West Science Interdisciplinary Studies
Vol. 03, No. 05, May and 2025: pp. 887-902

Analysis of The Suitability of Frequency Distribution of Rainfall Data and Rainfall Return Period at PT. X, Kutai Kartanegara Regency, East Kalimantan

Muhammad Tri Aditya[1], Helik Susilo[2]
[1] Politeknik Negeri Malang and muhammadtri@polinema.ac.id
[2] Politeknik Negeri Malang and susilohelik@polinema.ac.id

ABSTRACT

Annual maximum rainfall is a critical parameter in mine planning, particularly for the design of drainage systems and hydrometeorological disaster mitigation. This study aims to analyze the frequency distribution of rainfall and estimate design rainfall for various return periods in the operational area of PT X, located in Kutai Kartanegara Regency, East Kalimantan. The dataset comprises 11 years of annual maximum rainfall data from 2013 to 2023. Four probability distribution models were evaluated: Gumbel, Normal, Log-Normal, and Log Pearson Type III. Goodness-of-fit testing was conducted using the Chi-Square and Kolmogorov-Smirnov methods. The Chi-Square test results indicate that all distributions are acceptable at a 5% significance level, but only the Gumbel distribution is valid at a 1% level. The Kolmogorov-Smirnov test confirms that all distributions are statistically acceptable at both significance levels. Design rainfall values were calculated for return periods of 2, 5, 10, 25, 50, and 100 years. The highest rainfall for the 2-year return period was produced by the Normal distribution (307.24 mm), while the Gumbel distribution yielded the highest values for return periods from 5 to 100 years, reaching a peak of 493.86 mm for the 100-year return period. These results suggest that the Gumbel distribution is the most suitable model for representing extreme rainfall events in the study area and is recommended for application in mine drainage system design and surface runoff control planning.

Keywords: *Annual Maximum Rainfall, Probability Distribution, Return Period, Chi-Square Test, Kolmogorov-Smirnov Test, Gumbel Distribution.*

1. INTRODUCTION

Rainfall is one of the most critical climatic variables impacting both natural systems and human activities. It serves as an essential resource for agriculture and ecosystem sustainability but may also trigger hazards such as flooding and landslides when extreme values occur [1] . In tropical regions like Indonesia, rainfall patterns are primarily governed by large-scale climate drivers, including the Asian–Australian Monsoon system, ENSO, Walker, and Hadley circulations [2] . These complex interactions lead to high variability and frequent extreme events.

In mining operations, comprehensive water management is indispensable—not only for maintaining operations but also for preventing environmental degradation and geotechnical failures [3][4]. Rainfall directly influences slope stability, site dewatering, and infrastructure resilience [5]. In particular, the pronounced variability of rainfall in tropical climates presents challenges for designing drainage systems, settling ponds, and erosion control structures [6].

Estimating design rainfall—or extreme rainfall associated with specific return periods—is key to hydrologic infrastructure planning. This is typically accomplished using frequency analysis of annual maximum rainfall via probabilistic models such as Gumbel, Normal, Log-Normal, and Log-Pearson Type III distributions [7]. Recent studies continue to endorse these models, comparing them using goodness-of-fit tests like Chi-Square and Kolmogorov–Smirnov, as well as newer diagnostics (e.g., RMSE, L-Moments) [8]. International reviews confirm that methods remain consistent—but performance can vary by region, with Gumbel and Log-Pearson III often outperforming others for tropical or monsoonal climates [9].

This study investigates rainfall frequency characteristics and derives design rainfall values for multiple return periods in PT X's operational area in Kutai Kartanegara, East Kalimantan. The objectives are to Evaluate the fit of four probabilistic models (Gumbel, Normal, Log-Normal, Log-Pearson III) using Chi-Square and K-S goodness-of-fit tests, Estimate design rainfall depths for return periods of 2, 5, 10, 20, 50, and 100 years, using each model, and Compare the performance of each method and recommend the most suitable distribution for designing water-management infrastructure in tropical mine sites. The outcomes aim to provide robust, empirical support for hydrologic design practices in mining contexts with complex tropical rainfall regimes.

2. LITERATURE REVIEW

2.1 *Rainfall Analysis*

Rainfall refers to the volume of precipitation that falls within a specific region [10]. It plays a crucial role in designing drainage systems, particularly in mining operations, as the volume of rainfall directly affects the quantity of mine water that must be managed [11]. However, raw rainfall data cannot be used immediately for the design of mine water management structures [12]. Instead, it requires further processing to derive accurate and representative rainfall values. This processed data becomes a primary input in planning drainage systems for open-pit mines [13].

Rainfall measurement is typically conducted using a rain gauge instrument [14], which is available in two main types: manual and automatic. These instruments are usually installed in unobstructed outdoor areas to ensure rainwater is not hindered by surrounding structures or vegetation. Rainfall data obtained from these instruments is critical for estimating design rainfall. Two primary statistical methods are commonly used in analyzing this data: the Annual Series method, which involves selecting the single highest rainfall value recorded each year; and the Partial Duration Series, which includes all rainfall events that exceed a predefined threshold, regardless of when they occur.

Design rainfall refers to the maximum anticipated rainfall that could occur over the projected lifespan of a given drainage facility [15]. To determine rainfall corresponding to a specific return period and to assess short-duration rainfall intensity (e.g., one-hour rainfall), a detailed rainfall analysis must be conducted.

2.2 *Normal Distribution Method*

The Normal distribution is widely used in hydrological analysis [16], for example in rainfall frequency analysis, calculation with the normal distribution can practically be approached by the following equation:

$$X_T = \bar{x} + z.s$$

Description:

X_T = expected value with T-year return period

$\bar{x}$ = calculated mean value of the variate

S = standard deviation of the variate value

Z = frequency factor of the normal distribution

A function of chance or return period is a type of mathematical model of chance distribution used for chance analysis.

2.3 Log-Normal Distribution

The log-normal distribution is the result of the transformation of the normal distribution [17], namely by changing the value of X to the logarithmic value of X. If Y= log X, then the calculation with the normal distribution can practicipally be approached by the following equation:

$$\mathbf{Y_T = \overline{Y} + z.s}$$

Description:

Y_T = expected value for the T-annual return period

$\overline{Y}$ = calculated mean value of the variate

S = standard deviation of the variate value

Z = frequency factor of the normal distribution

A function of chance or return period type.

2.4 Gumble Distribution

The Gumble distribution also called the extreme distribution is generally used for maximum data analysis calculation of planned rainfall according to the Gumble method [18], has the following formula:

$$\mathbf{X_T = \overline{x} + s.K}$$

Description:

X_T = estimated value expected to occur with a return period of T-Years

$\overline{x}$ = calculate mean value of the variate

S = standard deviation of the variate value

K = frequency factor

The probability factor K for extreme Gumble prices can be expressed by the following equation:

$$K = \frac{Y_{T_r} - Y_n}{S_n}$$

Description:

Y_T = reduced variated

Yn = reduced mean which depends on the number of samples n

Sn = reduced standard deviation, depending on the amount of data

2.5 Log Pearson Distribution

The log pearson type III distribution is widely used in hydrological analysis, especially in analyzing data with extreme values [19]. The Log Pearson type III distribution form is the result of a transformation of the Pearson type III distribution by replacing the variate with logarithmic distribution can practically be approximated by the following eaquation:

$$\mathbf{Y_T = \overline{Y} + K._{TS}}$$

Description:

Y_T = expected value with T-years return period
$\overline{Y}$ = calculate mean value of the variate
S = standard deviation of the variate value
K_T = frequency factor, the value of K_T depends on the slope coefficient and probability

2.6 *Chi-Square Test*

The Chi-Square test is intended to determine whether the distribution equation that has been selected can represent the statistical distribution of the data sample being analyzed [20]. Chi-Square test statistics are determined based on the following equation:

$$\chi_h{}^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$$

Description:
$X_h{}^2$ = calculated Chi-Square parameter
M = number of categories
O_i = frequency of observation
E_i = expected frequency

2.7 *Smirnov Kolmogorof Frequency*

This suitability test is used to test horizontal deviations. The Smirnov-Kolmogorof suitability test, often called the non-parametric suitability test, because the test does not use a specific distribution function [21]. This test is used to test the largest deviation/difference between observational (empirical) opportunities and theoretical opportunities.

This test is carried out with the following stages:

1. Sort the maximum daily rainfall data from the smallest to the largest value
2. Plot the maximum daily rainfall price Xt, with the probability price, Sn (x), as in the equation above
3. Test the suitability of the data using the available table with parameters of the amount of data (n), confidence level/significance level (α), and Δcr
4. Calculate the maximum difference value between the theoretical distribution and the empirical distribution with the equation:

$$\Delta_{max} = | P_e - P_t |$$

Where to compare the values of Δcr and Δmax with the provisions if:
Δcr > Δmax, then the distribution is not accepted
Δcr < Δmax, then the distribution is accepted

2.8 *Rainfall Return Period*

According to Triatmodjo (2008), the return period is defined as a hypothetical time at which a certain amount of discharge or rainfall (XT) will be equal to or exceeded once in that period. Based on discharge or rainfall data for several years of observation, it can be estimated that the expected discharge/rain will be equal to or exceeded once in T years; and the discharge/rain is known as discharge/rain with a return period of T years or annual discharge/rain T. To find the probability, we can use the Weibull equation, namely:

$$P = \frac{m}{n+1}$$

while the recurrence period can be found using the equation :

$$T_r = \frac{1}{P}$$

With:

m = data ranking sequence number after being sorted from largest to smallest,

n = amount of data or number of events,

P = probability,

Tr = recurrence period

3. METHODS

This study uses annual maximum rainfall data for the period 2013 to 2023 obtained from the operational area of PT X in Kutai Kartanegara, East Kalimantan. The analysis was conducted to determine the best probability distribution that can represent rainfall data and calculate the planned rainfall based on a specified return period. The four probability distribution models used are Gumbel, Normal, Log-Normal, and Log Pearson Type III. The Gumbel model is used to analyze extreme events, while the Log-Normal and Log Pearson III distributions are used for data that shows skewness. The Normal distribution is used as a comparison because it assumes that the data is symmetrically distributed.

The suitability test was carried out using two statistical approaches, namely the Chi-Square (χ^2) test and the Kolmogorov-Smirnov (K–S) test, to assess the suitability between the distribution model and the actual data. The values of statistical parameters such as mean, standard deviation, coefficient of variation, skewness, and kurtosis are calculated as the basis for calculating the distribution. The return period is calculated using the Weibull method to obtain the planned rainfall value for periods of 2, 5, 10, 20, 50, and 100 years. The distribution model with the best test results is used as the basis for planning a mine water management system and mitigating hydrometeorological risks in the study area.

4. RESULTS AND DISCUSSION

4.1 The Rainfall Data

The rainfall data used for the analysis is the annual maximum rainfall data from 2013 - 2023 obtained from PT. X data, the rainfall data used is as follows:

Table 1. Rainfall Data 2013-2023

Rainfall 2013-2023 (mm)											
Month	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
January	269.1	141.2	137.9	287.8	81.3	200.9	258.2	197	189	261.6	274
February	115.4	176.2	29.8	146	10.7	67.8	147.2	79	30.7	335.1	179
March	181.9	189.4	58.2	136.2	42.5	292.9	230.8	223.5	117.5	331.5	386.5
April	169.9	184.2	296.1	151.3	185.2	216.7	321	206.7	79.5	213	142.6
May	177.8	254.2	142.7	258	184.6	326.3	159.5	264.8	273	160.4	241
June	152.1	173.2	164.1	350.2	192.5	125.5	166.3	211.7	177.5	143.5	174.8
July	80.8	136	112	42.5	119.3	251.3	71	77.2	172.5	169.5	133.3
August	67.3	132.5	178.2	1.2	55.5	125.2	104.6	102	121.1	306.7	165.5
September	94.5	190.9	57.2	39	252.5	187.2	44.5	37.5	158.1	194.5	195.6
October	75.2	137.8	58.4	27.5	184.9	154.6	184.8	272.8	218.1	149.9	262
November	128.3	94.6	156.7	170	226.8	291.2	201.3	161	181.8	244.5	258.4
December	134	120.5	282.6	124.1	114.7	64.6	198	298	212.2	352.7	116
Rainfall Max	269.1	254.2	296.1	350.2	252.5	326.3	321	298	273	352.7	386.5

The table above shows the monthly fluctuations in rainfall in PT. X, Kutai Kartanegara Regency over the past 11 years, with a general pattern of the rainy season occurring from November to April and the dry season from June to September. The highest rainfall was consistently recorded in January and March, with a peak in March 2023 of 386.5 mm. Conversely, the lowest rainfall occurred in dry months, such as February 2017 (10.7 mm). Variations between years are quite significant, reflecting certain climate anomalies, such as high rainfall in August 2022 (306.7 mm). Overall, these data indicate dynamic climate trends that are important to monitor in environmental and water resource planning.

4.2 Frequency Analysis of Rainfall Data

1. Calculation of the Gumbel Distribution Method

Table 2. Gumbel Distribution Method

No	Years	R_{MAX}(Xi)	X	(Xi-X)	$(Xi-X)^2$
1	2013	386.5	307.2	79.26	6282.7
2	2014	352.7	307.2	45.46	2066.9
3	2015	350.2	307.2	42.96	1845.9
4	2016	326.3	307.2	19.06	363.4
5	2017	321	307.2	13.76	189.4
6	2018	298	307.2	-9.24	85.3
7	2019	296.1	307.2	-11.14	124.0
8	2020	273	307.2	-34.24	1172.1
9	2021	269.1	307.2	-38.14	1454.4
10	2022	254.2	307.2	-53.04	2812.9
11	2023	252.5	307.2	-54.74	2996.1
Sum		3379.6	Sum		19393.2
Average		307.24	Average		1763.0
Standard Deviation		44.04			

Source: Calculation Result

Gumbel Distribution Analysis in Table 2 shows the annual maximum rainfall (Rmax) data from 2013 to 2023, with an average of 307.2 mm and a standard deviation of 44.04 mm. The difference between each data point and the average is calculated to determine the data variation, which is then squared to obtain a total variance of 19,393.2 mm². These results are the basis for estimating the planned rainfall using the Gumbel formula, which is important in hydrological planning to determine extreme rainfall based on a certain return period.

2. Calculation of Normal Distribution Method

Table 3. Normal Distribution Method

No	Years	R_{MAX}(Xi)	X	(Xi-X)	$(Xi-X)^2$
1	2013	386.5	307.236	79.26	6282.724
2	2014	352.7	307.236	45.46	2066.942
3	2015	350.2	307.236	42.96	1845.874
4	2016	326.3	307.236	19.06	363.422
5	2017	321	307.236	13.76	189.438
6	2018	298	307.236	-9.24	85.310
7	2019	296.1	307.236	-11.14	124.019
8	2020	273	307.236	-34.24	1172.129
9	2021	269.1	307.236	-38.14	1454.382
10	2022	254.2	307.236	-53.04	2812.856
11	2023	252.5	307.236	-54.74	2996.070
Sum		3379.6	Sum		19393.165
Average		307.23636	Average		1763.015
Standard Deviation		44.037672			

Source: Calculation Result

The normal distribution analysis of the annual maximum rainfall frequency from 2013 to 2023 shows that the average maximum rainfall ($\bar{X}$) is 307.24 mm, with a standard deviation of 44.037 mm. The difference between each rainfall value (Xi) and the average is calculated to determine the distribution of the data, and the squared result of the difference produces a total variance of 19,393.17 mm². Since these values are relatively symmetrically distributed around the average, this data is suitable for analysis using a normal distribution, which can be used to estimate the probability of extreme rainfall events based on a certain frequency or return period.

3. Calculation of Log Normal Distribution Method

Table 4. Log Normal Distribution Method

No	Years	R_{MAX}(Xi)	Log Xi	Log Xi - Log $X_{Average}$	(Log Xi - Log $X_{Average}$)^2
1	2013	386.5	2.587	0.1037	0.01075
2	2014	352.7	2.547	0.0639	0.00409
3	2015	350.2	2.544	0.0608	0.00370
4	2016	326.3	2.514	0.0301	0.00091
5	2017	321.0	2.507	0.0230	0.00053
6	2018	298.0	2.474	-0.0093	0.00009
7	2019	296.1	2.471	-0.0120	0.00014
8	2020	273.0	2.436	-0.0473	0.00224
9	2021	269.1	2.430	-0.0536	0.00287
10	2022	254.2	2.405	-0.0783	0.00613
11	2023	252.5	2.402	-0.0812	0.00659
Sum		3379.6	27.318	Sum	0.038040408
Average		307.2	2.483		
S LogX		0.0617			

Source: Calculation Result

The Log Normal Distribution Method table shows the analysis of the annual maximum rainfall (Rmax) in 2013 to 2023 using the log normal distribution approach. Each Rmax value is converted to logarithmic form (log Xi), then the difference with the log average (log $\bar{X}$ = 2.483), and squared to obtain the logarithmic variance. The results show a log average rainfall of 2.483 and a logarithmic standard deviation (S LogX) of 0.0617, reflecting the log-normal data distribution. This analysis is used to predict extreme rainfall in a certain return period more accurately on asymmetric data.

4. Calculation of Log Pearson Type III Distribution Method

Table 5. Log-Pearson Type III Distribution Method

No	Years	MAX(Xi)	LOG Xi	LogXi - Log $X_{Average}$	(Log Xi - Log $X_{Average}$)^2	(Log Xi - Log $X_{Average}$)^3
1	2013	269.1	2.429913698	-0.05356	0.00287	-0.00015
2	2014	254.2	2.405175546	-0.07829	0.00613	-0.00048
3	2015	296.1	2.471438407	-0.01203	0.00014	0.00000
4	2016	350.2	2.544316142	0.06085	0.00370	0.00023
5	2017	252.5	2.402261382	-0.08121	0.00659	-0.00054
6	2018	326.3	2.513617074	0.03015	0.00091	0.00003
7	2019	321	2.506505032	0.02304	0.00053	0.00001
8	2020	298	2.474216264	-0.00925	0.00009	0.00000
9	2021	273	2.436162647	-0.04731	0.00224	-0.00011
10	2022	352.7	2.54740546	0.06394	0.00409	0.00026

11	2023	386.5	2.587149498	0.10368	0.01075	0.00111
Sum		3379.6	27.31816115	Sum	0.03804	0.00036
Average		307.2363636				
S LOG X		0.0617				
Cs		0.1892				

Source: Calculation Result

The Log-Pearson Type III Distribution Method table presents the analysis of the annual maximum rainfall (Rmax) from 2013 to 2023 using the Log-Pearson Type III distribution approach, which is commonly used in hydrology to model extreme data with asymmetric tendencies. The Rmax value is converted to the logarithm (Log Xi), then the difference is calculated from the log mean (Log $\bar{X}$ = 2.483), the square of the difference for the variance (σ^2), and the cube of the difference to calculate the skewness coefficient (Cs). The analysis results show a logarithmic standard deviation (S Log X) of 0.0617 and a skewness (Cs) of 0.1892, indicating slight positive asymmetry. This distribution allows for more accurate estimation of design rainfall based on return periods for data that does not follow a perfect normal distribution.

4.3 Frequency Distribution Goodness-of-Fit Test

1. Chi-Square Test

Table 6. Chi Square value of Gumbel distribution method

No	**Class Limit**	**Number of data**		**Oj-Ej**	**(Oj- Ej)²/Ej**
		Ej	Oj		
1	<269.68	2.75	3	0.25	0.022727273
2	269.68 - 300.21	2.75	3	0.25	0.022727273
3	300.21-338.93	2.75	2	-0.75	0.204545455
4	>338.93	2.75	3	0.25	0.022727273
	Sum	11	11	0	0.272727273

Conclusion		
α	X² critical	Description
5%	3.841	Accepted
1%	0.635	Accepted

Source: Calculation Result

The Gumbel Distribution Chi-Square Test Table shows the suitability test of the annual maximum rainfall data to the Gumbel distribution. The data is divided into four classes based on the distribution value limits, then compared between the expected frequency (Ej = 2.75) and the observed frequency (Oj). The difference (Oj–Ej) is calculated and squared to obtain the value $(Oj-Ej)^2/Ej$, which is then added up to a total Chi-Square value of 0.273. This value is compared to the critical Chi-Square value at the 5% (3.841) and 1% (0.635) significance levels, and because the test results are smaller than both, the Gumbel distribution is accepted as an appropriate model for the maximum rainfall data.

Table 7. Chi Square value of Normal distribution method

No	**Class Limit**	**Number of data**		**Oj-Ej**	**(Oj- Ej)²/Ej**
		Ej	Oj		
1	<277.73	2.75	4	1.25	0.568
2	277.73 - 307.24	2.75	2	-0.75	0.205
3	307.24 - 336.74	2.75	2	-0.75	0.205
4	>336.74	2.75	3	0.25	0.023

Sum			11	11	0	1.00
Conclusion						
α	X² critical	Description				
5%	3.841	Accepted				
1%	0.635	Rejected				

Source: Calculation Result

The total Chi-Square value obtained is 1.00. This result is compared with the critical Chi-Square value at the significance level:

α = 5% (critical value 3.841): accepted, because 1.00 < 3.841
α = 1% (critical value 0.635): rejected, because 1.00 > 0.635

The normal distribution can still be accepted at the 5% significance level, but is rejected at a stricter level of 1%, which indicates that the fit of the normal distribution to the data is not as strong as other distributions.

Table 8. Chi Square value of Log-Normal distribution method

No	Class Limit	Number of data		Oj-Ej	(Oj- Ej)²/Ej
		Ej	Oj		
1	<276.79	2.75	4	1.25	0.568
2	276.79 - 304.42	2.75	3	0.25	0.023
3	304.42 - 334.81	2.75	2	-0.75	0.205
4	>334.81	2.75	2	-0.75	0.205
Sum		11	11	0	1.00
Conclusion					
α	X² critical	Description			
5%	3.841	Accepted			
1%	0.635	Rejected			

Source: Calculation Result

The results of all classes are summed up and produce a total Chi-Square value of 1.00. This result is compared to the critical value at the significance level:

5% (χ^2 = 3.841): accepted, because 1.00 < 3.841
1% (χ^2 = 0.635): rejected, because 1.00 > 0.635

The log normal distribution is still appropriate and accepted at the 5% significance level, but rejected at the 1% significance level, which means this model is quite good, but not as accurate as when tested at a higher level of confidence.

Table 9. Chi Square value of Log-Pearson Type III distribution method

No	Class Limit	Number of data		Oj-Ej	(Oj- Ej)²/Ej
		Ej	Oj		
1	<276.65	2.75	4	1.25	0.568
2	276.65 - 302.35	2.75	2	-0.75	0.205
3	302.35-338.23	2.75	2	-0.75	0.205
4	>338.23	2.75	3	0.25	0.023

Sum			11	11	0	1.00
Conclusion						
α	X^2 critical	Description				
5%	3.841	Accepted				
1%	0.635	Rejected				

Source: Calculation Result

The total number of chi-square values produced is 1.00. These results are compared with the critical value of Chi-Square:

α = 5% (critical value 3.841): accepted, because 1.00 < 3.841
α = 1% (critical value 0.635): rejected, because 1.00 > 0.635

The Log Pearson III distribution is accepted at the 5% significance level, but not accepted at the 1% level, which means that this model is quite suitable to represent maximum rainfall data, but is less strong when tested at a stricter significance level.

2. **Kolmogorov-Smirnov Test (K-S Test)**

Table 10. K-S Test Value of Gumbel Distribution Method

No	Years	Max(Xi)	R_{Max} (mm)	m	*Pe(R)*	K	Y_T	Tr	*P_t[R]*	*(Pe-Pt)*
1	2013	269.1	386.5	1	0.08	1.80	2.24	9.91	0.10	0.018
2	2014	254.2	352.7	2	0.17	1.03	1.50	4.99	0.20	0.034
3	2015	296.1	350.2	3	0.25	0.98	1.44	4.76	0.21	0.040
4	2016	350.2	326.3	4	0.33	0.43	0.92	3.04	0.33	0.004
5	2017	252.5	321	5	0.42	0.31	0.80	2.77	0.36	0.055
6	2018	326.3	298	6	0.50	-0.21	0.30	1.91	0.52	0.024
7	2019	321	296.1	7	0.58	-0.25	0.25	1.85	0.54	0.044
8	2020	298	273	8	0.67	-0.78	-0.25	1.38	0.72	0.057
9	2021	273	269.1	9	0.75	-0.87	-0.34	1.33	0.75	0.004
10	2022	352.7	254.2	10	0.83	-1.20	-0.67	1.17	0.86	0.024
11	2023	386.5	252.5	11	0.92	-1.24	-0.70	1.15	0.87	0.049
Sum		3379.60				Δ_{Calc}				0.057
Average		307.24								
Standard Deviation		44.04								
lots of data		11								
Yn		0.4995								
Sn		0.9676								
Conclusion										
α	Δcritical	Δcalc								
5%	0.41	Accepted								
1%	0.49	Accepted								

Source: Calculation Result

The following table of Smirnov-Kolmogorov Calculation of Gumbel Distribution shows the suitability test between the empirical distribution of the annual maximum rainfall data and the theoretical Gumbel distribution, using the Kolmogorov-Smirnov (K-S) method. The data is sorted by the maximum value (R), then the empirical probability (Pe) and theoretical probability (Pt) of the Gumbel distribution are calculated. The absolute difference between Pe and Pt at each point (|Pe -

Pt|) is called Δ, and the maximum value of this difference is used as the calculated Δ, which is 0.057. The calculated Δvalue is compared with the critical Δvalue at the significance level:

α = 5% (Δcritical = 0.41): accepted, because 0.057 < 0.41
α = 1% (Δcritical = 0.49): accepted, because 0.057 < 0.49

Based on the Smirnov-Kolmogorov test, the Gumbel distribution is accepted as a model that fits the annual maximum rainfall data at both the 5% and 1% significance levels, because the calculated Δvalue is smaller than the critical limit. This indicates that the Gumbel distribution is statistically representative enough to model the rainfall data.

Table 11. K-S Test Value of Normal Distribution Method

NO	Years	RMAX(Xi)	$R_{sequence}$	m	Pe [R]	Z/Kt	*Pt [R]*	*(Pe[R] - Pt[R])*
1	2013	269.1	386.5	1	0.083	1.80	0.08	0.003
2	2014	254.2	352.7	2	0.167	1.03	0.15	0.017
3	2015	296.1	350.2	3	0.250	0.98	0.22	0.030
4	2016	350.2	326.3	4	0.333	0.43	0.23	0.101
5	2017	252.5	321	5	0.417	0.31	0.32	0.100
6	2018	326.3	298	6	0.500	-0.21	0.48	0.020
7	2019	321	296.1	7	0.583	-0.25	0.54	0.047
8	2020	298	273	8	0.667	-0.78	0.58	0.089
9	2021	273	269.1	9	0.750	-0.87	0.71	0.040
10	2022	352.7	254.2	10	0.833	-1.20	0.79	0.045
11	2023	386.5	252.5	11	0.917	-1.24	0.89	0.027
Sum		3379.60			Δ_{Calc}			0.101
Average		307.24						
Standard Deviation		44.04						
lots of data		11						
	Conclusion							
α	Δcritical	Δcalc						
5%	0.41	Accepted						
1%	0.49	Accepted						

Source: Calculation Result

The maximum rainfall data from 2013 to 2023 is sorted by value (Rurut), then the empirical probability is calculated (Pe[R] = m/n+1) and compared with the theoretical probability of the normal distribution (Pt[R]) obtained from the Z value (standard normal). The absolute difference between Pe and Pt is calculated at each point, and the highest value is used as Δcalc, which is 0.101. This value is compared with Δcritical at two levels of significance:

α = 5% (Δcritical = 0.41) → 0.101 < 0.41 → accepted
α = 1% (Δcritical = 0.49) → 0.101 < 0.49 → accepted

The normal distribution is accepted as a suitable model for the annual maximum rainfall data based on the Smirnov-Kolmogorov test, because the calculated Δvalue is much smaller than the Δcritical value at both the 5% and 1% significance levels. This shows that the normal distribution can statistically represent the data distribution well.

Table 12. K-S Test Value of Log Normal Distribution Method

No	Years	MAX(Xi)	$R_{sequence}$	Log Xi	$Rlog_{sequence}$	m	*Pe [R]*	Z/Kt	*Pt[R]*	*(Pe[R] - Pt[R])*

1	2013	269.1	386.5	2.430	2.587	1	0.08	1.6800	0.0515108	0.031822533
2	2014	254.2	352.7	2.405	2.547	2	0.17	1.0316	0.157115607	0.00955106
3	2015	296.1	350.2	2.471	2.544	3	0.25	0.9829	0.167946399	0.082053601
4	2016	350.2	326.3	2.544	2.514	4	0.33	0.4966	0.320983008	0.012350326
5	2017	252.5	321	2.402	2.507	5	0.42	0.3831	0.361078627	0.05558804
6	2018	326.3	298	2.514	2.474	6	0.50	-0.1518	0.550100831	0.050100831
7	2019	321	296.1	2.507	2.471	7	0.58	-0.2004	0.567284667	0.016048666
8	2020	298	273	2.474	2.436	8	0.67	-0.7678	0.658002818	0.008663848
9	2021	273	269.1	2.436	2.43	9	0.75	-0.8650	0.67828267	0.07171733
10	2022	352.7	254.2	2.547	2.405	10	0.83	-1.2703	0.762782055	0.070551279
11	2023	386.5	252.5	2.587	2.402	11	0.92	-1.3189	0.772921981	0.143744686
Sum					27.3170			ΔCalc		0.143744686
Average					2.4834					
Standard Deviation					0.0617					
lots of data					11					

Conclusion		
α	Δcritical	Δcalc
5%	0.41	Accepted
1%	0.49	Accepted

Source: Calculation Result

Table 12 shows the results of the Kolmogorov–Smirnov test for the log-normal distribution of the annual maximum rainfall data. The maximum value of the difference between the empirical probability (Pe) and the theoretical probability (Pt) is Δcalc = 0.144. Since Δcalc is smaller than the critical value at the 5% (0.41) and 1% (0.49) significance levels, the log-normal distribution is accepted as an appropriate model for the data.

Table 13. K-S Test Value of Log-Pearson III Distribution Method

NO	Years	MAX(Xi)	LOG Xi	$Log_{sequence}$	m	$P_e[R]$	K	*Pt[R]*	*(Pe[R] - Pt[R])*
1	2013	269.1	2.43	2.587	1	0.08	1.679	0.056	0.02749
2	2014	254.2	2.41	2.547	2	0.17	1.030	0.158	0.00915
3	2015	296.1	2.47	2.544	3	0.25	0.981	0.168	0.08212
4	2016	350.2	2.54	2.513	4	0.33	0.479	0.322	0.01087
5	2017	252.5	2.40	2.506	5	0.42	0.365	0.362	0.05470
6	2018	326.3	2.51	2.474	6	0.50	-0.154	0.528	0.02770
7	2019	321	2.51	2.471	7	0.58	-0.202	0.539	0.04461
8	2020	298	2.47	2.436	8	0.67	-0.770	0.667	0.00062
9	2021	273	2.44	2.422	9	0.75	-0.997	0.719	0.03128
10	2022	352.7	2.55	2.405	10	0.83	-1.272	0.781	0.05217
11	2023	386.5	2.59	2.402	11	0.92	-1.321	0.792	0.12448
Sum			27.31816115			ΔCalc			0.12448
Average			2.483469196						
Standard Deviation			0.061676907						
lots of data			11						

CS	0.189256946	
Conclusion		
α	Δcritical	Δcalc
5%	0.41	Accepted
1%	0.49	Accepted

Source: Calculation Result

Table 13 shows the results of the Kolmogorov–Smirnov test on the annual maximum rainfall data using the Log Pearson Type III distribution. The empirical probability (Pe[R]) is compared with the theoretical probability (Pt[R]) of the log Pearson III model. The maximum difference between the two produces a value of Δcalc = 0.124.

Because Δcalc is smaller than the critical value at the significance level:

5% (0.41) → accepted
1% (0.49) → accepted

The Log Pearson Type III distribution is accepted as a suitable model for the annual maximum rainfall data, because the results of the Kolmogorov–Smirnov test show suitability at the 5% and 1% significance levels.

4.4 Planning Rainfall Data

1. Gumbel Distribution Plan Rainfall Data

Table 14. Gumbel Distribution Plan Rainfall

Rainfall Plan Period Year – Gumbel Method			
Return Period	YT	K	Rainfall
2	0.366	-0.1379702	301.160476
5	1.499	1.03296817	352.725877
10	2.2502	1.80932203	386.914694
25	3.1985	2.78937578	430.073979
50	3.9019	3.51632906	462.087309
100	4.6001	4.23790823	493.863976

Source: Calculation Result

This table presents the results of the design rainfall calculation based on the Gumbel Method for a return period of 2 to 100 years. As the return period increases, the frequency factor (K) and maximum rainfall values increase, indicating that greater extreme rainfall is likely to occur over a longer period of time. For example, the maximum rainfall for a 2-year period is 301.16 mm, while for 100 years it reaches 493.86 mm.

2. Normal Distribution Plan Rainfall Data

Table 15. Normal Distribution Plan Rainfall

Rainfall Plan Period Year – Normal Method		
Return Period	KT	Rainfall
2	0	307.2363636
5	0.84	344.228008
10	1.28	363.6045837
20	1.64	379.4581455
50	2.05	397.513591
100	2.33	409.8441391

Source: Calculation Result

As the return period increases, the KT value and maximum rainfall increase. For example, for the 2-year Return Period the maximum rainfall is estimated to be 307.24 mm, while for the 100-year Return Period it reaches 409.84 mm. This reflects that the less frequent the extreme rainfall events, the higher the rainfall intensity is expected.

3. **Log Normal Distribution Plan Rainfall Data**

Table 16. Log Normal Distribution Plan Rainfall

Rainfall Plan Period Year – Log Normal Method			
Return Period	KT	LOG XT	Rainfall
2	0	2.483469196	304.4172059
5	0.84	2.535277797	342.9871086
10	1.28	2.562415636	365.1031972
20	1.64	2.584619322	384.2548187
50	2.05	2.609906854	407.2929138
100	2.33	2.627176388	423.8150629

Source: Calculation Result

As the return period increases, LOG XT and maximum rainfall increase. For example, for the 2-year Return Period, the rainfall is 304.42 mm, while for the 100-year Return Period it increases to 423.82 mm. The log normal distribution is used because it is more suitable for hydrological data that is often skewed to the right (positive skew).

4. **Log Person III Distribution Plan Rainfall Data**

Table 16. Log Person III Distribution Plan Rainfall

Rainfall Plan Period Year – Log Person III Method		
Return Period	KTR	Rainfall
2	-0.031	303.068
5	0.831	342.547
10	1.300	366.168
25	1.814	393.931
50	2.153	413.364
100	2.464	432.028

Source: Calculation Result

The table shows that the larger the rainfall return period, the higher the design rainfall, reflecting a higher probability of extreme rainfall. For example, the design rainfall for a 2-year rainfall return period is 303.07 mm, while for 100 years it reaches 432.03 mm. The Log Pearson III distribution is very useful for describing rainfall data that is not symmetrical and has positive skewness.

CONCLUSION

The conclusions of this research are:

1. Based on the Chi-Square test, all probability distribution models (Gumbel, Normal, Log-Normal, and Log Pearson Type III) are statistically acceptable at the 5% significance level, as the calculated chi-square values are less than the critical value (χ^2calc < χ^2crit). However, at the 1% significance level, only the Gumbel distribution meets the acceptance criteria, while the Normal, Log-Normal, and Log Pearson Type III distributions do not. Furthermore, the Kolmogorov–Smirnov test indicates that all four

distributions are acceptable at both significance levels, as the maximum deviation values (Δcalc) are below the critical thresholds.
2. From the design rainfall calculations for return periods of 2, 5, 10, 20, 50, and 100 years using the four distribution methods, it is found that for the 2-year return period, the Normal distribution yields the highest rainfall value of 307.23 mm. For the 5- to 100-year return periods, the Gumbel distribution consistently produces the highest values: 352.72 mm (5 years), 386.91 mm (10 years), 430.07 mm (20 years), 462.08 mm (50 years), and 493.86 mm (100 years). These results suggest that the Gumbel distribution is the most suitable model for representing extreme rainfall events in the study area.

ACKNOWLEDGEMENTS

Author would like to express our deepest gratitude to all parties who have assisted in completing the research entitled " Analysis of The Suitability of Frequency Distribution of Rainfall Data and Rainfall Return Period at Pt. X, Kutai Kartanegara Regency, East Kalimantan ". We would also like to thank our colleagues who have provided direction and input that have been very helpful in the process of this research. This research would not have been possible without the cooperation of all parties. Thank you.

REFERENCES

[1] F. J. Acero, S. Parey, J. A. García, and D. Dacunha-Castelle, "Return level estimation of extreme rainfall over the Iberian Peninsula: Comparison of methods," *Water (Switzerland)*, vol. 10, no. 2, 2018, doi: 10.3390/w10020179.
[2] J. GhoshDastider, D. Pal, and P. K. Mishra, "Evidence of Kolmogorov like scalings and multifractality in the rainfall events," pp. 1–20, 2024, [Online]. Available: http://arxiv.org/abs/2405.03463
[3] A. Akcil and S. Koldas, "Acid Mine Drainage (AMD): causes, treatment and case studies," *J. Clean. Prod.*, vol. 14, no. 12, pp. 1139–1145, 2006, doi: https://doi.org/10.1016/j.jclepro.2004.09.006.
[4] P. L. Younger, S. A. Banwart, and R. S. Hedin, "Mine Water Hydrology BT - Mine Water: Hydrology, Pollution, Remediation," P. L. Younger, S. A. Banwart, and R. S. Hedin, Eds. Dordrecht: Springer Netherlands, 2002, pp. 127–270. doi: 10.1007/978-94-010-0610-1_3.
[5] T. Kempka *et al.*, "Best-practice guidelines on Hybrid Pumped Hydropower Storage of excess energy in open-pit lignite mines," 2024.
[6] C. Asdak, Yulizar, and Subiyanto, "a National Policy on Indonesia'S Integrated Water Resource Conservation Management," *Indones. J. For. Res.*, vol. 10, no. 2, pp. 151–162, 2023, doi: 10.59465/ijfr.2023.10.2.151-162.
[7] R. Montes-Pajuelo, Á. M. Rodríguez-Pérez, R. López, and C. A. Rodríguez, "Analysis of Probability Distributions for Modelling Extreme Rainfall Events and Detecting Climate Change: Insights from Mathematical and Statistical Methods," *Mathematics*, vol. 12, no. 7, pp. 1–24, 2024, doi: 10.3390/math12071093.
[8] A. C. de S. Matos, F. E. O. E. Silva, and G. de O. Corrêa, "Estimating the parameters of a monthly hydrological model using hydrological signatures," *Rev. Bras. Recur. Hidricos*, vol. 29, pp. 1–11, 2024, doi: 10.1590/2318-0331.292420230121.
[9] A. Al Mamoon and A. Rahman, "Selection of the best fit probability distribution in rainfall frequency analysis for Qatar," *Nat. Hazards*, vol. 86, no. 1, pp. 281–296, 2017, doi: 10.1007/s11069-016-2687-0.
[10] M. Basinger, F. Montalto, and U. Lall, "A rainwater harvesting system reliability model based on nonparametric stochastic rainfall generator," *J. Hydrol.*, vol. 392, no. 3–4, pp. 105–118, Oct. 2010, doi: 10.1016/J.JHYDROL.2010.07.039.
[11] C. M. Côte, C. J. Moran, C. J. Hedemann, and C. Koch, "Systems modelling for effective mine water management," *Environ. Model. Softw.*, vol. 25, no. 12, pp. 1664–1671, Dec. 2010, doi: 10.1016/J.ENVSOFT.2010.06.012.
[12] M. T. Aditya, H. Susilo, and Y. Fanani, "Analysis of Decreased Limestone Production on the Effect of Rainfall with the Linear Regression Method," vol. 01, no. 10, pp. 1092–1101, 2023.
[13] T. A. Cahyadi, S. Magdalena, D. Haryanto, W. D. Ratminah, P. E. Rosadi, and P. E. Asmara, "Planning of mining water management costs," *AIP Conf. Proc.*, vol. 2245, no. 1, p. 90008, Jul. 2020, doi: 10.1063/5.0007076.
[14] A. Gires, I. Tchiguirinskaia, D. Schertzer, A. Schellart, A. Berne, and S. Lovejoy, "Influence of small scale rainfall variability on standard comparison tools between radar and rain gauge data," *Atmos. Res.*, vol. 138, pp. 125–138, Mar. 2014, doi: 10.1016/J.ATMOSRES.2013.11.008.
[15] A. L. A. Martins, G. R. Liska, L. A. Beijo, F. S. de Menezes, and M. Â. Cirillo, "Generalized Pareto distribution applied to the analysis of maximum rainfall events in Uruguaiana, RS, Brazil," *SN Appl. Sci.*, vol. 2, no. 9, p. 1479, 2020, doi: 10.1007/s42452-020-03199-8.
[16] A. Kumar and M. Saharia, "Exploratory Analysis of Hydrological Data BT - Python for Water and Environment," A. Kumar and M. Saharia, Eds. Singapore: Springer Nature Singapore, 2024, pp. 23–41. doi: 10.1007/978-981-99-9408-3_4.
[17] R. Diwakar, "An evaluation of normal versus lognormal distribution in data description and empirical analysis," *Pract. Assessment, Res. Eval.*, vol. 22, no. 13, pp. 1–15, 2017.
[18] C. G. Anghel, "Revisiting the Use of the Gumbel Distribution: A Comprehensive Statistical Analysis Regarding

Modeling Extremes and Rare Events," *Mathematics*, vol. 12, no. 16, 2024, doi: 10.3390/math12162466.
[19] C.-G. Anghel and D. Ianculescu, "An In-Depth Statistical Analysis of the Pearson Type III Distribution Behavior in Modeling Extreme and Rare Events," *Water*, vol. 17, no. 10, 2025, doi: 10.3390/w17101539.
[20] B. Oseni and F. Ayoola, "Fitting the Statistical Distribution for Daily Rainfall in Ibadan, Based on Chi-Square and Kolmogorov-Smirnov Goodness-of-Fit," *Eur. J. Bus. …*, vol. 4, no. 17, pp. 62–70, 2012, [Online]. Available: http://www.iiste.org/Journals/index.php/EJBM/article/view/3202
[21] F. Soehardi and M. Dinata, "Recent analysis of maximum rain period," *Int. J. Eng. Technol.*, vol. 7, no. September, pp. 63–67, 2018, doi: 10.14419/ijet.v7i2.3.12323.

BIOGRAPHIES OF AUTHORS

Muhammad Tri Aditya,S.T.,M.T Mining Engineering, Lecturer at Malang State Polytechnic, Mining Technology Study Program muhammadtri@polinema.ac.id