Effect of Missing Rainfall Data on the Uncertainty of Design Floods

N Nurhamidah¹, S Hanwar², A Junaidi¹, Februarman¹, Sunaryo¹
¹Department of Civil Engineering, Faculty of Engineering, University of Andalas, Padang, Indonesia
²Department of Civil Engineering, Padang State Polytechnics, Padang, West Sumatra, Indonesia

Corresponding author: nurhamidah@eng.unand.ac.id

Abstract. Hydrological data is a collection of information or facts about the hydrological phenomenon. Mistakes in the hydrological analysis have a fatal effect on the accuracy or accuracy of hydraulic building planning, whether in the form of planning for irrigation buildings, flood control structures, drainage systems, weirs, and others hydraulic structure design. As we already know, rainfall data is the most important data in the hydrological analysis. The minimum amount of rainfall data required in the hydrological statistical analysis is 10 years. To obtain this 10-year data, often encountered incompleteness in recording the rainfall data. Meanwhile, hydrological analysis is always required in any water construction planning.

1. Introduction
Mistakes in the hydrological analysis have a fatal effect on the accuracy or accuracy of hydraulic building planning, in the form of planning for irrigation buildings, flood control structures, drainage systems, weirs, others design of hydraulics structures. As we already know, rainfall data is the most important data in the hydrological analysis. The minimum amount of rainfall data needed in hydrological statistical analysis is 10 years [1]. To obtain this 10-year data, incompleteness is often encountered in recording the rainfall data. Meanwhile, hydrological analysis is always required in any water construction planning.
Due to data limitations and incomplete data availability, it is necessary to research to analyze how much influence the completeness of the data and the availability of rainfall data in an area on the accuracy of the data to predict the planned discharge in different return periods and to make a Rating Curve of the Kuranji Batang cross-section.

Rainfall data recording at rain stations are not always complete. Then it is necessary to estimate the lost rainfall data. This estimate can be calculated using the inverse square distance method. The following is the formula for the inverse fair distance method.

\[ R_A = \frac{1}{(d_{AB})^2} R_B + \frac{1}{(d_{AC})^2} R_C \]

Which is:
- \( R_A \) = Rainfall data at station A (mm)
- \( R_B \) = Rainfall data at station B (mm)
- \( R_C \) = Rainfall data at station C (mm)
- \( d_{AB} \) = Distance from A to B (km)
- \( d_{AC} \) = Distance from A to C (km)

The average rainfall can be calculated using several methods:

1.1. Algebra Average Method

According to [2,3], this method is the simplest one and is obtained by calculating the arithmetic average of all total rain gauges in all areas. This method can be used in areas that are relatively flat and have a large number of rain stations, assuming that the rainfall in the watershed is even. The following equation can formulate the algebraic mean method:

\[ \bar{R}_{\text{average}} = \frac{R_1 + R_2 + R_3 + \cdots + R_n}{n} \]

- \( \bar{R}_{\text{average}} \) = Average rainfall (mm)
- \( n \) = Number of station
- \( R_1, R_2, \ldots, R_n \) = Rainfall data at point 1, 2, 3, ..., n (mm)

1.2. Polygon Thiessen Method

The Thiessen Polygon method is suitable for areas with uneven precipitation measurements requiring observation stations in and near the site, and this method does not take topography into account. The following equation can formulate the Thiessen Polygon method:

\[ \bar{R}_{\text{average}} = \frac{A_1 \times R_1 + A_2 \times R_2 + \cdots + A_n \times R_n}{A_1 + A_2 + \cdots + A_n} \]

After calculating the average rainfall with both methods. Furthermore, error checking is carried out on each process to determine which way is the best. Error checking can be calculated using the following equation:

\[ \text{Error} = \frac{\sum_{i=1}^{n} |\bar{R} - R_i|}{\bar{R}} \]

- \( R_i \) = Rainfall data at \( i^{th} \) station
- \( \bar{R} \) = Average rainfall (mm)
- \( n \) = Number of station

1.3. Design Rainfall

To determine the designed rainfall can be calculated by several methods, i.e. the Normal distribution method, Log-Normal distribution, Gumbel distribution, and Log Pearson Type III distribution. In this study, the calculated rainfall design was carried out using four methods. After that, a dispersion test is
carried out to select an approach that meets the requirements. Where the calculated values that meet the criteria can be seen in table 1 below:

| Distribution Methods | Requirement | Distribution Methods | Requirement |
|----------------------|-------------|----------------------|-------------|
| Normal               | $Cs = 0$    | Log Pearson Type III | $Cs = 0$   |
|                      | $Ck = 3$    |                      |             |
| Gumbel               | $Cs \leq 1,1396$ | Log-Normal           | $Cs = 3Cv + Cv^2$ |
|                      | $Ck \leq 5,4002$ |                      | $Ck = 5,383$ |

In this research, the value obtained meets the boundary conditions is the Gumbel distribution method. Then the results of the calculated rainfall plan used for further calculations are the results of the computed rainfall plan using the Gumbel distribution method.

Furthermore, the distribution harmony test was carried out using the Chi-square method, where the Chi-Square test is used to test the vertical deviation [4]. In the calculation of the unit hydrograph, the method used is the Nakayasu method. The following are the steps for calculating the unit hydrograph using the Nakayasu method [5].

2. Materials and methods
Rainfall data collection was obtained by accessing data from BMKG and river sections taken from the journal Flood Control Analysis in the Batang Kuranji River Basin [6]. The location and the cross-section of the Kuranji river can be seen in figure 1.

The rainfall data was taken from the recording of three rain stations, which is located at Gunung Nago station, Batu Busuk station, and Ladang Padi station. The three rain stations are the closest stations to the observed river basin. The maximum daily rainfall data used is the maximum daily rainfall data per month.

In this project, the rainfall data is divided into three conditions. The first condition is the rainfall data for 20 years from 1994-2013, which is equipped with an estimated rainfall loss. The second condition is the rainfall data which is also complemented with the estimated loss of rain, namely from 1994-2004, which consists of 11 data. The third condition, rainfall data regardless of the lost rainfall, namely from 1996-2004, 2006-2007, and 2009-2013 which consisted of 16 data.

According to the literature review that has been carried out, the hydrological analysis that needs to be done in relation to estimating lost rainfall data is as follows:

1. Determine the average rainfall using the Average Algebra method and the Thiessen Polygon method.
2. Rainfall analysis plan using four methods, namely the Normal distribution method, Log-Normal distribution, Log Pearson Type III distribution, and the Gumbell distribution method.
3. Calculation of the debit plan using the Rational Method.
4. Determine the unit hydrograph using the Nakayasu method.
5. Determine the percentage difference in discharge accuracy.
6. Creating a Rating Curve.
3. Results and discussion

3.1 Rainfall Frequency

The calculation of the frequency of rainfall is used to determine the factors required in determining the type of rainfall distribution at the study location. In this calculation, the value of Standard Deviation (S), Skewness Coefficient (Cs), Kurtosis Coefficient (Ck), and Variation Coefficient (Cv) will be obtained that meet the requirements of each distribution.

In selecting the distribution for the first, second, and third conditions, the value that meets the requirements is the Gumbel distribution. Then the planned rainfall data used to determine the planned discharge is the result of calculations using the Gumbel distribution method. Next, the distribution harmony test was carried out using the Chi-square method for the first condition. The results of the conformity test calculation can be seen in table 2.

Table 2. Chi-square.

| Interval P     | Oi | Ei | (Oi-Ei) | (Oi-Ei)^2 | (Oi-Ei)^2/Ei |
|---------------|----|----|---------|-----------|-------------|
| P<119.613     | 5  | 4  | 1       | 1         | 0.25        |
| 119.613<P<139.627 | 7  | 4  | 3       | 9         | 2.25        |
| 139.627<P<159.640 | 3  | 4  | -1      | 1         | 0.25        |
| 159.640<P<179.653 | 1  | 4  | -3      | 9         | 2.25        |
| 179.653<P<199.667 | 4  | 4  | 0       | 0         | 0           |
| Total         | 20 | 20 | 0       | 20        | 5           |

Figure 1. The study area and river cross-section.
The conformity test is acceptable if the value \((O_i - E_i)^2 / E_i \leq X^2\). From the above calculation, it is obtained \((5 \leq 7,815)\), then the Gumbel distribution used is acceptable. For the second condition, the suitability test is obtained \((4.64 \leq 5.991)\), then the Gumbel distribution used is acceptable. For the third condition, the conformity test is obtained \((4 \leq 7,815)\), then the Gumbel distribution used is acceptable.

In other calculations using the maximum daily volume per year, a more safety figure is obtained. Because the results obtained, that is a greater value than the calculation using the maximum daily rainfall per month. However, in this final project, the discharge value is obtained from a focused rainfall calculation using the maximum monthly rainfall data. In the discharge calculation with the two calculations, the percentage difference for each condition is obtained, namely 5% -19% for the first condition, 4% -9% for the second condition, and 11% -14% for the third condition in the return period of 2 years to 200 years.

### 3.2 Rating Curve

Rating Curve is a graph that shows the relationship between discharge data in the cross-section of the Kuranji river and water level.

![RATING CURVE](Image)

Figure 2. Rating Curve.

Which can be describe for 3 conditons, such as:

- **I** = 1st condition, data was filled by 20 years (1994-2013)
- **II** = 2nd condition, data was filled by 11 years. (1994-2004)
- **III** = 3rd condition, missing rainfall data was neglected and only 16 data available from 1994-2013

The rating curve above is obtained from the calculation of the cross-sectional discharge of Batang Kuranji with a height value per 0.5 meter.
Table 3. Discharge and water level relationships measurement.

| h (m) | Q (m³/s) | h (m) | Q (m³/s) |
|-------|----------|-------|----------|
| 0     | 0        | 5     | 334.121  |
| 0.5   | 7.143    | 5.5   | 413.190  |
| 1     | 24.676   | 6     | 508.444  |
| 1.5   | 51.483   | 6.5   | 620.605  |
| 2     | 86.945   | 7     | 750.426  |
| 2.5   | 130.555  | 7.5   | 898.673  |
| 2.89  | 169.948  | 8     | 1066.116 |
| 3     | 170.981  | 8.5   | 1253.522 |
| 3.5   | 188.491  | 9     | 1461.653 |
| 4     | 222.083  | 9.5   | 1691.262 |
| 4.5   | 270.585  | 10    | 1943.097 |

From the rating curve, (figure 2). So from the return discharge time of 2, 5, 10, 25, 50, 100, and 200, the water level can be determined.

Table 4. Discharge and water level relationships calculation for 3 conditions.

| Return period (T) | Condition 1 | Condition 2 | Condition 3 |
|-------------------|-------------|-------------|-------------|
|                   | Q (m³/s)    | h (m)       | Q (m³/s)    | h (m)       | Q (m³/s)    | h (m)       |
| T2                | 414.404     | 4.806       | 437.669     | 4.980       | 405.350     | 4.737       |
| T5                | 518.111     | 5.566       | 543.734     | 5.748       | 515.420     | 5.547       |
| T10               | 586.774     | 6.046       | 613.958     | 6.231       | 588.296     | 6.057       |
| T25               | 673.530     | 6.626       | 702.686     | 6.814       | 680.375     | 6.670       |
| T50               | 737.890     | 7.036       | 768.510     | 7.226       | 748.684     | 7.103       |
| T100              | 801.775     | 7.428       | 833.848     | 7.617       | 816.489     | 7.515       |
| T200              | 865.427     | 7.801       | 898.947     | 7.991       | 884.046     | 7.907       |

According to the calculation mentioned above, it shows that the discharge for the first condition is 801,775 m³ / s, for the second condition it is 833,848 m³ / s, and for the third condition it is 816,489 m³ / s. In determining the difference in discharge accuracy obtained, it is reviewed against the second condition because the second condition has more prevalent rainfall data than the first and third conditions. Where the difference inaccuracy in the first condition is 3.846% against the second condition, and the difference inaccuracy in the third condition is 2.082% for the second condition.

4. Conclusions
Based on the description in the analysis and discussion, it can be concluded that the effect of the availability of rainfall data on the planned discharge is not very influential. From the unit hydrograph curve, it can be seen that in the first condition with a return period of 100 years, a discharge of 801.775 m³/s is obtained, for the second condition consisting of 11 rainfall data, a discharge of 833.848 m³/s is obtained, and for the third condition by ignoring the rain data that is lost, it is obtained a debit of 816.489 m³/s. Where the difference inaccuracy in the first condition is 3.846% against the second condition, and the difference inaccuracy in the third condition is 2.082% for the second condition. In the rating curve section of Batang Kuranji, the normal water level discharge is 169.948
m³/s with an elevation of 2.89 m. In comparison, the maximum discharge at the cross-section of Batang Kuranji is 748,260 m³/s with an elevation of 6.99 m.

Acknowledgement
Publication of this article is supported by the Faculty of Engineering, University of Andalas, and we would like to thank all those who have involved in this research.

References
[1]. Canchala-Nastar, T., Carvajal-Escobar, Y., Alfonso-Morales, W., Cerón, W. L., & Caicedo, E. 2019 Estimation of missing data of monthly rainfall in southwestern Colombia using artificial neural networks Data in brief 26 104517.
[2]. Hasyimzoem, E.F. 2019 Perbandingan analisis data curah hujan yang hilang menggunakan metode Reciprocal, Normal Ratio, dan Rata-rata Aljabar (Studi kasus: Curah Hujan Beberapa Stasiun Hujan Daerah Pringsewu). [In Indonesian]
[3]. Chowdhury, M. A. I., Kabir, M. M., Sayed, A. F., & Hossain, S. 2016 Estimation of rainfall patterns in Bangladesh using different computational methods (arithmetic average, Thiessen polygon and isohyet) J. Biodivers. Environ. Sci 8 43-51.
[4]. Arvind, G., Kumar, P. A., Karthi, S. G., & Suribabu, C. R. 2017 Statistical Analysis of 30 Years Rainfall Data: A Case Study In IOP Conference Series: Earth and Environmental Science (Vol. 80, No. 1, p. 012067).
[5]. Idfi, G., Yulistyorini, A., & Suryoputro, N. 2020 The comparative study of peak discharge at Ngotok watershed by using the method of SCS, Snyder and Nakayasu for flood control needs In IOP Conference Series: Materials Science and Engineering (Vol. 930, No. 1, p. 012075). IOP Publishing.
[6]. Syofyan, Z. 2017 Analisa pengendalian banjir pada Daerah Aliran Sungai Batu Busuk (Batang Kuranji) Kota Padang Jurnal Teknik Sipil ITP, ISSN 2354-8452 E-ISSN 2614-414X, 3(2). [In Indonesian]