Application of the Length–Biased Weibull–Rayleigh Distribution to Fit the Rainy Season Rainfall for the Upper Ping River in Northern Thailand

Tanachot Chaito, Nawapon Nakharutai and Manad Khamkong

Tanachot Chaito

Department of Statistics, Faculty of Science,

Chiang Mai University, Chiang Mai, 50200 Thailand

4 - 5 AUGUST 2021

CONTENTS

INTRODUCTION

- The world is facing climate change, including floods and drought.
- The flood could give damage life, habitat and economy, and they are expected to become more severe in the future.
- In the 21st century, heavy rainfall will occur more frequently in many areas of the world cause the increased risk of flooding that contributes to damaging infrastructure and the economy (Intergovernmental Panel on Climate Change, 2012).
- In Thailand, there were several extreme floods occurred. For example, the northern and central regions had heavily flooded in 2011, which resulted in damage of agricultural, industry and economy sectors (Hydro - Informatics Institute, 2016).

INTRODUCTION

Figure 1: Map of the Chao Phraya River drainage basin showing the Ping River

- The Upper Ping River basin consists of Chiang Mai and Lamphun provinces.
- Chiang Mai and Lamphun provinces are still experiencing continuous flooding, which severely affects the agriculture and industry in the areas.
- An approach that will prevent or reduce the severity of flood is to monitor the areas using flood irrigation (Paladkong et al., 2019) or to introduce suitable models for determining the return levels of the highest rainfall during the return period (Chaleeraktrakoon, 2008).

INTRODUCTION

- Rainfall data is often right–skewed and in some situations, outlier or extreme values can occur.
- Common statistical distributions that have been applied to model rainfall data are Gumbel, Weibull, gamma, lognormal, Pearson type III and Frechet distributions, among others (Cordeiro et al., 2019).

INTRODUCTION

- Several authors (Yusof and Hui-Mean, 2012; Hussain et al., 2019) claimed that the Weibull distribution could be the best choice for fitting rainfall.
- Weibull distribution has been developed in order to fit hydrological data. For example, Ganji et al. (2016) developed the Weibull–Rayleigh distribution, which is mixed distribution, and suggested that it could provide a better fitting for flood data compared with the beta–Pareto, Weibull, and Pareto distributions.

INTRODUCTION

- Hydrological data is categorized as environment data that are usually non–random and non–replicated which can lead to bias recorded observations (Nanuwong, 2015).
- A weighted distribution is a common method using when the probabilities of observations recorded from a random process are not equal. The weighted distribution was first proposed by Fisher (1934) and further extended by Rao (1965) as length–biased distribution.

INTRODUCTION

- In 2020, Chaito and Khamkong presented the length–biased Weibull–Rayleigh (LBWR) distribution, which modified the Weibull–Rayleigh distribution using the length–biased distribution.
- The LBWR distribution could provide more efficiency of fitting to flood datasets than the Rayleigh, Weibull, Pareto, and Weibull–Rayleigh distributions.
- Therefore, The LBWR distribution might be potential to apply the LBWR distribution to fit rainfall data.

INTRODUCTION

The probability density function (pdf) and cumulative distribution function (cdf) of the LBWR distribution are given by

$$f_L(x) = \frac{\alpha x^2}{\beta\delta^2\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)}\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha-1}\exp\left[-\left(\frac{x^2}{2\beta\delta^2}\right)^{\alpha}\right], \quad x>0, \;\; \alpha,\beta,\delta>0, \qquad (1)$$

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

where α is a shape parameter, β and δ are scale parameters, $\Gamma(\alpha)=\int_0^{\infty} u^{\alpha-1}e^{-u}du$ is a gamma function and $\gamma(\alpha,x)=\int_0^{x} u^{\alpha-1}e^{-u}du$ is the lower incomplete gamma function.

OBJECTIVES

- To apply the length–biased Weibull–Rayleigh (LBWR) distribution for fitting the rainy season rainfall data.

- To predict the return levels of the rainy season rainfall data.

METHODOLOGY

Study area

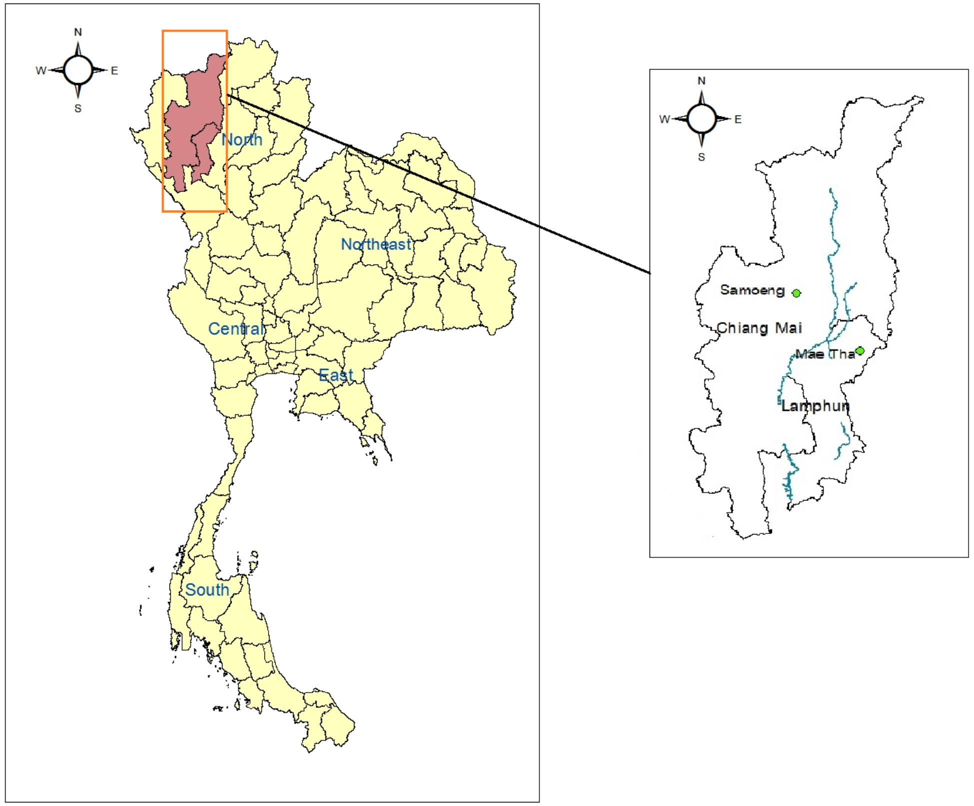


Figure 2: Location of two stations on upper Ping River in northern Thailand

Samoeng District, Chiang Mai Province, and Mae Tha District, Lamphun Province, were chosen at study cases, since floods were frequently occurred in both areas.

METHODOLOGY

Data collection

Monthly rainfall data of Samoeng and Mae Tha stations from January 1957 to December 2020
(Hydrology and Water Management Center for the Upper Northern Region of Thailand)

- **Summer season** (February to May)
- **Rainy season** (June to September)
- **Winter season** (October to January of the following year)

This study handled the missing monthly rainfall data by replacing it with the average of the past five years of such month.

METHODOLOGY

Maximum likelihood estimation

Let $X_1, X_2, \dots, X_n$ be a random sample from the LBWR distribution with parameter vector $\Theta = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta})$, $x_1, x_2, \dots, x_n$ be the sample values. The likelihood and log–likelihood functions are given by

$$L(\Theta) = \prod_{i=1}^{n} \left\{ \frac{\alpha x_i^2}{\beta\delta^2\sqrt{2\beta\delta^2}\,\Gamma\left(1+\frac{1}{2\alpha}\right)} \left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha-1} \exp\left[-\left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha}\right] \right\}, \tag{3}$$

$$\log L(\Theta) = \sum_{i=1}^{n} \left\{ \log\alpha + \log x_i^2 - \log\beta - 2\log\delta - \frac{1}{2}\log(2\beta\delta^2) - \log\Gamma\left(1+\frac{1}{2\alpha}\right) + (\alpha-1)\log\left(\frac{x_i^2}{2\beta\delta^2}\right) - \left(\frac{x_i^2}{2\beta\delta^2}\right)^{\alpha} \right\}. \tag{4}$$

To obtain the maximum likelihood estimation, this study uses <u>mle</u> function in <u>stats4</u> package in the R statistical software (R Core Team, 2020).

METHODOLOGY

Model selection criteria

Selecting the appropriate distribution for the rainy season rainfall data, this study uses Kolmogorov–Smirnov test, Anderson–Darling test (Anderson and Darling, 1952) and Akaike information criterion (Akaike, 1973).

- The Kolmogorov–Smirnov (KS) test is defined as

$$\mathrm{KS} = \sup_{x}[G_0(x) - G(x)]\,, \tag{5}$$

where $G_0(x)$ is empirical distribution function of the observed data and $G(x)$ is the cdf of the hypothesized distribution.

METHODOLOGY

Model selection criteria

- The Anderson–Darling (AD) test is given by

$$\mathrm{AD} = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln G(x_i) + \ln\left(1 - G(x_{n-i+1})\right)\right], \tag{6}$$

where $G(x)$ is the cdf of the hypothesized distribution, n is the sample size and x_i are the ordered data.

- The Akaike information criterion (AIC) is written as

$$\mathrm{AIC} = 2k - 2\log L\left(\widehat{\Theta}\right), \tag{7}$$

where k is the number of parameters and $L\left(\widehat{\Theta}\right)$ is the maximized value of the likelihood function. The best fit distribution for the rainy season rainfall data can be selected from minimum of KS test, AD test and AIC values.

METHODOLOGY

Return level

The T−year return level of the LBWR distribution can be calculated as follows:

$$x_T = \sqrt{2\hat{\beta}\hat{\delta}^2 A^{\left(\frac{1}{\hat{\alpha}}\right)}}, \tag{8}$$

where $A = \gamma^{-1}\left[1 + \frac{1}{2\hat{\alpha}}, \Gamma\left(1 + \frac{1}{2\hat{\alpha}}\right)\left(\frac{1}{T}\right)\right]$, when γ^{-1} is inverted of the lower incomplete gamma function, Γ is the gamma function, T is return period, $\hat{\alpha}$ is a shape parameter, and $\hat{\beta}$ and $\hat{\delta}$ are scale parameters, which were estimated via maximum likelihood estimation method.

METHODOLOGY

Return level

The profile likelihood of $1-\omega$ confidence interval of return levels (x_T) for the LBWR distribution can be written as

$$\left\{x_T: 2\left[\log L(\alpha, \beta, \delta) - \max_{\alpha,\delta} \log L(x_T, \alpha, \delta)\right] \leq C_{1-\omega}\right\}, \tag{9}$$

where $C_{1-\omega}$, is $1-\omega$ quantile of the chi–square distribution with on degree of freedom and ω is the significance level.

RESULTS AND DISCUSSIONS

Descriptive statistics of the rainy season rainfall data

Table 1: Descriptive statistics of the rainy season rainfall data for Samoeng and Mae Tha stations.

Station	Rainfall (mm)								
	Min.	Max.	Q_1	Q_2	Q_3	Mean	SD	Skewness	Kurtosis
Samoeng	272.40	1920.40	608.10	747.60	896.60	773.60	266.4708	1.5241	7.7506
Mae Tha	381.50	1933.60	567.00	674.90	800.00	698.20	221.2011	2.6457	16.2500

mm denotes millimeter; Q_i denotes the i^{th} quartile of data and SD denotes the standard deviation.

RESULTS AND DISCUSSIONS

Descriptive statistics of the rainy season rainfall data

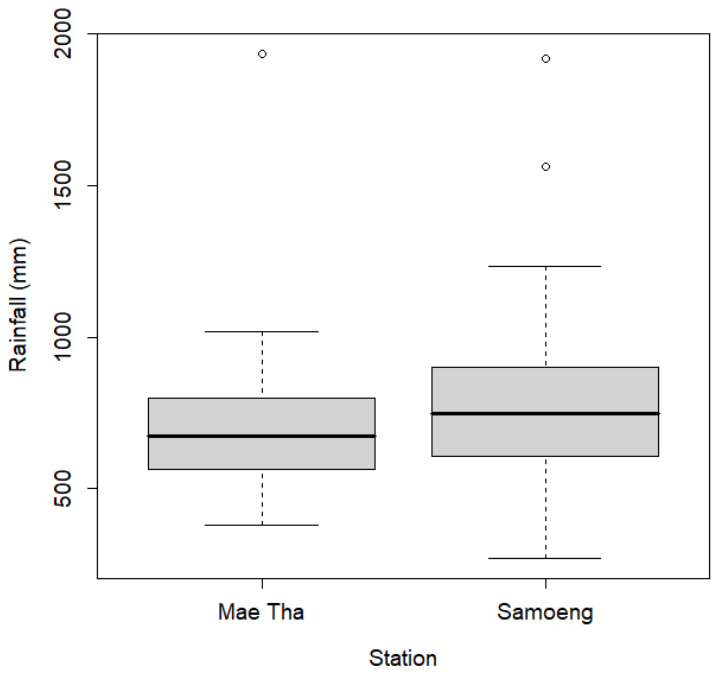


Figure 3: Boxplot of the rainy season rainfall data for Samoeng and Mae Tha stations.

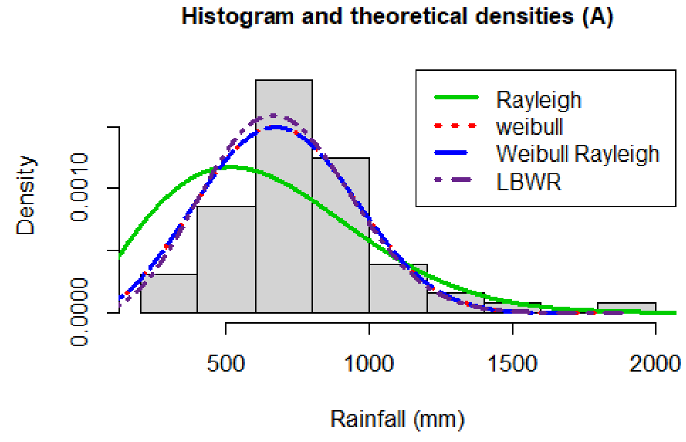
- ➢ Samoeng station has outlier in 1963 (1562.50 mm.) and 1994 (1920.40 mm.).
- ➢ Mae Tha station has outlier in 1958 (1933.60 mm.).

RESULTS AND DISCUSSIONS

Model selection criteria of rainy season rainfall data in Samoeng station

Table 2: Summary of selected distributions using the KS test, AD test and AIC for the rainy season rainfall data in Samoeng station.

Distribution	KS	AD	AIC
Rayleigh	0.2392	5.3347	913.6498
Weibull	0.1203	1.7352	900.4815
Weibull- Rayleigh	0.1206	1.7364	902.4815
LBWR	**0.1098**	**1.3798**	**898.6661**



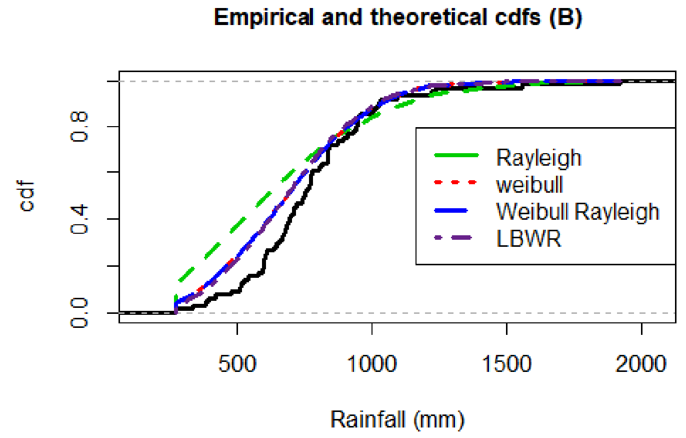


Figure 4: Histograms and theoretical densities (A) and empirical and theoretical cdfs (B) for the rainy season rainfall data in Samoeng station.

RESULTS AND DISCUSSIONS

Model selection criteria of rainy season rainfall data in Mae Tha station

Table 3: Summary of selected distributions using the KS test, AD test and AIC for the rainy season rainfall data in Mae Tha station.

Distribution	KS	AD	AIC
Rayleigh	0.2470	7.3541	896.6363
Weibull	0.1402	2.8046	880.8026
Weibull- Rayleigh	0.1403	2.8083	882.8026
LBWR	**0.1284**	**2.2462**	**876.8741**

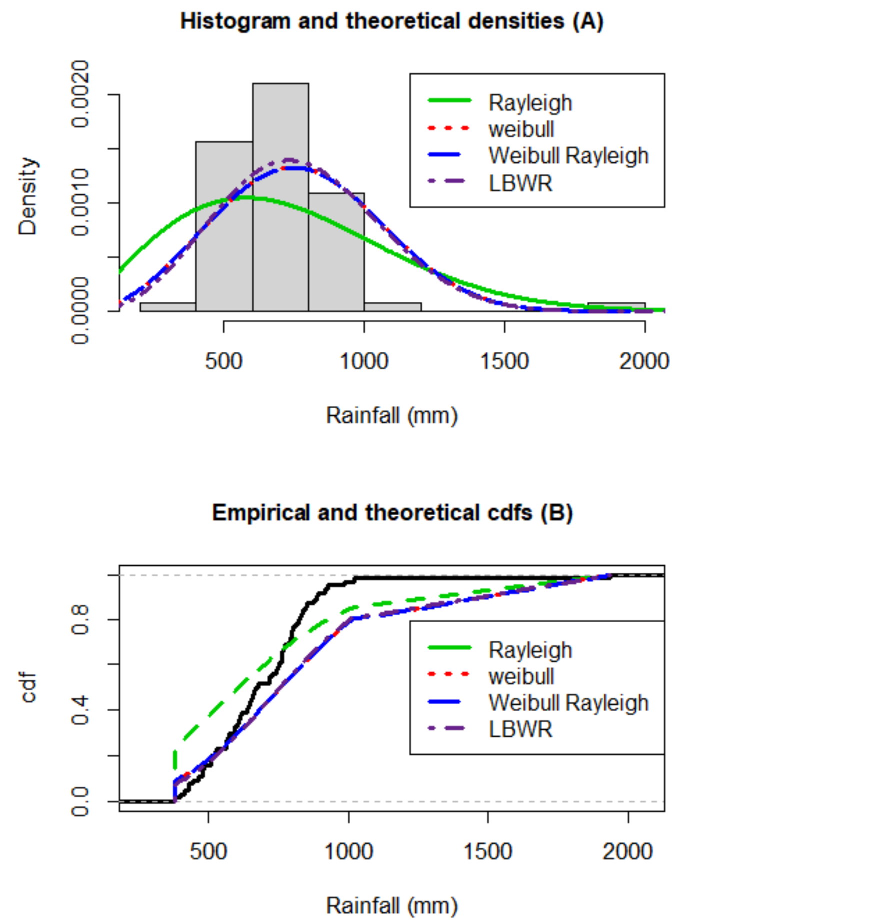


Figure 5: Histograms and theoretical densities (A) and empirical and theoretical cdfs (B) for the rainy season rainfall data in Mae Tha station.

RESULTS AND DISCUSSIONS

Estimating parameters of the LBWR distribution

Table 4: Estimated parameters of the LBWR distribution for the rainy season rainfall data in Samoeng and Mae Tha stations.

Station	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$
Samoeng	1.2298	151.8907	41.9395
Mae Tha	1.2713	147.1836	38.6006

RESULTS AND DISCUSSIONS

Return levels of rainy season rainfall data

Table 5: The estimates of return levels and 95% confidence intervals of return levels based on the profile likelihood method for the rainy season rainfall data in Samoeng and Mae Tha stations.

Return period	Samoeng			Mae Tha		
	Lower confidence limit	Return level	Upper confidence limit	Lower confidence limit	Return level	Upper confidence limit
1	246.25	331.01	410.97	233.75	304.56	371.08
2	669.55	757.28	839.31	608.43	682.07	752.84
3	798.78	882.88	966.91	720.19	791.86	865.03
4	872.10	955.42	1042.97	783.41	855.04	931.32
5	921.92	1005.62	1096.91	826.38	898.67	978.05
10	1049.95	1139.08	1246.29	937.16	1014.33	1106.27
15	1111.82	1206.20	1324.75	990.94	1072.33	1173.01
20	1151.61	1250.28	1377.45	1025.61	1110.36	1217.64
30	1203.08	1308.37	1448.21	1070.56	1160.41	1277.37
40	1236.91	1347.15	1496.24	1100.14	1193.78	1317.79
50	1261.83	1376.02	1532.40	1121.95	1218.60	1348.16

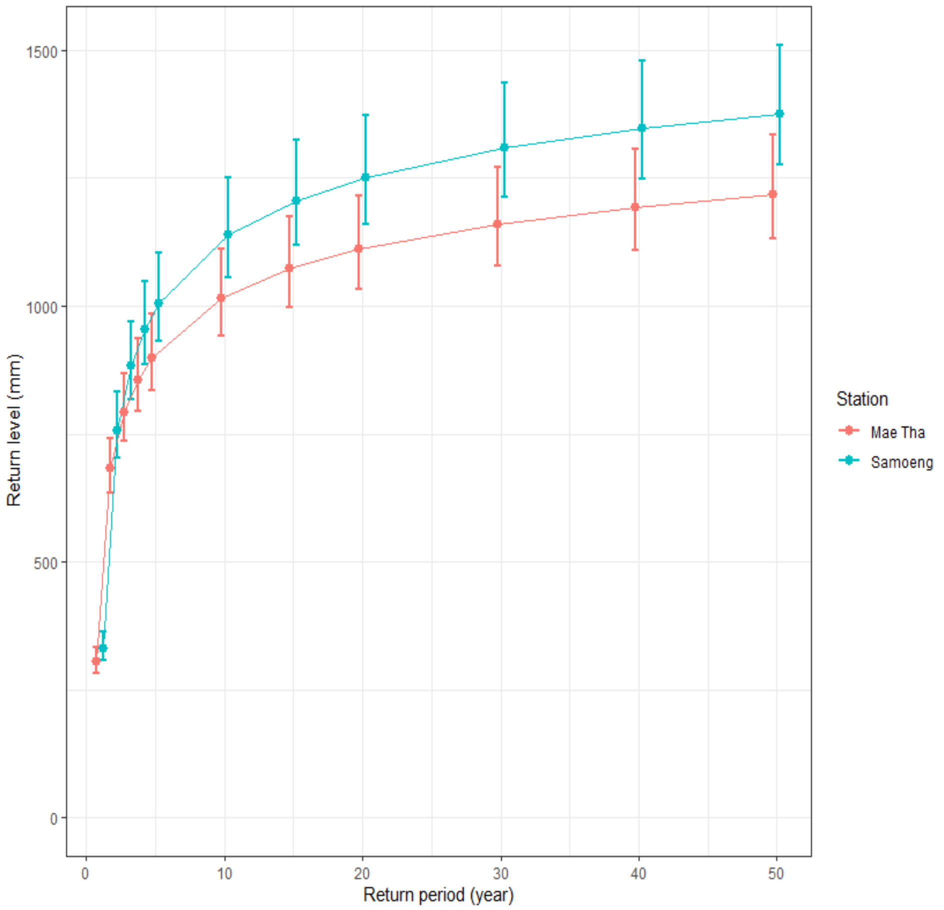


Figure 6: Comparison of return levels and 95% confidence intervals of return levels for the rainy season rainfall data in Samoeng and Mae Tha stations.

CONCLUSIONS

- ❖ Statistical distributions are important and useful to determine the appropriate distributions for rainfall data and predict the return levels of rainfall data.
- ❖The return levels can be useful for flood irrigation planning and water management.
- ❖ Comparing the LBWR distribution with Rayleigh, Weibull and Weibull–Rayleigh distributions showed that the LBWR distribution is outperformed and suitable in fitting the rainy season rainfall data.

CONCLUSIONS

- Using the LBWR distribution to predict the return levels of the rainy season rainfall data showed that the levels at Samoeng station might be higher than the levels at Mae Tha station. This can imply that Samoeng district may have a higher risk of flooding compared to Mae Tha district.
- The LBWR distribution can be another choice for modelling the rainy season rainfall data.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Proceedings of 2nd International Symposium on Information Theory, Budapest, Hungary.

Anderson, T. and Darling, D. (1952). Asymptotic theory of certain “goodness-of-fit” criteria based on stochastic processes. Annals of Mathematical Statistics, 23:193–212.

Chaito, T. and Khamkong, M. (in press). The length-biased Weibull-Rayleigh distribution for application to hydrological data. Lobachevskii Journal of Mathematics.

Chaleeraktrakoon, C. (2008). Analysis of the flood frequency. Thammasat Printing house, Thammasat University, Bangkok, Thailand.

Cordeiro, G., Mansoor, M., and Provost, S. (2019). The Harris extended Lindley distribution for modelinghydrological data. Chilean Journal of Statistics, 10:77–94.

Fisher, R. (1934). The effect of methods of ascertainment upon the estimation of frequencies. Annals of Eugenics, 6:13–25.

REFERENCES

Ganji, M., Bevrani, H., Hami, G., and Zabihi, S. (2016). The weibull–rayleigh distribution, some properties, and applications. Journal of Mathematical Sciences, 218:269–277.

Hussain, T., Bakouch, H., and Chesneau, C. (2019). A new probability model with application to heavy-tailed hydrological data. Environmental and Ecological Statistics, 26:127–151.

Hydro - Informatics Institute (2016). Flood event record. https://www.thaiwater.net/web/index.php/archive.Html. Accessed: 2021-02-05.

Hydrology and Water Management Center for the Upper Northern Region of Thailand (2020). Rainfall data. https://www.hydro-1.net/. Accessed: 2021-02-01.

Intergovernmental Panel on Climate Change (2012). Managing the risks of extreme events and disasters to advance climate change adaptation. Cambridge University Press, Cambridge, Cambridge University.

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

REFERENCES

Paladkong, T., Nanthaprut, P., and Chaisee, K. (2019). Extreme value analysis and return estimation of rainfall data in Bangkok metropolitan region. In Proceedings of the National Undergraduate Conference on Statistics, Chiang Mai University, Chiang Mai, Thailand.

R Core Team (2020). R: Language and environment for statistical computing. https://www. R-project.org/. Accessed: 2021-03-12.

Rao, C. (1965). On discrete distributions arising out of methods of ascertainment. Sankhy¯a: TheIndian Journal of Statistics, Series A, 27:311–324.

Yusof, F. and Hui-Mean, F. (2012). Use of statistical distribution for drought analysis. Applied Mathematical Sciences, 6:1031–1051.

THANK YOU

INTERNATIONAL CONFERENCE ON COMPUTING,
MATHEMATICS AND STATISTICS