International Journal of Mathematics Trends and Technology
ISSN: 2231-5373/ https://doi.org/10.14445/22315373/IJMTT-V69I1P511

Volume 69 Issue 1, 71-78, January 2023

Original Article

The Effect of Simulation Storm Amount Hourly Rainfall Based on Increase Various Rates Parameters to Mean and Maximum by Using Best Distribution Model

Halimah Tun Satdiah[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: 05 December 2022 Revised: 07 January 2023 Accepted: 18 January 2023 Published: 31 January 2023

Abstract *- Flood events are very difficult to predict early, this is because the value of the amount of rainfall that occurs for each month in the next few years cannot be properly estimated, of course, will result in difficulty estimating the strength of the dam in holding back the flow rate water, in addition to the difficulty in estimating the size of the ditch as a rain reservoir for the next few years. This event will continue to occur every year along with an increased amount of rainfall. Amount rainfall variables can be determined using short-scale rainfall such as hourly obtained using storm theory. This study focuses on using hourly rainfall data to be modeled using several probability distribution models such as Gamma, Weibull, and Log Normal. Once the best model can be determined, the parameters will be modified by increasing every 10% so that it reaches 50% and using the quantile function the Amount of rainfall data simulation based on parameter modification will be carried out and the two characteristics of the simulated data such as the mean and maximum amount values will be compared to show the effect of increasing storm amount rainfall. In this study, it was found that the log-normal distribution model was the best and an increase in parameter values would affect the increase in the mean and maximum amount of rainfall for each year.*

Keywords *- Amount Storm, Storm Theory, Gamma, Weibull, Log Normal, Quantile Function.*

1. Introduction

Numerous fields, including hydrology, engineering, and environmental science, rely heavily on probability models like the Gamma, Weibull, and Log Normal Distribution [1]. Damages to human lives and property caused by rainfall events like floods or landslides can be reduced or avoided by analyzing rainfall data, which is part of hydrology. The storm theory has been applied to the analysis of a long series of hourly rain data that is dominated by non-rain data (zero data). The definition of a storm is determined by the interevent time definition (IETD), which is the minimal length of a dry interval between two concurrent storm events. As a result, the dry period needed to separate two storm events must at least match the IETD value. They wouldn't be considered as two separate events, but rather as parts of the same storm if this weren't the case. In order to reduce the serial correlation between the two different storms, the IETD value was chosen [2]. The IETD for small urban catchments is commonly taken as 6 hours since it would appear that the runoff response of subsequent storms would be independent if the time concentration of rainfall was less than 6 hours [3]. The storm amount is defined as the total quantity of rain that has at least one wet hour at the start and conclusion and either contains dry spells that last fewer than six hours or none at all. Storm depth to duration, which is the period between storm episodes, is what determines a storm's severity. Figure 1 explains what a storm is and how it is defined.. Let x_{ij} is *j*th rainfall (mm) with black circle sign on *i*-th storm and n_i is duration (hour) on *i*th storm. The storm amount on storm *i*-th is

$$\sum_{j=1}^{n} x_{ij}\,,\ \ i = 1,2,\cdots$$

Several previous studies have used several methods of seeing storms in their analysis, including [4 – 7], The determination of the best opportunity distribution model is the main issue in this study, and the relationship between two variables is mutually dependent given the marginal distribution associated with the copula model, for example, the marginal is assumed to be normal [8] may possess the exact same kind of probability distribution. Bivariate exponential [9], Bivariate gamma [10], Bivariate lognormal [11], and Extreme Value Distribution in Two Variables [12] are a few examples. The marginal distributions of storm rainfall are actually a complex phenomenon that aren't necessarily comparable or typical. It's crucial to consider alternative distributions since they could provide more accurate rainfall estimates. The copula technique is a versatile strategy that allows for the use of more marginal distributions and dependence structures in multivariate issues. [13]. Hydrology has utilized a variety of copulas, including the Archimedean copulas [14–17]. This study aims to (1) use the storm event to derive short-scale rainfall variables, (2) determine some modeling for the storm amount 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 and (4) simulate the storm amount based on the best model. 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. The best distribution model in this study will be used to simulate rainfall. The quantile function of this distribution model will be used in obtaining the simulation results. Several researchers have previously used several distribution models to simulate rainfall data. Several other researchers have used the Gamma distribution with two parameters in simulating rainfall data [18 – 21]. The three-parameter Kappa distribution was also used for this purpose, among researchers who have used the distribution [22], using the mixed exponential distribution to simulate rain for several regions in the U.S. [23], and have also done the same for the Quebec area [24].

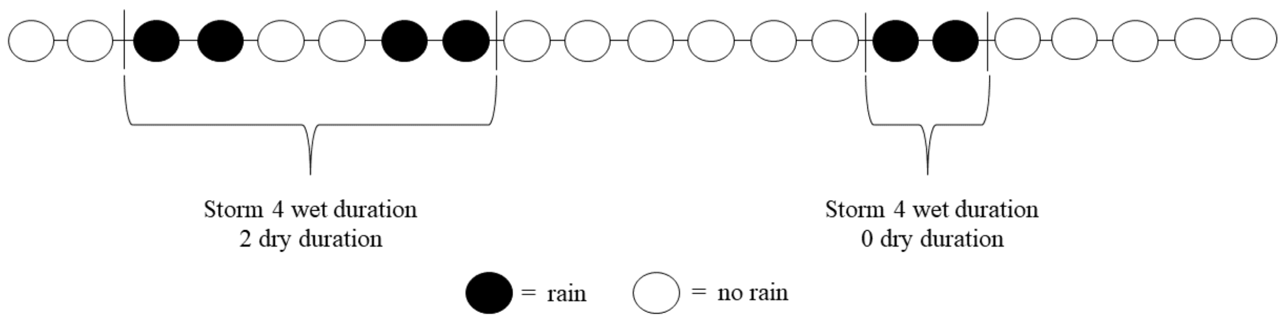

Fig. 1 Definition of storm

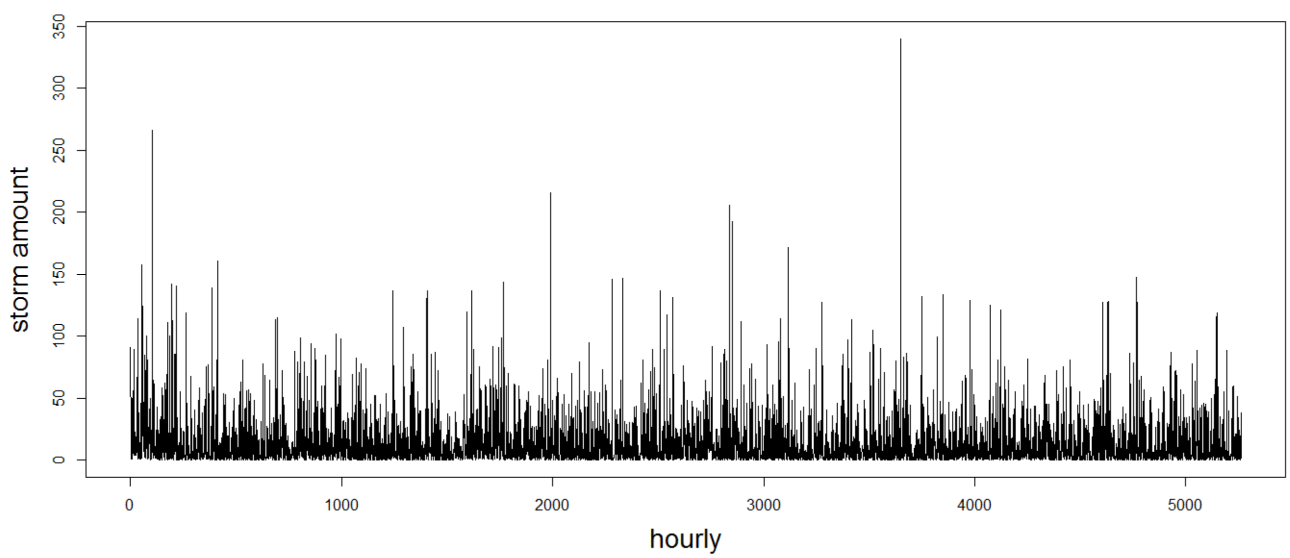

Fig. 2 The Storm amount short scale rainfall (hourly)

1.1. The Data of Storm Amount

The data that has been extracted based on storm events will be broken down into storm amount data for each storm that is generated as shown in Figure 2. Initial information about storm amounts in the form of statistical information can be described in table 1. The values shown are descriptive statistics that generally describe the storm amount of 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 storm amount 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.

Table 1. Desciptive Statistics of storm amount data

mean	var	skewness	kurtosis	minimum	maximum
13.83	441.64	4.67×10^{-9}	0.14	0.1	340.1

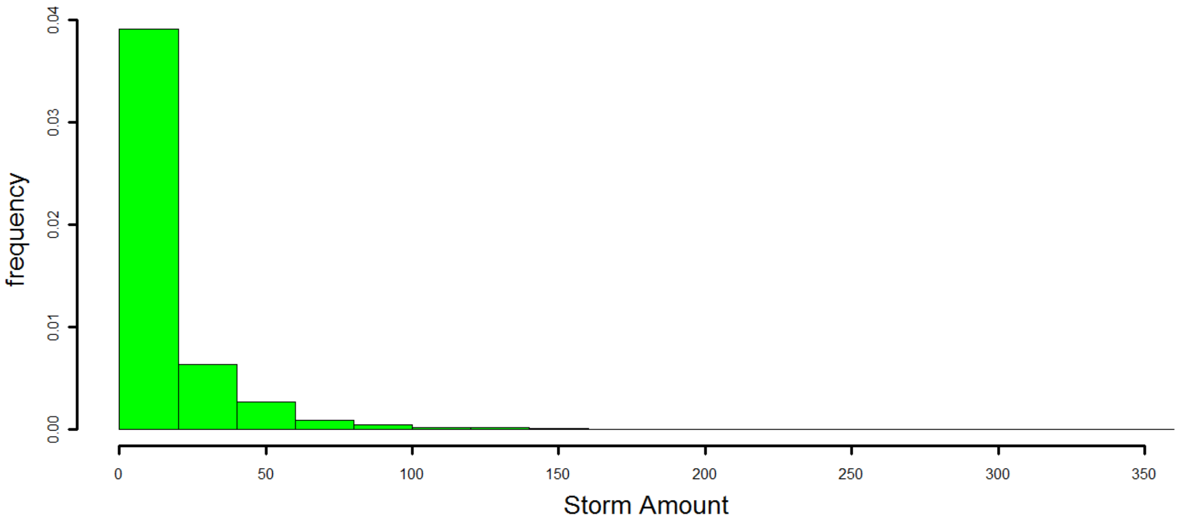

Fig. 3 The Histogram of Storm Amount

2. Methods

The amount that results from the extraction of short-scale (hourly) rainfall data is needed for the storm amount modeling. Probability density functions are the most important tools for describing the characteristics of the storm amount. This study will make use of three probability density functions with two parameters, such as Weibull, Gamma, and Log Normal. Table 2 displays the pdf and CDF for each distribution we take into consideration, with x denoting the observed values of the random variable that corresponds to 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. We will not go into detail in this paper about the implicit and intricate MLE function for this model. Newton's Rhapshon is a numerical method for solving the nonlinear equation resulting from the maximal log-likelihood function (ln L). However, iteration systems have utilized this strategy to locate the solution. For this procedure, several different initial values have been tested. It is possible to use a particular value as the selected estimation parameter if the initial value used results in iterations that converge to a particular 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.

Concepts to describe the probability distribution of a random variable *Y i*nclude distribution function $F(y) = P[Y \leq y)$ and the quantile function is $Q(u) = F^{-1}(u)$, where u = uniform (0,1). The simulation technique using the quantile function is quite simple, just by determining the inverse distribution function and changing the random variable to a uniformly distributed random number (0,1)

Let $(x_1, x_2, \ldots, x_n)$ be a random sample taken from three pdf in this study, as shown in Table 2, with parameters θ. The likelihood function (L) is a joint probability density function which can be written as follows $L = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$ and the log-likelihood function is the logarithm natural of the likelihood function (ln (L)). The nonlinear equation that comes after is also solved by the MLE $\hat{\theta}\ of\ \theta$ because it is the solution of the equation $\frac{d \ln L}{d\theta} = 0$ itself. Using the results of several goodness-of-fit tests (GOF), the most suitable distribution is determined. Number-based standards The goodness-of-fit criteria of the distributions were assessed using Akaike's information criterion (AIC) and Bayesian information criterion (BIC), with the GOF tests pdf, CDF, and QDF being taken into account. The majority of the time, graphical inspection revealed the same outcome, despite the fact that the AIC, BIC, and ln (L) findings were inconsistent. It was decided that the distribution that fit the data the best had the lowest AIC and BIC values. AIC and BIC can be calculated using the formula below:

$$AIC = -2 \log L + 2j;$$
$$BIC = -2 \log L + j \log m,$$

where j = the number of parameters, m = the sample size

3. Results

In this study, the parameters of the three probability distributions will be estimated using the MLE method as presented in table 3. 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 storm amount 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 4 shows that the two parameters distribution such as Weibull, Gamma, and Log Normal are able to approach the histogram or the frequency of storm amount data that occurs in Malaysia.

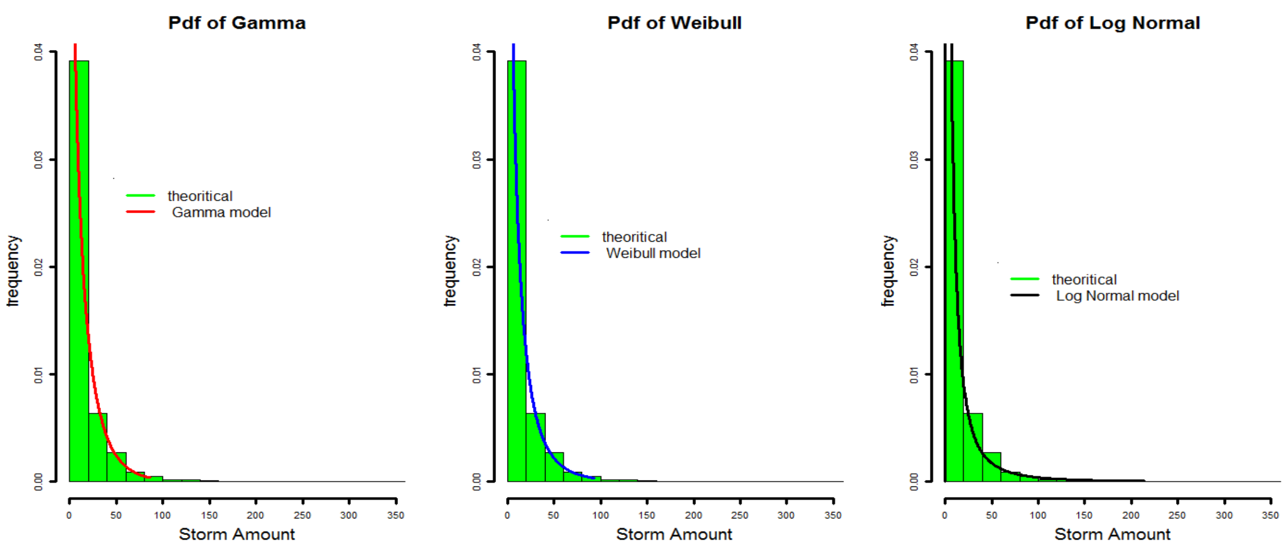

Fig. 4 Pdf plot for comparisons Estimated and observed Storm Amount

Table 3. Computed parameters of different distribution

	η	**κ**	**α**	**β**	**μ**	**σ**
Weibull	0.73	11.15				
Gamma			0.64	21.57		
Log Normal					1.67	1.49

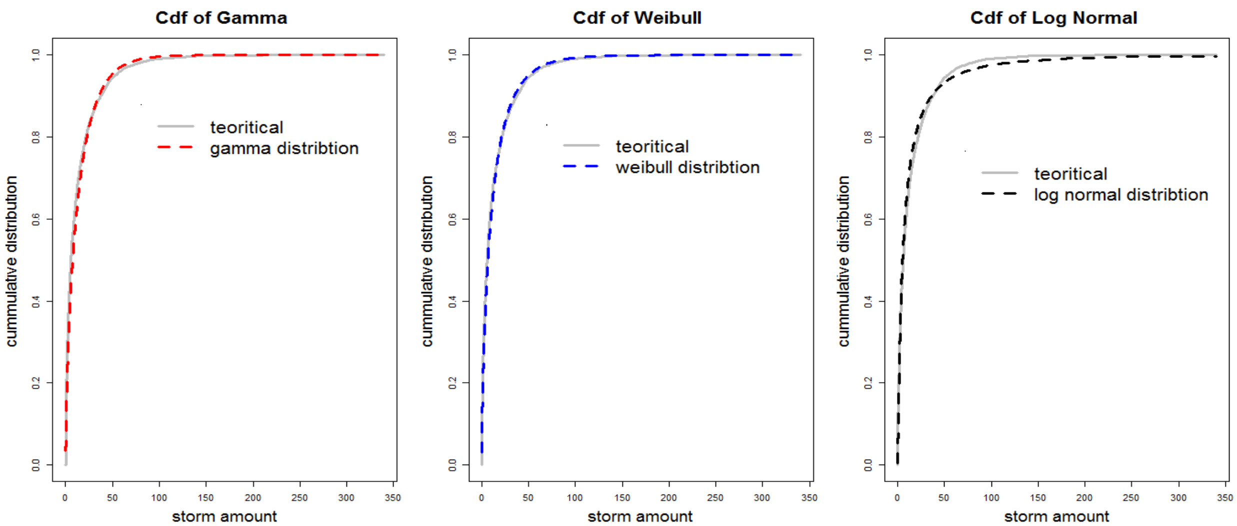

Fig. 5 Cdf plot for comparisons Estimated and observed for Storm Amount

Furthermore, the plot of the cumulative distribution function will be used to make more convincing conclusions in choosing the best distribution in this study. Therefore Figure 5 is also presented for this purpose. From the figure, it is very clear that the two-parameter distribution such as Weibull, Gamma, and Log Normal is very good at capturing the observation distribution function. Figure 3 also clarifies the conclusion that the distributions in this paper are the best models produced in this study.

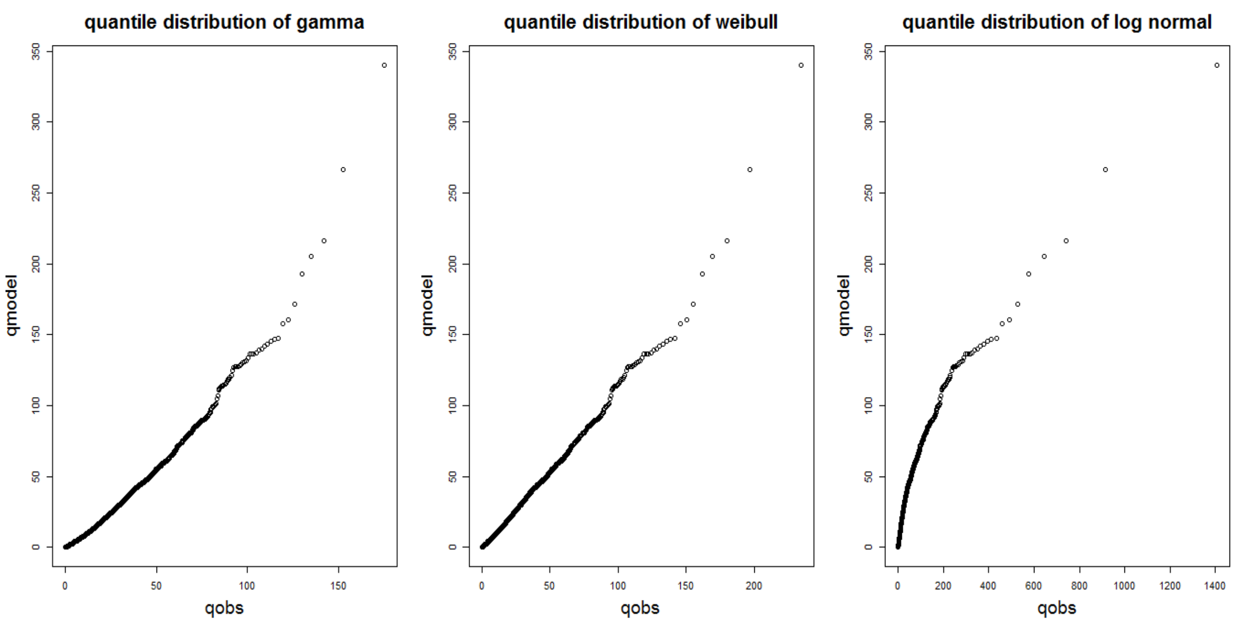

Fig. 6 Qdf plot for comparisons Estimated and observed for Storm Amount

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 QDF (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 storm amount 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 4. 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.

Table 4. The GOF test result

	Weibull	Gamma	Log Normal
AIC	37111.99	37350.99	36769.95
BIC	37125.12	37364.12	36783.08
Log (L)	-18553.99	-18673.49	-18382.97

Based on the best model in this research, storm amount simulation is run by modifying the average parameters. Modifications are executed by increasing the average from 10% to 50% while maintaining the standard deviation parameter values. The simulation is run using the quantile function of the log-normal distribution. The results of the simulation that has been run for the storm amount will produce the average and maximum storm amount for each modification of the average parameters made, as shown in figure 7. 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 [25, 26] 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.

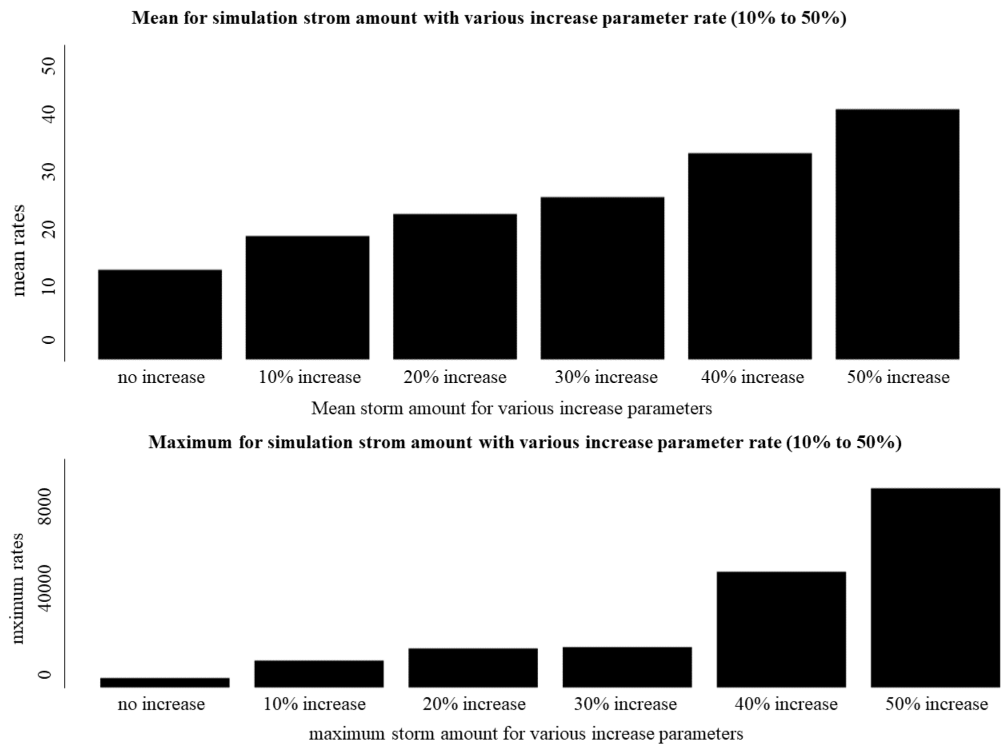

Fig. 7 Mean and maximum storm amount for various increase parameters

4. Conclusion

The Gamma, Weibull, and Log-Normal probability distributions were used to examine the occurrence probability of storm amount events in this study. The types of data in this work used MLE, especially to estimate the parameters of the three probability distributions. It is concluded that the Log Normal two parameters distribution returned better results when compared with other well-known distributions. In addition, through the best model in this study, we can use the distribution of the quantile function to simulate the storm amount for various increased parameters rate. The results of the simulations run show that the increase in parameters is accompanied by an increase in the average and maximum storm amount.

Acknowledgments

The Department of Irrigation and Drainage staff provided hourly rainfall data that made this paper possible, and the authors owe them a debt of gratitude. Without the financial support of Universitas Islam Negeri Sultan Syarif Kasim Riau, this study would not have been possible.

References

[1] R.D. Reiss, and M. Thomas, *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhauser, 1997.

[2] 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

[3] 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

[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] Barry J. Adams et al., "Meteorological Data Analysis for Drainage System Design," *Journal of Environmental Engineering*, vol. 112, no. 2, pp. 827-848, 1986. *Crossref,* https://doi.org/10.1061/(ASCE)0733-9372(1986)112:5(827)

[6] 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. 3421-3431, 1998. *Crossref,* https://doi.org/10.1029/98WR02448

[7] 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

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

[9] 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, pp. 1307. 2002. *Crossref,* https://doi.org/10.1029/2002WR001343

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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

[15] 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)

[16] 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

[17] 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)

[18] N.T. Ison, A.M. Feyerherm, and L.Dean Bark, "Wet Period Precipitation and the Gamma Distribution," *Journal of Applied Meteorology and Climatology,* vol. 10, no. 4, pp. 658-665, 1971. *Crossref,* https://doi.org/10.1175/1520-0450(1971)010%3C0658:WPPATG%3E2.0.CO;2

[19] Richard W. Katz, "Precipitation as a Chain-Dependent Process," *Journal of Applied Meteorology and Climatology*, vol. 16, no. 7, pp. 671-676, 1977. *Crossref,* https://doi.org/10.1175/1520-0450(1977)016%3C0671:PAACDP%3E2.0.CO;2

[20] T. A. Buishand, "*Stochastic Modeling of Daily Rainfall Sequences*," Medidlingen Agricultural College Wageningen, vol. 77, no. 3, 1977.

[21] E.D. Waymire, and Vijay K. Gupta, "The Mathematical Structure of Rainfall Representations: 1. A Review of the Stochastic Rainfall Models," *Water Resources Research*, vol. 17, no. 5, Pp. 1261-1272, 1981. *Crossref,* https://doi.org/10.1029/WR017i005p01261

[22] Paul W. Mielke, "Another Family of Distributions for Describing and Analyzing Precipitation Data," *Journal of Applied Meteorology and Climatology*, vol. 12, no. 2, pp. 275-280, 1973. *Crossref,* https://doi.org/10.1175/1520-0450(1973)012%3C0275:AFODFD%3E2.0.CO;2

[23] Jose Rolda'n, and David A. Woolhiser, "Stochastic Daily Precipitation Models: I. A Comparison of Occurrence Processes," *Water Resources Research,* vol. 18, no. 5, pp. 1451-1459, 1982. *Crossref,* https://doi.org/10.1029/WR018i005p01451

[24] Van-Thanh-Van Nguyen, and Alfred Mayabi, "Probabilistic Analysis of Summer Daily Rainfall for the Montreal Region," *Canadian Water Resources Journal*, vol. 16, no. 1, pp. 65-80, 1991. *Crossref,* https://doi.org/10.4296/cwrj1601065

[25] 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

[26] 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