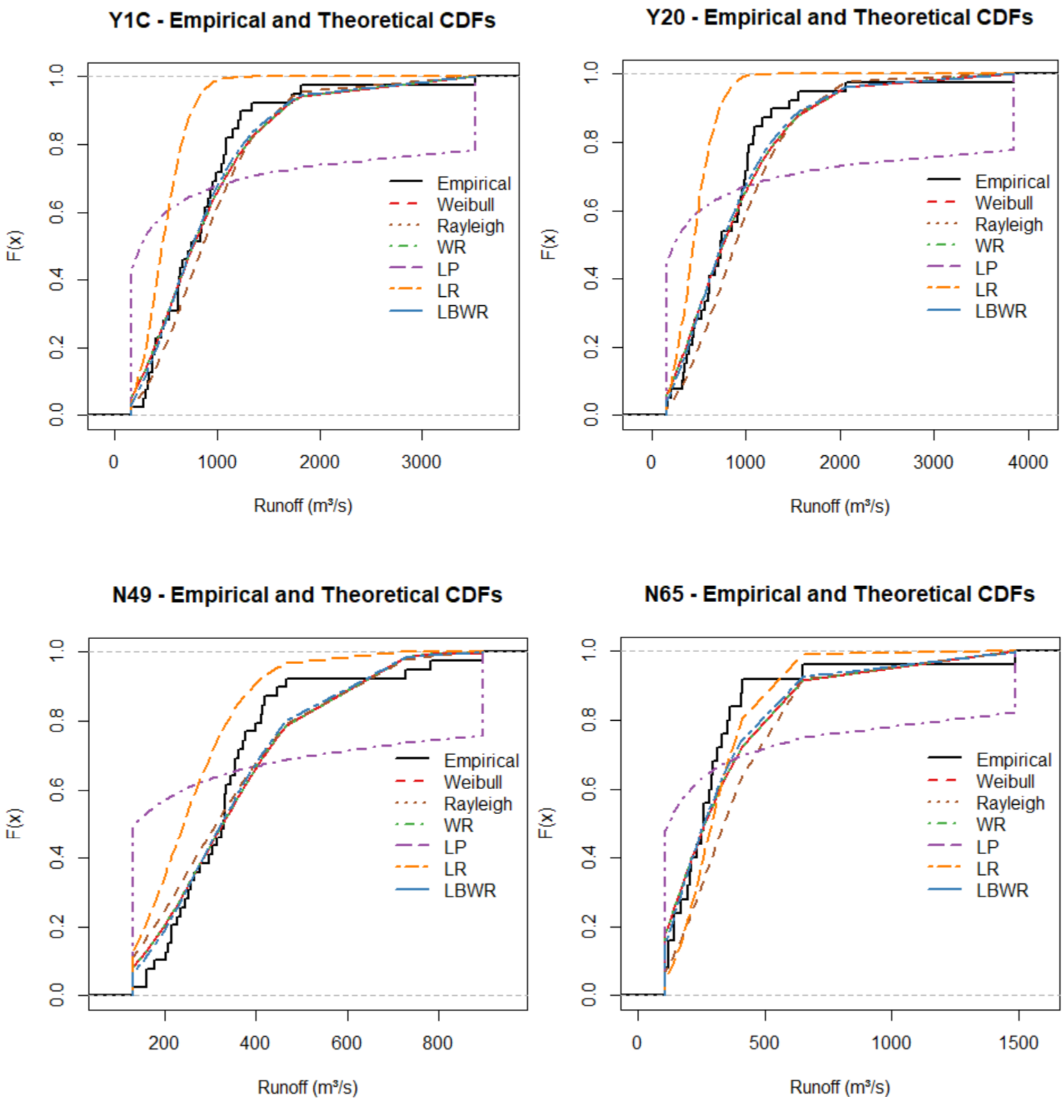

Figure 5. Theoretical cdfs for the annual maximum runoff data at stations Y.1C, Y.20, N.49, and N.65.

Table 8. Return levels of the annual maximum runoff data for four stations.

Station	Return level (m³/s*)							
	2-year	**5-year**	**10-year**	**20-year**	**30-year**	**40-year**	**50-year**	**100-year**
Y.1C	745.87	1,165.37	1,422.37	1,652.98	1,779.31	1,865.88	1,931.48	2,127.82
Y.20	738.78	1,180.03	1,455.15	1,704.53	1,842.04	1,936.60	2,008.44	2,224.33
N.49	316.00	436.20	502.85	559.38	589.23	609.29	624.28	668.18
N.65	253.93	402.74	495.00	578.36	624.23	655.74	679.66	751.44

Note: *Cubic meter per second.

Table 9 displays the estimated return periods for the design capacity runoff at four gauging stations. Station Y.20 exhibits the highest capacity at 2,131.00 m^3/s, with a return period of 73.81 years, categorizing it as a moderate hazard. Conversely, stations Y.1C (992.00 m^3/s), N.49 (319.00 m^3/s), and N.65 (189.00 m^3/s) demonstrate notably shorter return periods of 3.30, 2.04, and 1.48 years, respectively. The reduced return periods result in a high-hazard classification for these locations, indicating that the capacities at three of the four stations are likely to be frequently exceeded.

Table 9. Return periods of the annual maximum runoff data four stations.

Station	Capacity value (m^3/s*)	Return period (year)	Hazard class
Y.1C	992.00	3.30	High
Y.20	2,131.00	73.81	Moderate
N.49	319.00	2.04	High
N.65	189.00	1.48	High

Note: *Cubic meter per second.

The application of percentiles from the LBWR distribution, along with ADE for parameter estimation, facilitates the categorization of runoff variability index values, as demonstrated in Table 10. Stations Y.1C and Y.20 demonstrate considerable interannual variability, with pronounced peaks during 1995-1996 signifying extremely wet circumstances. Station N.49 exhibits fluctuation, with multiple years categorized as moderately wet, and a peak between 2014-2015 nearing the very wet criterion. Station N.65 documented a significant precipitation event early in its lifespan (about 1995-1996), subsequently experiencing predominantly reduced runoff levels. All stations underwent intervals classified as normal and somewhat dry, underscoring the dynamic characteristics of runoff in the region and detailed in Figure 6.

Table 10. Runoff variability index values (m^3/s*) for four stations.

Station	Severely wet	Moderately wet	Normal	Moderately dry	Severely dry
Y.1C	153.43 and less	153.44 to 248.92	248.93 to 1,254.90	1,254.91 to 1,893.10	1,893.11 and above
Y.20	141.87 and less	141.88 to 234.76	234.77 to 1,275.51	1,275.52 to 1,966.40	1,966.41 and above
N.49	97.01 and less	97.02 to 139.97	139.98 to 459.90	459.91 to 615.52	615.53 and above
N.65	49.84 and less	49.85 to 81.95	81.96 to 434.79	434.80 to 665.66	665.67 and above

Note: *Cubic meter per second.

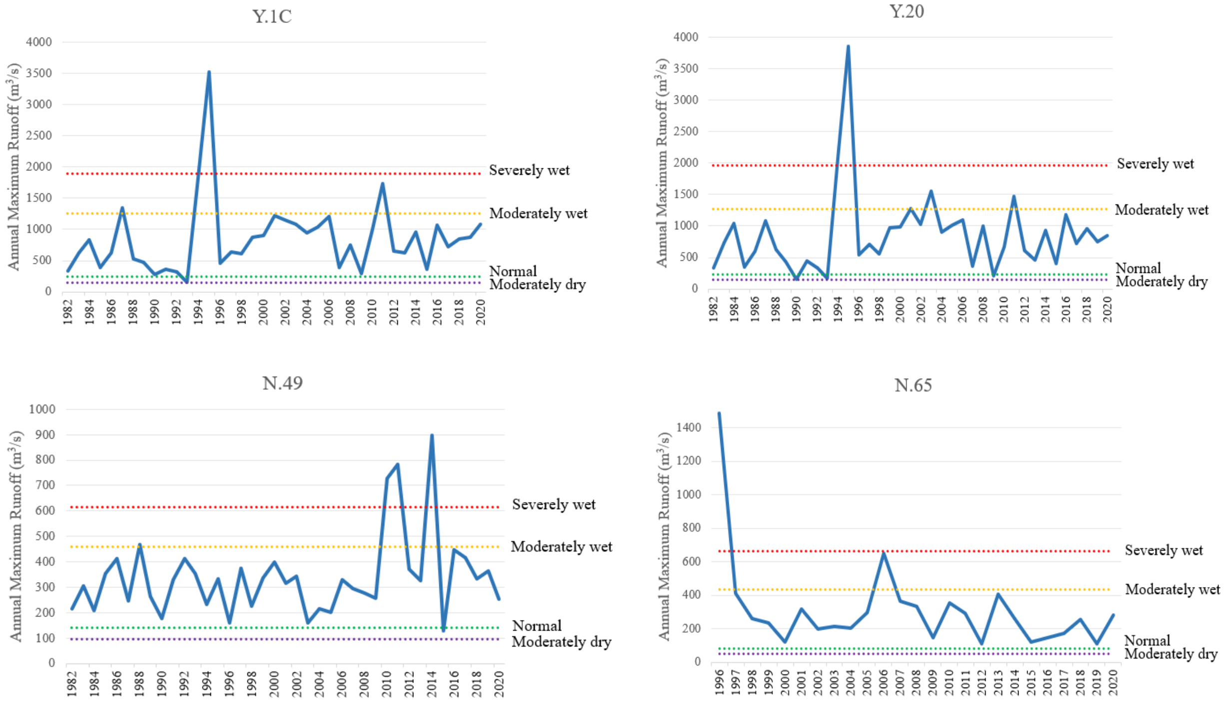

Figure 6. Annual runoff variability indices for four stations.

4. Conclusions and future directions

The runoff in the upper Northern region of Thailand primarily demonstrates a right-skewed pattern. Creating a valuable runoff variability index is crucial for assisting agencies in comprehending alterations in runoff patterns and managing water resources efficiently. In this article, we delineate the procedure for developing a runoff variability index utilizing the LBWR distribution, with parameters derived by the ADE method. Furthermore, percentiles from the LBWR distribution were employed to categorize the runoff variability index, demonstrating its capacity to monitor fluctuations in runoff within Thailand's upper Northern Region. The Wang River basin encountered very wet runoff conditions from 1995 to 1996 as a result of exceptionally heavy rainfall in the middle part of the upper Northern Region [15]. In the Nan River basin, station N.49 in Pua District, Nan Province, documented runoff levels nearing the extremely wet threshold during 2014-2015. Moreover, the Nan River basin exhibits a significant likelihood of recurrence for events with return intervals of under three years. Likewise, the Wang River basin, specifically at station Y.1C in Mueang Phrae District, exhibits a significant likelihood of recurrence for events with return durations of less than three years. The future utilization of this runoff variability index should encompass other basins within the Upper Northern Region to assess both return periods and severity levels of runoff episodes. This will enable entities engaged in water resource management in Northern Thailand to utilize this information for enhanced water management practices.

Author contributions

Tanachot Chaito: Accountable for data acquisition and analysis, interpretation of findings, composition of the research report; Manad Khamkong: Directed the research design, supervised the study, offered recommendations based on the findings, performed the review and editing of the research report. All authors have examined and sanctioned the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was partially supported by Chiang Mai University.

Conflict of interest

The author declares no conflict of interest.

References

1. D. Alford, Streamflow and sediment transport from mountain watersheds of the chao phraya basin, northern Thailand: a reconnaissance study, *Mt. Res. Dev.*, **12** (1992), 257–268. https://doi.org/10.2307/3673669
2. C. Hugenschmidt, J. Ingwersen, W. Sangchan, Y. Sukvanachaikul, A. Duffner, S. Uhlenbrook, et al., A three-component hydrograph separation based on geochemical tracers in a tropical mountainous headwater catchment in northern Thailand, *Hydrol. Earth Syst. Sci.*, **18** (2014), 525–537. https://doi.org/10.5194/hess-18-525-2014
3. A. D. Ziegler, R. A. Sutherland, T. W. Giambelluca, Runoff generation and sediment production on unpaved roads, footpaths and agricultural land surfaces in northern Thailand, *Earth Surf. Proc. Land.*, **25** (2000), 519–534. https://doi.org/10.1002/(SICI)1096-9837(200005)25:5<519::AID-ESP80>3.0.CO;2-T
4. T. Piman, M. Babel, Prediction of rainfall-runoff in an ungauged basin: case study in the mountainous region of northern Thailand, *J. Hydrol. Eng.*, **18** (2013), 285–296. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000573
5. A. F. A. Tinumbang, K. Yorozu, Y. Tachikawa, Y. Ichikawa, H. Sasaki, T. Nakaegawa, Analysis of runoff characteristics generated by land surface models and their impacts on river discharge, *J. Jpn. Soc. Civil Eng. Ser. B1 (Hydraul. Eng.)*, **75** (2019), 271–276. https://doi.org/10.2208/jscejhe.75.2_I_271
6. M. Khamkong, P. Bookkamana, Development of statistical models for maximum daily rainfall in upper northern region of Thailand, *Chiang Mai J. Sci.*, **42** (2015), 1044–1053.

7. T. Chaito, M. Khamkong, P. Murnta, Appropriate transformation techniques to determine a modified standardized precipitation index for the Ping River in northern Thailand, *EnvironmentAsia*, **12** (2019), 32–42. https://doi.org/10.14456/ea.2019.43
8. T. Chaito, M. Khamkong, The length-biased Weibull Rayleigh distribution for application to hydrological data, *Lobachevskii J. Math.*, **42** (2021), 3253–3265. https://doi.org/10.1134/S1995080222010218
9. R Core Team, *R: A language and environment for statistical computing*, Vienna, Austria, 2025. Available from: https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf.
10. B. Ranneby, The maximum spacing method. An estimation method related to the maximum likelihood method, *Scand. J. Stat.*, **11** (1984), 93–112.
11. R. C. H. Cheng, N. A. K. Amin, Estimating parameters in continuous univariate distributions with a shifted origin, *J. R. Stat. Soc. Ser. B (Methodol.)*, **45** (1983), 394–403. https://doi.org/10.1111/j.2517-6161.1983.tb01268.x
12. G. Salvadori, F. Durante, C. De Michele, M. Bernardi, L. Petrella, A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities, *Water Resour. Res.*, **52** (2016), 3701–3721. https://doi.org/10.1002/2015WR017225
13. World Meteorological Organization (WMO), *Drought and agriculture*, WMO.TN138, Geneva, 1975. Available from: https://library.wmo.int/idurl/4/30669.
14. Upper Northern Region Irrigation Hydrology Center, *Rainfall and runoff data*, 2020. Available from: https://www.hydro-1.net/.
15. United Nations Department of Humanitarian Affairs (DHA), *Thailand: Floods-Oct 1995*, ReliefWeb, 1995. Available from: https://reliefweb.int/taxonomy/term/5389.