International Journal of Mathematics Trends and Technology
ISSN: 2231-5373/ https://doi.org/10.14445/22315373/IJMTT-V69I1P512
Volume 69 Issue 1, 79-85, January 2023

Original Article

The Short Scale (Hourly) Rainfall Modelling For Intensity Based On Strom Analysis Events (SEA) Using Some Probability Distributions

Lisa Rahayu[1], Rado Yendra[2], Muhammad Marizal[3], Ari Pani Desvina[4], Rahmadeni[5]

[1,2,3,4,5] *Department of Mathematics, Faculty of Science and Technology, Universitas Islan Negeri Sultan Syarif Kasim Riau, Indonesia*

Received: 07 December 2022 Revised: 09 January 2023 Accepted: 20 January 2023 Published: 31 January 2023

Abstract - *Storm amount (SA), storm intensity (SI), and storm duration (SD) are three external measurements that can be used to describe short-scale rainfall modeling based on storm process. This study displays the distribution that fits the SI series the best, which is based on hourly rainfall data from 1970 to 2008 at the Alor Setar station in Peninsular Malaysia. Gamma, Weibull, and Log Normal distributions with two parameters are taken into consideration. The Bayesian Maximum Likelihood (MLE) approach is used to determine these distributions' parameters. Then it is determined how well theoretical data and model distributions fit one another (GOF). The outcome demonstrates that the stations discovered that the MLE method may provide the best SI modeling, specifically for Log-Normal distribution. We can reliably forecast the future risks associated with the SI based on the stated model.*

Keywords - *Storm, Storm Amount, Storm Intensity, Storm Duration, MLE.*

1. Introduction

Statistical modeling is very important to overcome the problems that arise due to climate change issuesThe construction of structures like seawalls, bridges, and buildings might benefit from knowing the probabilities of certain natural phenomenon events, such as changes in sea level and wind speed. The data can be used to evaluate the threat posed by other phenomena, such as precipitation and pollution. Climate change is frequently linked to episodic rainfall events like hourly showers, which can be followed by a string of natural disasters like flash floods and landslides. Decision-makers can use scale rainfall data analysis to plan strategies to lessen or mitigate the effects of disasters by taking this occurrence into consideration. The issue of rain on a short scale becomes very important when it is connected with the occurrence of flooding in big cities, rain that occurs for 2 hours continuously has been able to flood some areas in big cities. Statistical modeling for short-scale rain requires a method that can extract short-scale rain data into important variables, namely amount, duration, and intensity. This method is known as the storm event analysis (SEA) theory [1-3]. Earlier research offers a few ways to view storm event analysis (SEA) in their analysis, including [4-12]. The research is dominated by the determination of the best probability distribution and the relationship between the marginal distribution of each variable associated with the copula model, A bivariate exponential [13], a bivariate gamma [14], a bivariate lognormal [15], or a bivariate extreme value distribution [16] are examples of probability distributions that are expected to be either normal [17] or to have the same type as the marginals. Storm rainfall is actually a complex event, and its marginal distributions aren't always comparable or distributed as usual. Other distributions should be taken into account since they might result in more accurate rainfall forecasts. With the copula approach, multivariate problems can be solved with a wider range of marginal distributions and dependent structures [18]. Over the past ten years, various copula forms have been applied to hydrology. Among them are the Archimedean copulas [19 – 22], the Farlie-Gumbel-Morgenstern (FGM) copula [23], and elliptic copulas like the Gaussian copula [24]. This study's goals are to: (1) use the SEA approach to derive short scale rainfall variables, (2) determine some modeling for the storm intensity (SI) from some probability models (3) test the model distributions using the Good of Fit Test for the short scale rainfall data from the rain gauge station Alor Setar, Malaysia The several methods for Good of Fit test in determining the best model, such as the graphical method and the numerical method, will be carried out in this study. The graphical method will be carried out by comparing the pdf, CDF, and QDF (the inverse of CDF, known as quantile function) of each model while the numerical values of AIC and BIC will complete this research as a representation of the numerical method. Model selection using the graphical method often gives the same results for each model but using the numerical method will result in the selection of the best model. easy based on the lowest value generated by AIC and BIC

2. Data and Definition of Storm

The Department of Drainage and Irrigation provided the data, which consists of hourly rainfall data from rain gauge sites in Peninsular Malaysia between 1970 and 2008. First, the SEA definition must be used to get the hourly rain data from Scala. The definition of the intervention period has a significant impact on how a SEA is defined. Inter-event time definition (IETD) is the shortest amount of time that must pass between two successive storm events. The dry interval between two separate storm episodes must therefore be at least as long as the IETD value. The serial correlation between two separate storms is minimized by the selection of the IETD values [25]. Because rainfall time concentrations of less than 6 hours will allow the runoff response from subsequent storms to look independent, the IETD for small urban watersheds is typically taken to be 6 hours [12]. A storm's depth is measured by the total amount of rain it has produced, with at least one wet hour both before and after any dry intervals of less than six hours or none at all. Storm depth is the ratio of storm depth to storm duration, while storm intensity is the time span during which a storm occurs. SA, SD, and SI are the data derived from the rainfall data. Figure 1 explains details and the description of this storm. Let n_i be the duration hour for the ith storm, and let X_{ij} be j-th rainfall *mm*.

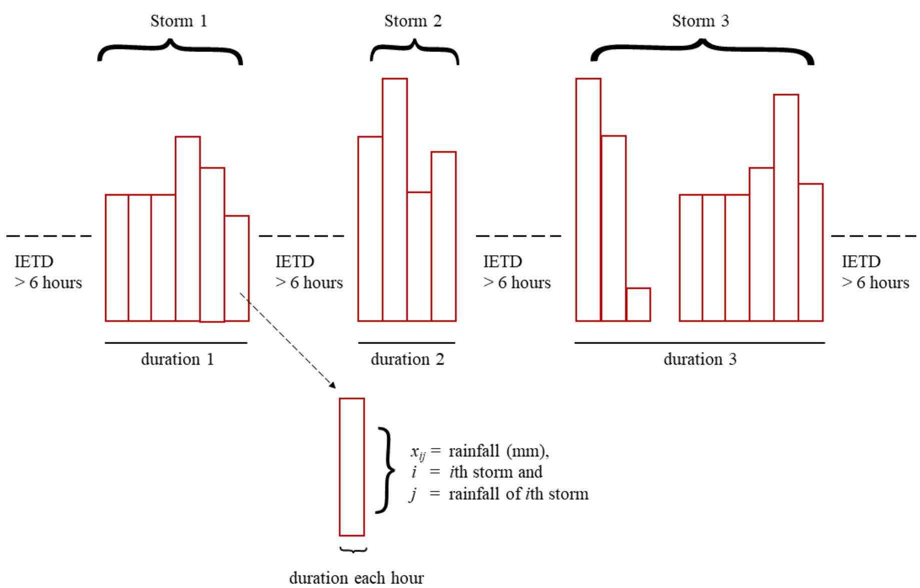

Fig. 1 Definition of storm

The data that has been extracted based on SEA will be broken down into storm intensity rate data for each storm that is generated as shown in Figure 2. Initial information about SI in the form of statistical information can be described in table 1. The values shown are descriptive statistics that generally describe the SI data obtained, such as the average value, variance, kurtosis, skewness, and maximum and minimum data. From table 1 it can be seen that the skewness value is very small and the kurtosis value is small than 1, this can mean that unequal opportunity models such as Weibull, Gamma, and Log Normal can be used in modeling SI data in this research. This is also clarified by the data histogram shown in Figure 3, in this figure it can be seen that the three probability models used in this study have the right reasons for their use.

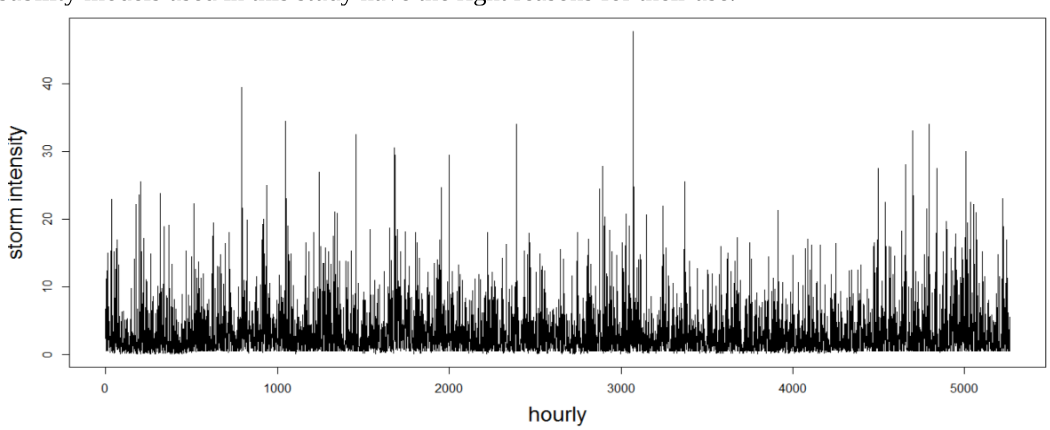

Fig. 2 The Storm Intensity short scale rainfall (hourly)

Table 1. Desciptive Statistics of storm intensity data

mean	var	skewness	kurtosis	minimum	maximum
3.28	15.85	5.37×10^{-7}	0.08	0.1	47.75

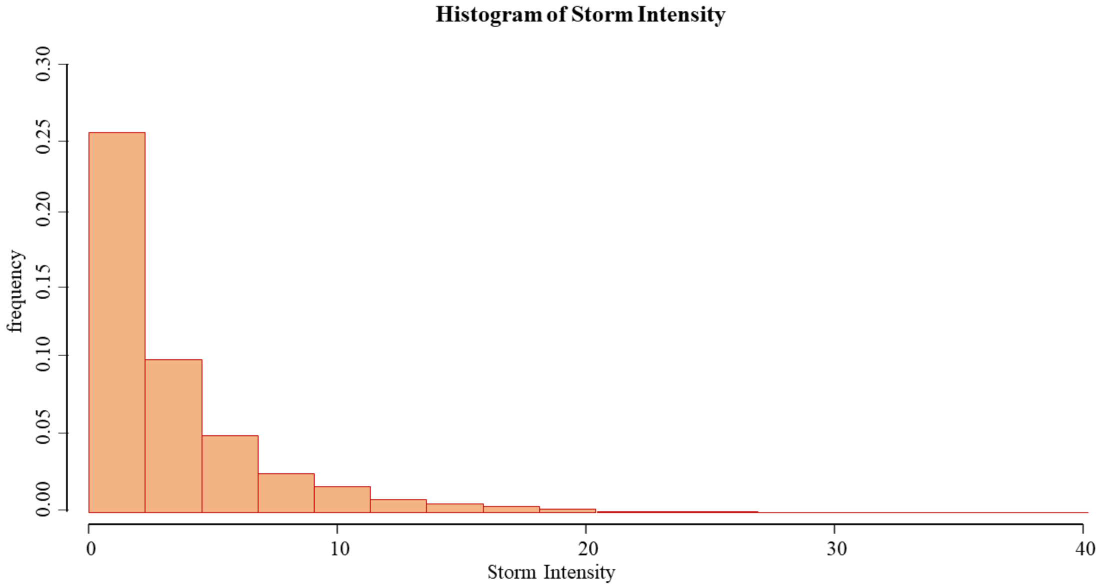

Fig. 3 The Histogram of Storm Intensity (SI)

3. Methods

3.1. Density Function and Cumulative Function

The Storm Intensity (SI) modeling requires analysis of the SI that comes to form the extraction of short scale rainfall data or hourly rainfall dataProbability density functions are the main instruments for describing the SI features. In this study, three two-parameter probability density functions like Weibull, Gamma, and Log Normal will be used. Table 2 shows the pdf and CDF for each distribution we take into account, where y stands for the observed values of the random variable used to represent the event of interest.

Table 2. Pdf and Cdf distributions model

Distribution	**PDF (*f(y)*) and CDF (*F(y)*)**
Weibull (x;η,κ)	$f(x) = \frac{\eta}{\kappa}\left(\frac{x}{\kappa}\right)^{\eta-1} e^{\left(-\left(\frac{x}{\kappa}\right)^{\eta}\right)}, x > 0,\ \ \eta,\kappa > 0$ $F(x) = 1 - e^{\left(-\left(\frac{x}{\kappa}\right)^{\eta}\right)}$
Gamma (x;α,β)	$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x > 0, \quad \alpha,\beta > 0$ $F(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_{o}^{x} x^{\alpha-1} e^{-\frac{x}{\beta}} dx$
Log Normal (x;μ,σ)	$f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{\frac{-(loglog\ x-\mu)}{2\sigma^2}}, x > 0, \quad \mu,\sigma > 0$ $F(x) = \Phi\left(\frac{log\ log\ x - \mu}{\sigma}\right)$ where Φ is the cumulative distribution function of the standard normal distribution (i.e., $N(0,1)$).

Parameter estimation is the first thing that must be done in probability modeling. Like most studies that have been done, it can be concluded that the maximum likelihood (MLE) method is the most dominant used in this case. Since the MLE function for this model is implicit and complicated, we won't go into great depth about it in this work.

Table 2. Computed parameters of different distribution

	η	κ	α	β	μ	σ
Weibull	0.96	3.21				
Gamma			1.02	3.22		
Log Normal					0.62	1.08

A numerical approach, termed Newton's Rhapshon, is needed to solve the nonlinear equation caused by the maximum log-likelihood function (ln L). However, iteration systems have employed this technique to identify solutions. For this method, a number of beginning settings have been tested. The chosen estimation parameter can be thought of as the value that is iterated if the initial value utilized does so or if the iterations converge to that value. In this study, the MLE method is also the main choice to be used in generating parameter estimates. After the parameter estimation is obtained, the next test in selecting the best model using the graphical method and numerical method will be carried out.

3.2. Maximum Likelihood and Goodness of Fit Tests

Suppose $(y_1, y_2, \cdots, y_n)$ is a randomly chosen sample from four PDFs, Table 3 displays the natural log probability (ln L). The following nonlinear problem is solved by the MLE $\hat{\theta}\ of\ \theta$ since it is the answer to the equation $\frac{d \ln L}{d\theta} = 0$. The best distribution is determined utilizing the findings from various goodness-of-fit tests. The GOF tests taken into consideration are based on numerical criteria and the graphical inspection probability density function (PDF). To ascertain the distributions' goodness-of-fit standards, Akaike's information criterion (AIC) and Bayesian information criterion (BIC) were used. The majority of the time, graphical inspection produced the same outcome, however the AIC and BIC findings varied. The distribution with the lowest AIC and BIC values was determined to be the best fit outcome. The following formulas are used to calculate AIC and BIC:

$$\text{AIC} = -2\ log\ L + 2P;$$

$$\text{BIC} = -2\ log\ L + P\ log\ i\ ,$$

where P = number of parameters, i = number of samples

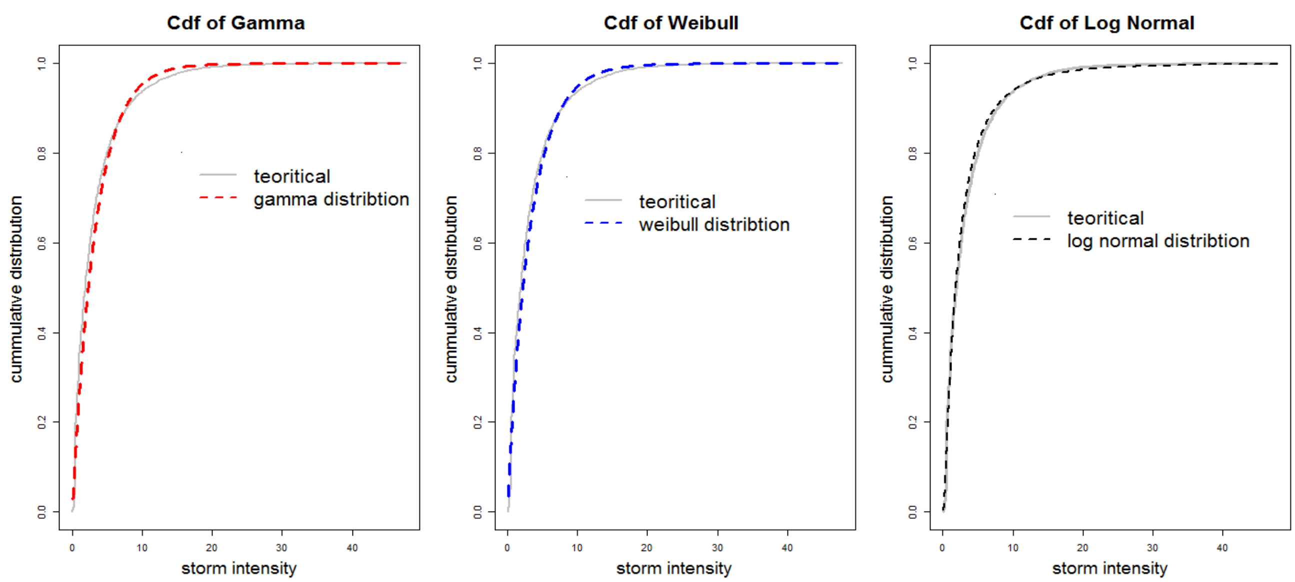

Fig 4. Cdf plot for comparisons Predicted and observed for SI data

4. Results

In this study, the parameters of the three probability distributions will be estimated using the MLE method as presented in table 2. Based on these parameters, pdf plots for all distributions in this paper will be created and will be used to test the goodness of fit of the model by looking at the ability of the plots to approach the SI histogram data. as depicted in Figure 4. In this figure, it can be seen that the models used have different capabilities in approaching the histogram. Figure 5 shows that the two parameters distribution such as Weibull, Gamma, and Log Normal are able to approach the histogram or the frequency of SI data that occurs in Malaysia.

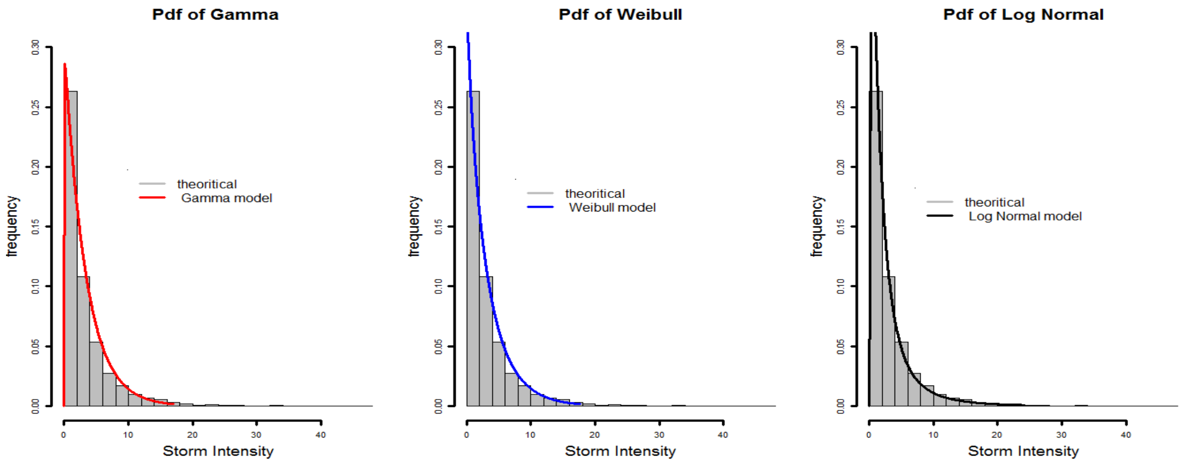

Fig. 5 Histogram Probability density fucntion

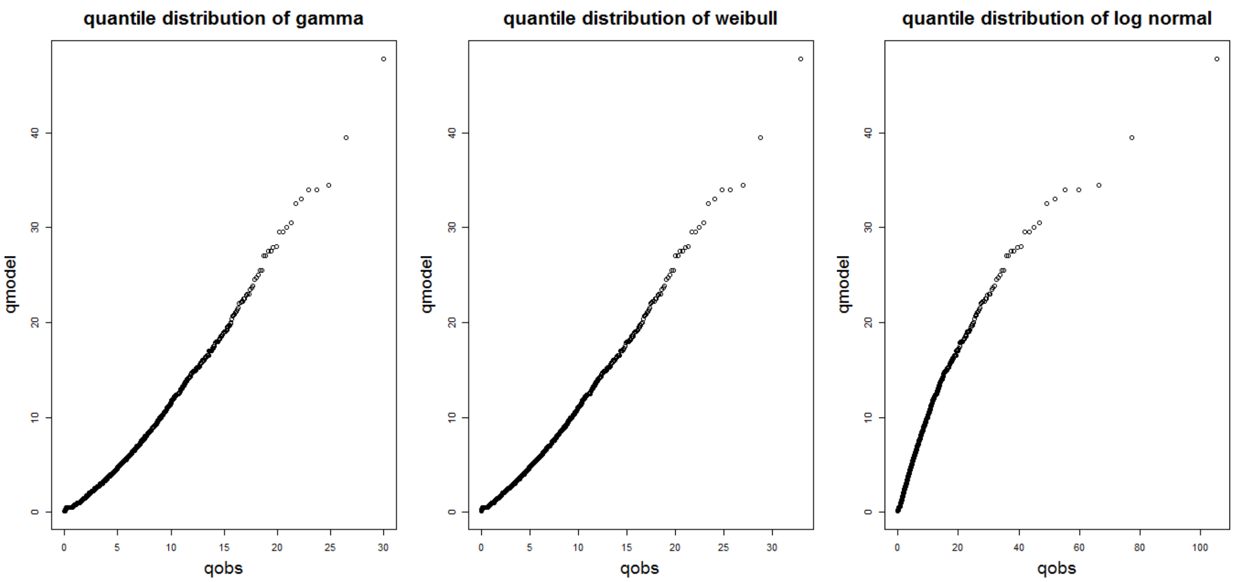

Fig 6. Qdf plot for comparisons Predicted and observed for SI data

Results that are not much different in testing the goodness of fit of the models are also shown by the graph of the inverse cumulative distribution or better known as the GQDF (quantile) plot as shown in Figure 6. From the goodness of fit test of the model using this graphical method, it is very clear that the pdf, CDF, and QDF plots give very clear results that the two-parameter distributions such as Weibull, Gamma, and Log Normal are the best model in analyzing the frequency of SI data. Numerical methods such as AIC, BIC, and ln (L) values for the goodness-of-fit test were also used in this study. These three values for each distribution used will be presented in table 3. Based on the values in the table, it can be concluded that the two-parameter distribution namely Log Normal distribution is the best model because it has the smallest AIC and BIC values. Table 3 is also equipped with tests of model goodness such as log-likelihood (Log L) mode, from the values presented it can also be concluded that the distribution of the Log-Normal two parameters is the best in this study. An increase in parameter values would affect the increase in the mean and maximum amount of rainfall for each year. This finding differs from studies [26, 27] for covid data with a comparison of the same distribution, finding that the correct distribution is the two-parameter Weibul distribution. it is possible that this occurs due to differences in the characteristics of the data.

Table 3. The GOF test result

	Weibull	**Gamma**	**Log Normal**
AIC	23019.25	23037.48	22311.46
BIC	23032.39	23050.62	22324.6
Log (L)	-11507.63	-11516.74	-11153.73

5. Conclusion

This study examined the likelihood that storm intensity (SI) events would occur at Peninsular Malaysian rain gauge sites. The Gamma, Weibull, and Log-Normal probability distributions were chosen to best suit the data. MLE performed an analysis of the several types of data included in this study, focusing on estimating the three probability distributions' parameters. The MLE performed well in this study, as could be shown in this publication. This study is focused to analyze the frequency of SI data caused short scale (hourly) rainfall, to identify the appropriate three models or distributions that can be used to describe the distribution of the SI data. It is concluded that the Log Normal two parameters distribution returned better results when compared with other well-known distributions. This conclusion is based on widely used goodness of fit test models such as AIC, BIC, and Log L. The graphical technique namely pdf, CDF, and QDF plots were also observed comparing the empirical distributions with the adjusted Log Normal two parameters distribution. In addition, through the best model in this study, we can use the distribution of the quantile function to simulate the SI data for the future.

References

[1] Barry J. Adams et al., “Meteorological Data Analysis for Drainage System Design,” *Journal of Environmental Engineering*, vol. 112, no. 5, pp. 827-848, 1986. *Crossref,* https://doi.org/10.1061/(ASCE)0733-9372(1986)112:5(827)

[2] Barry J. Adams, and Fabian Papa, *Urban Stormwater Management with Analytical Probabilistic Model*, Jhon Wiley and Sons Inc., Toronto, Ontario, 2000.

[3] William H. Asquith et al., *Statistical Characteristics of Storm Interevent Time, Depth, and Duration for Eastern New Mexico, Oklahoma, and Texas,* U.S. Geological Survey Publication, 2006. *Crossref,* https://doi.org/10.3133/pp1725

[4] P.S. Eagleson, “Dynamics of Flood Frequency,” *Water Resources Research*, vol. 8, no. 4, pp. 878-897, 1972. *Crossref,* https://doi.org/10.1029/WR008i004p00878

[5] Yiping Guo, and Barry J. Adams, “Hydrologic Analysis of Urban Catchments with Event-based Probabilistic Models: 2. Peak Discharge Rate,” *Water Resources Research*, vol. 34, no. 12, pp. 3433-3443, 1998. *Crossref*, https://doi.org/10.1029/98WR02448

[6] Yiping Guo, and Barry J. Adams, “Hydrologic Analysis of Urban Catchments with Event-based Probabilistic Models: 1. Runoff Volume,” *Water Resources Research,* vol. 34, no. 12, pp. 3421-3431, 1998. *Crossref,* https://doi.org/10.1029/98WR02449

[7] N.K. Goel et al., “A Derived Flood Frequency Distribution for Correlated Rainfall Rainfall Intensity and Duration,” *Journal of Hydrology*, vol. 228, no. 1-2, pp. 56-67, 2000. *Crossref,* https://doi.org/10.1016/S0022-1694(00)00145-1

[8] C. De Michele, and G. Salvadori, “A Generalized Pareto Intensity–duration Model of Storm Rainfall Exploiting 2-Copulas,” *Journal of Geophysical Research*, vol. 108, no. D2, 2003. *Crossref,* https://doi.org/10.1029/2002JD002534

[9] P. Rivera et al., “An Analytical Model for Hydrologic Analysis in Urban Watersheds,” *Proceedings of the 10th International Conference on Urban Drainage*, Copenhagen, Denmark, 2005.

[10] Shih-Chieh Kao, and Rao S. Govindaraju, “Probabilistic Structure of Storm Surface Runoff Considering the Dependence Between Average Intensity and Storm Duration of Rainfall Events,” *Water Resources Research*, vol. 43, no. 6, 2007. *Crossref,* https://doi.org/10.1029/2006WR005564

[11] L. Zhang, and Vijay P. Singh, “Bivariate Rainfall Frequency Distributions Using Archimedean Copulas,” *Journal of Hydrology,* vol. 332, no. 1–2, pp. 93–109, 2007. *Crossref,* https://doi.org/10.1016/j.jhydrol.2006.06.033

[12] Barry Palynchuk, and Yiping Guo, “Threshold Analysis of Rainstorm Depth and Duration Statistics at Toronto, Canada,” *Journal of Hydrology*, vol. 348, no.3–4, pp. 535–545, 2008. *Crossref,* https://doi.org/10.1016/j.jhydrol.2007.10.023

[13] A.C. Favre, A. Musy, and S. Morgenthaler, “Two-Site Modeling of Rainfall Based on the Neyman–Scott Process,” *Water Resources Research*, vol. 38, no. 12, 2002. *Crossref,* https://doi.org/10.1029/2002WR001343

[14] S. Yue, T.B.M.J. Ouarda, and B. Bobee, “A Review of Bivariate Gamma Distributions for Hydrological Application,” *Journal of Hydrology,* vol. 246, no. 1–4, pp. 1–18, 2001. *Crossref,* https://doi.org/10.1016/S0022-1694(01)00374-2

[15] Sheng Yue, “The Bivariate Lognormal Distribution for Describing Joint Statistical Properties of a Multivariate Storm Event,” *Environmetrics*, vol. 13, no. 8, pp. 811–819, 2002. *Crossref,* https://doi.org/10.1002/env.483

[16] J.T. Shiau, “Return Period of Bivariate Distributed Extreme Hydrological Events,” *Stochastic Environmental Research and Risk Assessment,* vol. 17, pp. 42–57, 2003. *Crossref,* https://doi.org/10.1007/s00477-003-0125-9

[17] Sheng Yue, “Joint Probability Distribution of Annual Maximum Storm Peaks and Amounts as Represented by Daily Rainfalls,” *Hydrological Sciences Journal*, vol. 45, no. 2, pp. 315–326, 2000. *Crossref,* https://doi.org/10.1080/02626660009492327

[18] Shih-Chieh Kao, and Rao S. Govindaraju, “Trivariate Statistical Analysis of Extreme Rainfall Events via the Plackett Family of Copulas,” *Water Resources Research,* vol. 44, no. 2, 2008. *Crossref,* https://doi.org/10.1029/2007WR006261

[19] C. De Michele, and G. Salvadori, “A Generalized Pareto Intensity–duration Model of Storm Rainfall Exploiting 2-Copulas,” *Journal of Geophysical Research,* vol. 108, no. D2, 2003. *Crossref,* https://doi.org/10.1029/2002JD002534

[20] C. De Michele et al., “Bivariate Statistical Approach to Check Adequacy of Dam Spillway,” *Journal of Hydrological Engineering,* vol. 10, no. 1, p. 50, 2005. *Crossref,* https://doi.org/10.1061/(ASCE)1084-0699(2005)10:1(50)

[21] G. Salvadori, and C. De Michele, “Frequency Analysis via Copulas: Theoretical Aspects and Applications to Hydrological Events,” *Water Resources Researach*, vol. 40, no. 12, 2004. *Crossref,* https://doi.org/10.1029/2004WR003133

[22] L. Zhang, and V.P. Singh, “Bivariate Flood Frequency Analysis Using the Copula Method,” *Journal of Hydrological Engineering,* vol. 11, no. 2, p. 150, 2006. *Crossref,* https://doi.org/10.1061/(ASCE)1084-0699(2006)11:2(150)

[23] Anne-Catherine Favre et al., “Multivariate Hydrological Frequency Analysis Using Copulas,” *Water Resources Research*, vol. 40, no. 1, 2004. *Crossref,* https://doi.org/10.1029/2003WR002456

[24] B. Renard, and M. Lang, “Use of a Gaussian Copula for Multivariate Extreme Value Analysis: Some Case Studies in Hydrology,” *Advances in Water Resources*, vol. 30, no. 4, pp. 897–912, 2007. *Crossref,* https://doi.org/10.1016/j.advwatres.2006.08.001

[25] Pedro J. Restrepo-Posada, and Peter S. Eagleson, “Identification of Independent Rainstorms,” *Journal of Hydrology,* vol. 55, no. 1–4, pp. 303–319, 1982. *Crossref,* https://doi.org/10.1016/0022-1694(82)90136-6

[26] Anisha Putri Ramadhani et al., “The Modelling Number of Daily Death Covid 19 Data Using Some of Two and One Parameter Distributions in Indonesia,” *International Journal of Mathematics Trends and Technology,* vol. 68, no. 12, pp. 106–111, 2022. *Crossref,* https://doi.org/10.14445/22315373/IJMTT-V68I12P512

[27] Vinny Anugrah et al., “Applied Some Probability Density Function for Frequency Analysis of New Cases Covid-19 in Indonesia,” *International Journal of Mathematics Trends and Technology,* vol. 68, no. 12, pp. 100–105, 2022. *Crossref,* https://doi.org/10.14445/22315373/IJMTT-V68I12P511