Effect of Data Length to the Consistency of Design Rainfall

I G Tunas* and G M Oka

1Department of Civil Engineering, Universitas Tadulako, Jalan Soekarno-Hatta Km. 9 Palu, Central Sulawesi, 94117, Indonesia

*Email: tunasw@yahoo.com

Abstract. Design rainfall is one of the hydrology quantities that can be used for analyzing design flood. This discharge can then be applied to design and manage civil engineering structures such as drainage channels in urban areas. Generally design rainfall is analyzed using maximum daily rainfall data with a minimum data length of 10 years obtained from a rainfall observation station. In various references, the use of rainfall data with a longer range is highly recommended. This paper intends to examine the use of variations in the length of rainfall data on the consistency of design rainfall. Rainfall data from the Mutiara Meteorological Station of Palu, Central Sulawesi (1990-2019) are grouped into five categories based on the length of the data: A (10 years), B (15 years), C (20 years), D (25 years) and E (30 years). Frequency analysis was performed to predict the design rainfall of each data group using four distribution methods: Normal, Normal Log, Gumbel and Log Pearson III. Design rainfall consistency is measured based on design rainfall deviations which are calculated using the longest data. The analysis shows that the consistency of design rainfall in the study area at various return periods for all statistical distribution tends to increase in proportion to the length of the data with high category. This consistency value indicates that the 10-year rainfall data is still sufficient for frequency analysis.

1. Introduction

Frequency analysis plays an important role in hydrological analysis to determine the design magnitude, one of which is design rainfall [1, 2, 3, 4, 5]. Determination of this parameter is generally used to predict design flood as a basis for planning and management of civil engineering construction, especially hydraulic structures, such as water use and flood control structures [6, 7, 8, 9, 10]. The dimensions and capacity of the buildings, in addition to being determined by the service volume, also depend on the design flood that corresponds to the selected return period.

Design rainfall as a frequency analysis output is generally determined based on the maximum daily rainfall data representing each data year [11, 12, 13]. This method is known as the Annual Maximum Series (AMS) by selecting only the largest rainfall data each year [14, 15, 16]. The amount of rainfall data for this analysis is generally recommended for more than 10 years. The minimum number of data relates to parametric statistical requirements that must be met in statistical analysis, especially confidence intervals and error margins.

Some researchers both in Indonesia and abroad have carried out many researches related to frequency analysis [11, 12, 13, 14, 16, 17]. Frequency analysis is more focused on rainfall data than on discharge data due to the still limited discharge data in a number of watersheds in Indonesia. Several frequency distribution methods such as Normal, Normal Log, Gumbel and Log Pearson III have been tested for the purpose of the analysis. Most daily rainfall data in Indonesia is consistent with the Normal and Log Pearson III Distribution.
Frequency analysis of rainfall data with variations in length of data has not been found in published references. The only paper reviewing this topic is as reported by Huddiankuwera (2016) regarding design rainfall deviations due to the use of different lengths of data in the Tabo-Tabo Watershed, Sulawesi, Indonesia [17]. Design rainfall deviations decrease with increasing data length. The study does not present in detail the deviations produced but is presented in a regression curve. In addition, the distribution methods tested for deviation analysis are limited to Log Pearson III.

The content in this paper will be very important because besides testing the four statistical distribution methods, the design rain deviations from each method are presented as it is. This presentation will show that the deviation curve does not actually produce a smooth curve as the results of Huddiankuwera's research [17]. Therefore, the results of this study are expected to be an additional reference on the use of data length for frequency analysis.

2. Material and Methods

2.1. Data
Rainfall data from Mutiara Meteorology Station, Palu, Sulawesi, Central can be downloaded for free from http://dataonline.bmkg.go.id/data_iklim. This meteorological station is located in Palu City at geographical coordinates 119°54'19.94"E, 0°54'57.20"S. This station was established by the Meteorology, Climatology, and Geophysical Agency of Indonesia primarily for navigation and weather forecasting purposes. Rainfall data recorded at this station is daily rain and it will be applied to this study for the period 1990-2019.

Located in the tropics, the study location has a very typical annual rainfall of less than 1000 mm/year. In contrast to other buffer zones such as Sigi, Donggala and Parigi, annual rainfall in the city of Palu is classified as the lowest with the number of rainfall days very volatile. This might be related to the location of the area which is located in the valley area which is directly bordered by mountains on the west and east sides and beaches on the north [18, 19, 20].

2.2. Methods
The study was initiated with daily rainfall data selection using the Annual Maximum Series (AMS) method. The maximum daily rainfall data was selected for 30 data and then grouped into 5 categories based on data length: A (10 years), B (15 years), C (20 years), D (25 years) and E (30 years). The five data groups are the basic data to be evaluated.

Frequency analysis was applied to each data group using four frequency distribution methods: Normal, Normal Log, Gumbel and Log Person III. Equations of the four methods are expressed by:

\[ X_T = \bar{X} + K_T S \]  

(1)

where \( X_T \) is design rainfall with a certain return period (mm), \( \bar{X} \) is average maximum daily rainfall (mm), \( K_T \) is frequency factor, and \( S \) is standard deviation. Rainfall data in the Log Normal and Log Pearson III Distributions are expressed in logarithmic form. The frequency factors of the four frequency distribution methods can be obtained from the frequency factor tables of each distribution. The frequency factor is closely related to the chosen return period [1].

The consistency of design rainfall as the output of frequency analysis is calculated by the number of deviations using the following equation:

\[ D = \text{abs}\left(\frac{R_{TE} - R_{TR}}{R_{TR}}\right) \times 100\% \]  

(2)

\( D \) = deviation of design rainfall (%), \( R_{TE} \) = design rainfall with a certain return period (mm), and \( R_{TR} \) = design rainfall predicted using longest data length (mm). The consistency of design rain can be interpreted from the number of deviations with a predicate: high (\( D < 20\% \)), moderate (\( 40\% \leq D \leq 60\% \)) and low (\( D > 60\% \)).
3. Result and Discussion

3.1. Selecting and Grouping Rainfall Data
As previously informed, the maximum daily rainfall data selected was 30 data in the 1990-2019 period. The data are grouped into five categories as illustrated in Figure 1. In this range, the selected rainfall data ranged from 17.7 mm to 115 mm with an average value of 53.97 mm. Actually the maximum daily rainfall for 30 years in this area shows a fairly low variability rate. If extreme rainfall data (> 100 m) and very low rainfall data (<20 mm) are excluded from the data group, then the maximum daily rainfall data includes very low fluctuations.

If observed in more detail (Figure 1), fluctuations in rainfall data at the study location tend to follow the sine or cosine curve. The curve will return to a stationary point every 10 or 11 years. This shows that the maximum daily rainfall during the observation period shows a regular pattern with slow change. However, to further strengthen this analysis, daily data information is very important to know.

![Figure 1](image)

**Figure 1.** Daily maximum rainfall in the study area: A (10 years), B (15 years), C (20 years), D (25 years) and E (30 years)

3.2. Design Rainfall
Design rainfall determined by frequency analysis has been applied to the five data groups A, B, C, D and E. The selected return period is 1 year to 1000 years. The selection of a long return period is intended to see the pattern of consistency in the design rainfall. In general, design rainfall in the four distribution methods shows high uniformity in the five data groups, although in data groups A and B, design rainfall is relatively different compared to the other three data groups (Figure 2). This might be due to the fact that the maximum and minimum rainfall data has not been accommodated in the two data groups. Based on Figure 2, the design rainfall in the return period of 1 year to 1000 years ranges from 10 mm to 160 mm with the lowest prediction in the Normal Method. In contrast, the Normal Log Method shows the highest prediction results.
Figure 2. Design rainfall based on data length (a) Normal, (b) Log Normal, (c) Gumbel, and (d) Log Pearson III

3.3. Consistency
Consistency of prediction results shows the level of similarity of an event. This can be analogous to the predicted results. The consistency level of the predicted results shows the level of goodness of a model, formula or equation used. In this paper, the consistency of prediction results using four distribution methods based on five data groups is presented in Table 1. The level of consistency is inversely proportional to the number of deviations. Consistency increases with with decreasing deviation.

| Return Period (year) | Deviation (%) | Normal | Log Normal | Gumbel | Log Pearson III |
|----------------------|---------------|--------|------------|--------|-----------------|
|                      |               | 10     | 15         | 20     | 25              | 10    | 15    | 20    | 25    | 10    | 15    | 20    | 25    |
| 1                    |               | 0.4    | 5.6        | 3.8    | 12.6            | 0.2   | 5.6   | 1.2   | 1.4   | 6.1   | 4.2   | 3.9   | 13.2  | 0.6   | 6.1   | 0.6   |
| 2                    |               | 4.6    | 7.9        | 0.3    | 4.1             | 5.4   | 6.9   | 1.3   | 3.7   | 4.3   | 7.8   | 0.7   | 4.1   | 1.1   | 12.0  | 0.3   | 5.7   |
| 5                    |               | 5.7    | 8.5        | 1.0    | 4.2             | 0.3   | 11.5  | 1.7   | 5.4   | 5.6   | 8.5   | 0.9   | 4.2   | 1.3   | 11.1  | 1.8   | 5.5   |
| 10                   |               | 6.1    | 8.8        | 1.5    | 4.3             | 2.4   | 14.0  | 3.3   | 6.3   | 6.1   | 8.8   | 1.5   | 4.3   | 5.4   | 8.2   | 1.9   | 4.4   |
| 20                   |               | 6.4    | 8.9        | 1.8    | 4.3             | 4.8   | 16.0  | 4.7   | 7.1   | 6.5   | 9.0   | 2.0   | 4.3   | 9.7   | 4.9   | 1.6   | 3.0   |
| 50                   |               | 6.6    | 9.0        | 2.1    | 4.3             | 7.5   | 18.4  | 6.2   | 7.9   | 6.9   | 9.2   | 2.5   | 4.3   | 15.3  | 0.4   | 0.8   | 1.1   |
| 100                  |               | 6.8    | 9.1        | 2.3    | 4.3             | 9.3   | 20.1  | 7.2   | 8.5   | 7.1   | 9.3   | 2.7   | 4.3   | 19.2  | 3.0   | 0.2   | 0.3   |
| 500                  |               | 7.0    | 9.3        | 2.6    | 4.3             | 13.1  | 23.4  | 9.4   | 9.6   | 7.5   | 9.5   | 3.2   | 4.4   | 27.1  | 10.1  | 1.6   | 3.7   |
| 1000                 |               | 7.1    | 9.3        | 2.7    | 4.3             | 14.6  | 24.7  | 10.2  | 10.1  | 7.6   | 9.6   | 3.4   | 4.4   | 30.0  | 12.9  | 2.4   | 5.1   |

As shown in Table 1, design rainfall deviations in the four distribution methods tested ranged from 0.2% to 30%. The largest and smallest deviations occur in the Log Pearson III Distribution at different data lengths. But in general the average deviations in the entire length of the data and the method of
distribution are classified as low with average deviations ranging from 2.2% to 15% (Table 1). This shows that the level of consistency of prediction results can be categorized as high.

If compared with the results of the Huddiankuwera’s study (2016) [17], the results of this study tend to correspond where deviations decrease with increasing data length or increase in consistency in proportion to increasing data (Figure 3). This shows that the use of data length of at least 10 years still shows good performance. However, the results of this study must be tested with rainfall data in other with different characteristics to generalize conclusions.

4. Conclusion
Five groups of maximum daily rainfall data: A (10 years), B (15 years), C (20 years), D (25 years) and E (30 years) obtained from Mutiara Meteorology Station of Palu, Central Sulawesi, Indonesia have been analyzed. Four statistical distribution methods: Normal, The Normal Log, Gumbel and Pearson III Logs have been applied to determine the design rainfall based on the five groups of rain data. The consistency of design rainfall is assessed using deviations to design rainfall predicted from the longest rainfall data.

The result of analysis indicates that the consistency of design rainfall in the study area at various return periods for all statistical distribution tends to increase in proportion to the length of the data with high category. This consistency value indicates that the 10-year rainfall data is still sufficient for frequency analysis.

Acknowledgements
The authors acknowledge the Mutiara Meteorological Station of Palu, Meteorology, Climatology, and Geophysical Agency (BMKG) for supporting rainfall data

References
[1] Brotowiryatmo S H 2000 Hydrology: Teory, Problem and Solving (Yogyakarta: Nafiri Offset)
[2] Aurdin Y 2019 J. Tekno Global 8(1) 35–39
[3] Ginting S and Putuhena W M A 2016 J. Sumber Daya Air 12(1) 1–16
[4] Handayani Y L, Hendri A, and Aditya A 2013 J. Teknik Sipil 12(3) 222–232
[5] Deraman W H A W, Mutalib N J A M, and Mukhtar N Z 2017 IOP Conf. Series: J. Phys 890 012162
[6] Bhakar S R, Bansal A K, Chhajed N and Purohit R C ARPN J. Eng. App. Sci 1 64–67
[7] Chowdhury J U and Stedinger J R 1991 J. Hyd. Eng 117 811–831
[8] Read L K, and Vogel R M 2015 Water Resour. Res 51 638–6398
[9] Faridzad M, Yang T, Hsu K, Sorooshian S and Xiao C 2018 J. Hydrol 563 123–142
[10] Madsen H, Gregersen I B, Rosbjerg D and Arnbjerg-Nielsen K 2017 Water Sci. Technol 75(8)
1971–1981

[11] Al-Houri Z, Al-Omari A and Saleh O 2014 Civil Enviro. Res 6(3) 44–57
[12] Saghafian B, Golian S and Ghasemi A 2014 Nat. Hazards 71 403–417
[13] Arulmozhi S, Arulraj and Prince G 2017 Ind. J. Geo-Marine Sci 46(1) 213–218
[14] Das S 2017 Int. J. Climatol 37 612–623
[15] Tung Y and Wong C 2014 Stoch. Environ. Res. Risk. Assess 28 583–592
[16] Vivekanandan N 2014 Int. J. Civil Eng 1(4) 6–12
[17] Huddiankuwera A 2016 J. Ilmiah Teknik & Informatika 1(2) 36–41
[18] Tunas I G 2019 J. Ecol. Eng 20(7) 112–125
[19] Tunas I G, Tanga A and Oktavia S 2020 J. Ecol. Eng 21(2) 190–200
[20] Tunas I G and Maadj R 2018 Int. J. Adv. Sci. Eng. Inform. Technol 8(6) 2413–2420