

# Correction and Validation of Daily Rainfall Satellite Data in the Wae Mese Watershed, Labuan Bajo, West Manggarai Regency

Pankrasio Mario Yudarsi<sup>1</sup>, Albert Wicaksono<sup>2</sup>, Obaja Triputera Wijaya<sup>3</sup>  
<sup>1,2,3</sup>Universitas Katolik Parahyangan, Bandung, Indonesia  
Email: [aripankrazio@gmail.com](mailto:aripankrazio@gmail.com)

## Abstract

This study aims to analyze rainfall correction in the Wae Mese watershed, located in Labuan Bajo, focusing on improving the accuracy of rainfall data for flood modelling and watershed management planning. In this research, corrections are performed using an objective function that compares observed rainfall data with satellite-based rainfall data from GPM (Global Precipitation Measurement). The objective function is used to evaluate prediction errors and determine the most effective correction factor to enhance the reliability of satellite rainfall data. The research methodology includes the application of the objective function to measure differences between satellite-derived rainfall data and observed data, and applying corrections based on the analysis results. The findings show that rainfall data corrected using the objective function provides a more accurate representation of rainfall events, leading to more precise flood predictions for the watershed. This study highlights the importance of using the objective function in rainfall correction, which plays a key role in effective flood risk management and planning. These findings provide a better understanding of rainfall patterns in the Wae Mese watershed and support the use of corrected satellite-based rainfall data for hydrological modelling and future flood risk assessments.

**Keywords:** *Rainfall, GPM, Satellite Data, Objective Function, Wae Mese.*



## A. INTRODUCTION

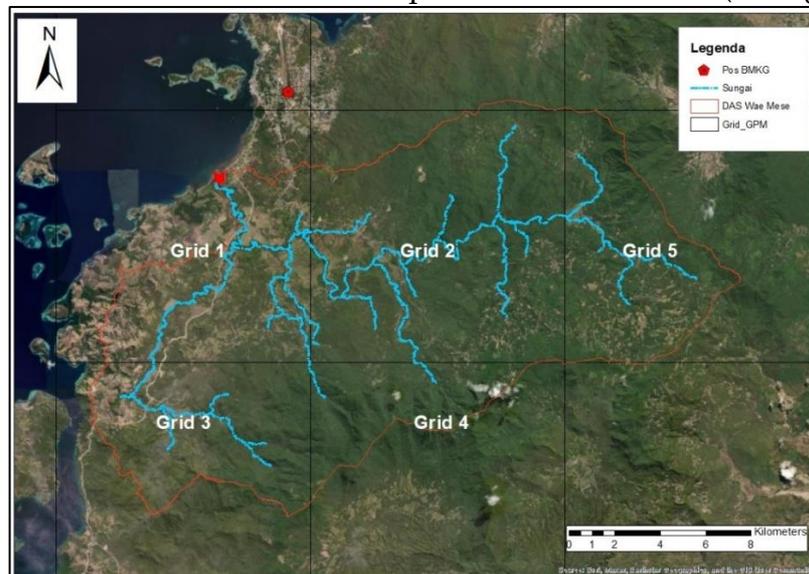
Rainfall data is a primary data source that can be transformed into streamflow discharge as a substitute for measured discharge data in hydrological analysis (Made et al., 2021). Therefore, the limitation of measured rainfall data poses a significant obstacle. In Indonesia, rainfall data, in terms of quantity, hourly intervals, and short durations, is still scarce (PUPR, 2022). The satellite data used to obtain rainfall is the GPM (Global Precipitation Measurement) satellite data, which utilizes simulated or weather-predicted rainfall data. The use of rainfall data obtained from satellites facilitates the analysis of long-term rainfall data for a river or watershed (Harsanto et al., 2022)

Various satellite sources, such as NASA's Tropical Rainfall Measuring Mission (TRMM) and the Global Precipitation Measurement (GPM), enable the monitoring of rainfall patterns worldwide with high resolution. TRMM was launched in 1997 with the aim of monitoring tropical rainfall and measuring precipitation in greater detail. TRMM provides rainfall data with a spatial resolution of  $0.25^\circ \times 0.25^\circ$ , or approximately  $28 \times 20$  km, and a temporal resolution ranging from daily rainfall to every 3 hours. TRMM officially ended on April 15, 2015, and its data processing concluded on December 31, 2019. Following the success of TRMM, the Global

Precipitation Measurement (GPM) was developed to provide space-based precipitation measurements to replace TRMM. GPM is a collaborative mission between NASA and JAXA and delivers improved products compared to TRMM. GPM has a spatial resolution of  $0.1^\circ \times 0.1^\circ$  and a superior temporal resolution, capturing data every 30 minutes. Although GPM and TRMM satellite data are increasingly accessible, and GPM data has better accuracy compared to rain gauge stations, these satellite data still need to be corrected to obtain more accurate data (Bolvin et al., 2020).

East Nusa Tenggara Province is considered an arid region where the dry season lasts longer than the wet season. The relatively short rainy season, combined with mountainous topography and sparse vegetation, results in low rainfall (an average of 1000 mm/year). This rainfall primarily manifests as surface water, accumulating in rivers, leading to flooding, and ultimately flowing into the sea (Welkis et al., 2021; Bees et al., 2021). As a region with low rainfall, rainfall data in East Nusa Tenggara is often incomplete. The challenges and issues still faced include limited availability of observational rainfall data, both spatially and temporally, insufficient and incomplete long-term rainfall time series data, an uneven distribution of rainfall stations, a lack of observation personnel, a manual data collection and input system, slow data transmission from certain regions to central levels, and non-standardized data formats (Mamenun et al., 2014)

The Wae Mese Watershed is one of the watersheds located in Labuan Bajo, East Nusa Tenggara, with a catchment area of 305.59 km<sup>2</sup>, where there is only one rainfall measurement station located in the lower part of the watershed (see figure 1).



**Figure 1 Wae Mese Watershed**

The limited rainfall measurement data results in a very low spatial representation of rainfall data. Therefore, with current technological advancements, satellite data can be used to obtain rainfall data in areas where rain gauges are not installed, without the need for direct field measurements. This study was conducted to evaluate how satellite rainfall data can be used for flood discharge calculations in

the Wae Mese Watershed. This study is expected to provide solutions for selecting reliable TRMM satellite imagery data by utilizing ground station rainfall data.

## B. METHOD

This study begins by collecting daily rainfall data from the GPM satellite. The data sources used in this study are categorized as secondary data. The secondary data collected in this study includes rainfall data from 2001 to 2022, obtained through satellite imagery in the form of GPM IMERG data and daily rainfall data from the rain gauge in Labuan Bajo. The GPM IMERG data can be accessed from NASA's official website at [giovani.gsfc.nasa.gov/giovani/](http://giovani.gsfc.nasa.gov/giovani/), while the rainfall data from the station can be obtained through the relevant authorities, such as the Meteorology, Climatology, and Geophysics Agency (BMKG) of West Manggarai Regency, Labuan Bajo. Once the data is collected, the satellite data will be tested based on the objective functions, including RMSE (Root Mean Square Error), Coefficient Correlation, POD (Probability of Detection), and FAR (False Alarm Ratio) (Sanjaya et al., 2022).

RMSE is a commonly used metric to evaluate the accuracy of a model by measuring the average difference between the observed and predicted values. In the context of rainfall prediction, RMSE calculates the difference between the satellite-based rainfall data and the actual rainfall recorded by ground stations. A lower RMSE indicates that the predicted rainfall values are closer to the observed values, thus indicating a more accurate model. RMSE shows enhanced performance in detecting essential anomalies, and calculated using Equation (1)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (e_i)^2}{n}} \quad (1)$$

Where  $n$  is the number of data and  $e_i$  is the error at  $i$ . The false alarm ratio (FAR) is quantifying how often the model falsely predicts rainfall when none occurs, which is known as a false positive. It compares the number of false positives (incorrectly predicted rainfall events) to the total number of predicted rainfall events. A lower FAR value indicates that the model is good at avoiding false alarms and not predicting rainfall when it doesn't occur. This metric is important for evaluating the specificity of the model, as it shows how well the satellite data avoids erroneous predictions of rainfall events that are not supported by actual observations. FAR is calculated using equation (2)

$$FAR = \frac{N_{Falsealarm}}{N_{hits} + N_{Falsealarm}} \quad (2)$$

POD is a metric used to assess how well the model is at detecting true rainfall events. It compares the number of rainfall events correctly predicted by the satellite data (true positives) to the total number of actual rainfall events, which includes both true positives and false negatives (missed rainfall events). A higher POD value indicates that the satellite data is effective at detecting actual rainfall events. This metric is especially useful for understanding the model's sensitivity—how often it correctly identifies rainfall when it occurs. POD is calculated using Equation (3)

$$POD = \frac{N_{hits}}{N_{hits} + N_{misses}} \quad (3)$$

Where  $N_{hits}$  is the number of correctly predicted daily GPM precipitation,  $N_{Falsealarm}$  is the number of daily classifieds as prediction by GPM when no precipitation is detected at rainfall station, and  $N_{misses}$  is the number of daily precipitation undetected by GPM. The correlation coefficient measures the strength and direction of the linear relationship between two variables—in this case, the satellite rainfall data and the ground station data. It ranges from -1 to 1, where 1 indicates a perfect positive relationship, -1 indicates a perfect negative relationship, and 0 suggests no linear relationship. A higher positive correlation means that as the rainfall predicted by the satellite increases, the ground station measurements also tend to increase in a similar manner. This metric helps assess whether the satellite data follows similar patterns as the ground-based data, which is crucial for validating the use of satellite data in hydrological analyses. The correlation is calculated using Equation (4)

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \quad (4)$$

Together, these metrics provide a comprehensive evaluation of satellite rainfall data's accuracy and performance in comparison to ground station data. They offer insights into the model's ability to detect, predict, and correlate rainfall events while minimizing errors, making them essential tools for hydrological modelling and flood prediction. For daily GPM correction (Senjaya et al., 2020) will be calculated using Equation (5).

$$Y_i = cX_i \quad (5)$$

Where  $X_i$  is the uncorrected GPM Data (in mm),  $Y_i$  is the corrected GPM Data (in mm), and  $c$  is the correction coefficient.

To test the accuracy of the corrected satellite rainfall data, a return period analysis is conducted as an evaluation method. This analysis aims to determine how well the corrected satellite data represents the pattern of extreme rainfall events in the study area. In this process, the corrected rainfall data will be analyzed using probability distributions such as Pearson Type III, Gumbel, or Log-Pearson III to determine the design rainfall for various return periods, such as 2 years, 5 years, up to 100 years (Pratiwi et al., 2023). The results of this analysis will be compared with historical data from rainfall stations to assess the level of accuracy and reliability of the corrected satellite rainfall data. Through return period analysis, it can be determined whether the correction process has improved the accuracy of extreme rainfall estimates, making it suitable for supporting hydrological planning, flood analysis, and more reliable infrastructure design. A commonly used probability distribution equation for extreme rainfall analysis is given by Equation (6)

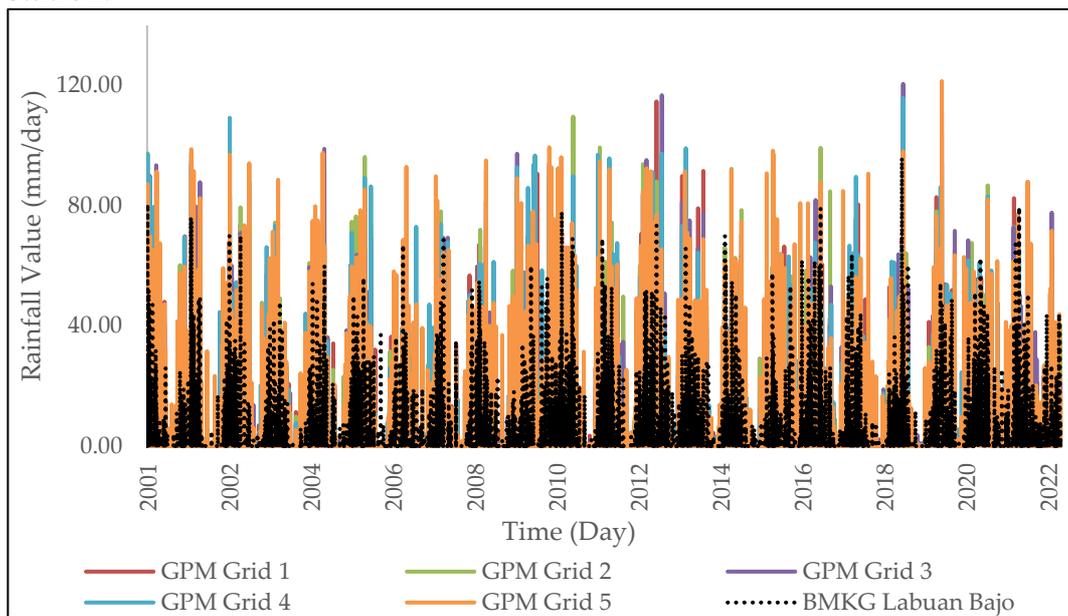
$$X_T = \bar{X} + K_T S \quad (6)$$

Where  $X_t$  is design rainfall for return period  $T$ ,  $\bar{X}$  is the mean of rainfall data,  $S$  is the standard deviation of the rainfall data, and  $K_T$  is frequency factor, which is depends on the return period and distribution type (Upomo et al., 2016; Pudyastuti et al., 2020)

### C. RESULT AND DISCUSSION

#### 1. Data Collection and Screening

At this stage, the daily rainfall data from the GPM satellite is collected. Then, this data will be compared with the data recorded at the BMKG Labuan Bajo station. After the GPM grid data is checked, the research will proceed with correcting the rainfall values recorded by the GPM satellite. The rainfall graph from the GPM satellite versus the BMKG Labuan Bajo station data, before correction, will be presented in Figure 2 below. Based on the graph bellow, the rainfall values recorded by the satellite tend to be higher when measuring rainfall events at the BMKG Labuan Bajo station.



**Figure 2 Comparison of Daily Rainfall Data Before Correction**

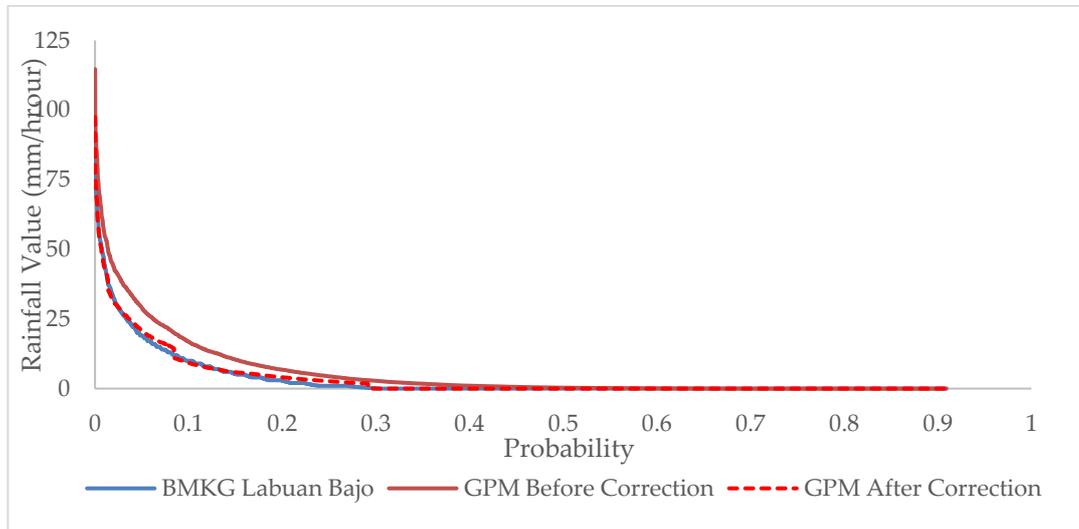
#### 2. Coefficient Correction of Daily Rainfall Data

The correction process is carried out by calculating the correction factor that will be applied to the GPM satellite data. This correction factor is obtained through a trial-and-error approach by optimizing the objective function until the result is closest to the rainfall data from the BMKG Labuan Bajo station. The correction factor obtained is expected to have a uniform value for all satellite grid data. The correction factor used to improve the satellite rainfall values in the Wae Mese watershed can be seen in Table 1 below:

**Tabel 1 Daily Rainfall Coefficient Correction**

Rainfall Range Data (mm)	Coefficient Correction
< 3	0
3 - 10	0,55
10 - 20	0,60
20 - 50	0,72
50 - 80	0,78
80 - 100	0,80
> 100	0,85

After the correction process is performed, the corrected rainfall results are compared with the observational data using the rainfall probability curve. In Figure 3 below, the rainfall correction is concluded to improve the estimation of high rainfall, so the corrected GPM curve is closer to the BMKG rainfall values.



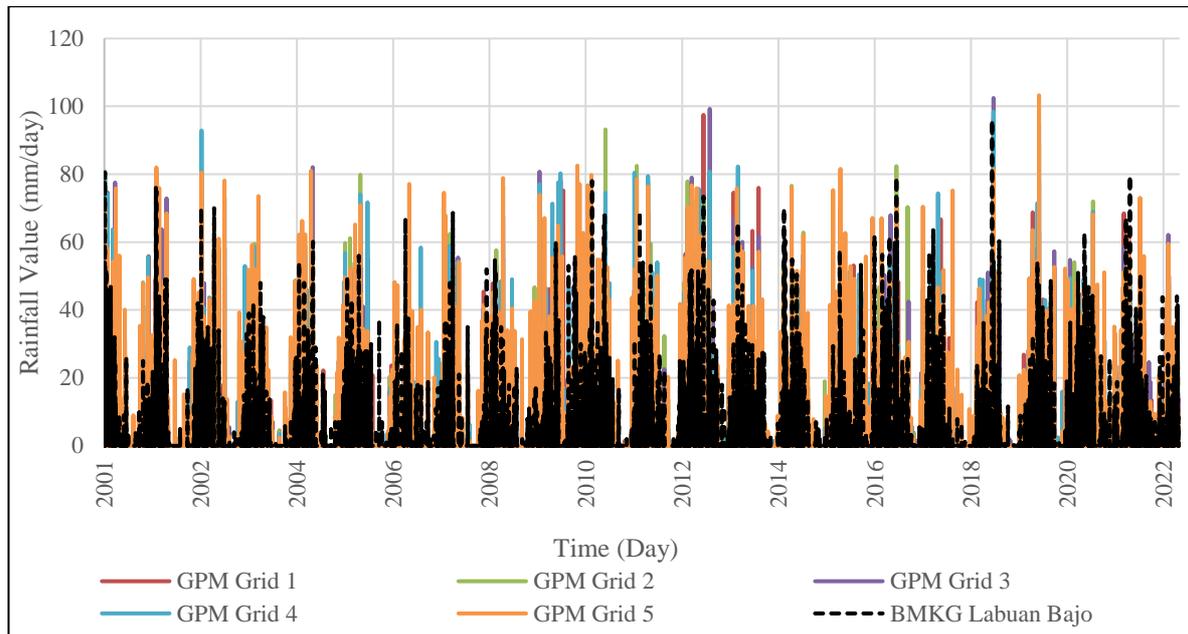
**Figure 3 Probability of Daily Rainfall Data**

After the rainfall correction, the objective function in the Wae Mese watershed showed changes in the RMSE, POD, and FAR values for each GPM grid. Based on the correction results, each grid showed significant improvements, where the RMSE value decreased by 26%-31%. This indicates that the rainfall prediction error by GPM has decreased, making the GPM estimates more accurate. Meanwhile, the POD value increased by 31%-40%, showing that GPM's ability to detect rainfall (both GPM and BMKG detected rainfall) improved after the correction, and the FAR value decreased by 23%-31%, meaning that the number of false alarms (incorrect rainfall predictions by GPM that did not occur according to BMKG) decreased, making the GPM predictions more reliable. On the other hand, the correlation value did not show significant changes before and after correction. However, based on the other three parameters, the corrected GPM rainfall values can be used for further analysis. The changes in the objective function values used can be presented in Table 3 below:

**Tabel 2 Objective Function of Daily Rainfall Data**

Objective Function Before Correction					
	Grid 1	Grid 2	Grid 3	Grid 4	Grid 5
<i>Korelasi</i>	0,57	0,56	0,51	0,49	0,49
<i>RMSE</i>	9,85	10,99	10,42	11,33	12,52
<i>POD</i>	0,40	0,41	0,40	0,41	0,42
<i>FAR</i>	0,51	0,51	0,50	0,53	0,54
Objective Function After Corection					
<i>Korelasi</i>	0,57	0,55	0,50	0,48	0,48
<i>RMSE</i>	7,80	8,37	8,41	8,96	9,55
<i>POD</i>	0,68	0,65	0,62	0,62	0,61
<i>FAR</i>	0,32	0,35	0,38	0,38	0,39
Diffrence (%)					
<i>RMSE</i>	26,35	31,27	23,85	26,41	31,08

<i>POD</i>	40,21	37,52	36,60	33,52	31,09
<i>FAR</i>	56,62	45,54	33,31	38,80	40,15



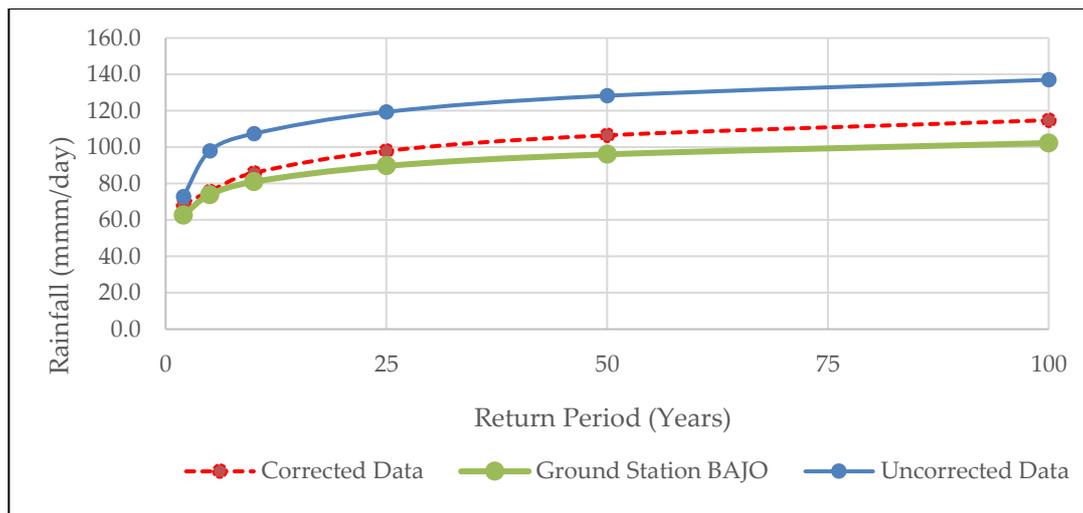
**Figure 4 Comparison of Daily Rainfall Data After Correction**

### 3. GPM Correction Application on Design Storm

Design storm analysis is done using annual maximum daily rainfall data. For design storm calculate using ground station data, then compare with both uncorrected GPM and corrected GPM at Wae Mese Watershed. Rainfall data, including both ground station and TRMM data, will undergo tests for independence, outliers, and trends. During the screening process, some rainfall stations were found to have outlier data. For upper outliers, the rainfall event of the specific year will be verified. If the recorded rainfall event is confirmed to have occurred, it will be included in the analysis. If not, it will be excluded. The calculations are performed using the Pearson III probability distribution, considering various return periods ranging from 2 years to 100 years. The results of the design storm analysis for Wae Mese Watershed are presented in table and figure below.

**Tabel 3 Design Storm Analysis for Return Period**

Return Period	Uncorrected Data	Corrected Data	BMKG Labuan Bajo
2	72.8	67.7	62.63
5	98.0	75,8	73.92
10	107.4	86,0	81.02
25	119.4	98,1	89.71
50	128.2	106,6	96.05
100	137.0	114,8	102.26



**Figure 5 Comparison of Return Period Design Storm**

#### D. CONCLUSIONS

GPM data correction based on ground station data using linear regression improves the errors between GPM data and ground station data by 31.08% of average RMSE, 31.09% of average POD and 40.15 % of average FAR for daily rainfall. GPM correction determination shows a satisfactory improvement for GPM data in the Wae Mese Watershed. GPM correction determination can be used to obtain the GPM correction coefficient for areas with similar rainfall characteristics. Applying the daily GPM correction coefficient to the maximum annual daily rainfall for design storm analysis yields better results compared to using uncorrected GPM data. While the corrected GPM results in a slightly higher design storm, this can be considered as an added safety factor.

#### ACKNOWLEDGEMENT

The author would like to express their gratitude to BMKG Labuan Bajo, the Department of Spatial Planning of Manggarai Barat, and everyone involved and assisting in the research process, both in terms of data provision and suggestions and criticisms, which made it possible to complete this paper.

#### REFERENCES

1. Bees, A., & Joko, S. (2021). Perbandingan Curah Hujan Satelit pada DAS Tilong Nusa Tenggara Timur Terkait Perubahan Iklim. *Prosiding CEEDRiMS 2021*, 25–32.
2. Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J., & Xie, P. (2020). *NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Prepared for: Global Precipitation Measurement (GPM) National Aeronautics and Space Administration (NASA)*. [https://pmm.nasa.gov/sites/default/files/imce/times\\_allsat.jpg](https://pmm.nasa.gov/sites/default/files/imce/times_allsat.jpg)
3. Harsanto, P., Prihatmanti, H. E., & Wisnulingga, B. K. (2022). Aplikasi Metode Nash Pada Perhitungan Limpasan Langsung Menggunakan Data Hujan GPM

- 3IMERGHH Studi Kasus SubDAS Winongo Hulu. *Jurnal Teknik Sumber Daya Air*, 1(2), 121–132. <https://doi.org/10.56860/jtsda.v1i2.27>
4. Made Candra Partarini, N., Sujono, J., & Prima Ari Pratiwi, E. (2021). Koreksi Dan Validasi Data Curah Hujan Satelit GPM-IMERGH dan Chirps di Das Selorejo, Kabupaten Malang. *Prosiding CEEDRiMS 2021*, 149–156. <https://giovanni.gsfc.nasa.gov/>
  5. Mamenun, M., Pawitan, H., & Sopaheluwakan, A. (2014). Validasi dan Koreksi Data Satelit Trmm Pada Tiga Pola Hujan di Indonesia. *Jurnal Meteorologi Dan Geofisika*, 15(1). <https://doi.org/10.31172/jmg.v15i1.169>
  6. Pratiwi, D. W., & Satria Negara, A. (2023). Perbandingan Pola Distribusi Hujan Terukur dan Metode Empiris dalam Perhitungan Debit Banjir Rencana DAS Jurug. *Jurnal Teknik Sumber Daya Air*, 29–42. <https://doi.org/10.56860/jtsda.v3i1.55>
  7. Pudyastuti, P. S., & Musthofa, R. A. (2020). Analysis of Maximum Daily Rainfall Distribution at Selected Rain Gauging Stations in Klaten Region in Years 2008-2018. *Dinamika Teknik Sipil*, 13(1), 10–15.
  8. PUPR. (2022). Bimbingan Teknis Analisis Debit Banjir Desain dengan Menggunakan Data Hujan Satelit. *Kementerian PUPR Balai Teknik Bendungan - Ditjen SDA*.
  9. Sanjaya, S., Yudianto, D., Adidarma, W., & Fitriana, F. (2022). Studi Pemanfaatan Curah Hujan Bulanan Satelit GPM di Kawasan Bandung Raya dengan Validasi Silang Mote-Carlo. *Jurnal Rekayasa Konstruksi Mekanika Sipil (JRKMS)*, 5(1), 31–40.
  10. Senjaya, T., Yudianto, D., Yuebo, X., & Adidarma, W. K. (2020). Application of TRMM in the Hydrological Analysis of Upper Bengawan Solo River Basin. *Journal of the Civil Engineering Forum*, 6(3), 309. <https://doi.org/10.22146/jcef.57125>
  11. Upomo, T. C., & Kusumawardani, R. (2016). Pemilihan Distribusi Probabilitas Pada Analisa Hujan-Togani dengan Metode Goodness of Fit Test. *Jurnal Teknik Sipil Dan Perencanaan*, 8(2), 139–148.
  12. Welkis, D. F. B., Sidharno, W., Wahyuni, S., & Krisnayanti, D. S. (2021). Analisis Curah Hujan pada Badai Siklon Seroja Terhadap Aliran Debit Sungai di DAS Temef. *Jurnal Ilmiah Desain & Konstruksi*, 20(2), 108–117. <https://doi.org/10.35760/dk.2021.v20i2.5048>