Civil and Environmental Science Journal (CIVENSE)

Journal homepage: https://civense.ub.ac.id/index.php/civense

Original research article

Evaluation of GPM IMERG v.06 rainfall product over the Lau Simeme Watershed in Indonesia

Bachtiar Malthus Hutagaol [abc*], James. E. Ball [b], Ery Suhartanto [c], Sri Wahyuni [c]

[a] Water Resources Department, Ministry of Public Works and Housings, Republic of Indonesia
[b] Water Engineering Department, University of Technology Sydney, Sydney, Australia
[c] Water Resources Engineering Department, Universitas Brawijaya, Malang, 65145, Indonesia

ARTICLE INFO

Keywords:
Evaluation
Rainfall
Remote sensing
GPM IMERG
Lau Simeme
Indonesia

ABSTRACT

In Indonesia, rainfall is still significant spatially and temporally. To gain optimal results from utilizing water resources, we have to ensure that the precipitation data is provided in good quality and quantity. Several spatial rainfall measurement sources, such as GPM data (Global Precipitation Measure), have become available in recent years. This study evaluated the GPM IMERG V.06 product using rain gauge measurements in the Lau Simeme watershed in North Sumatra Province, Indonesia. The relevance of the GPM IMERG was tested by direct comparison with observations at different time scales (daily, monthly, annual, and seasonal) between 2005 and 2019. Results show that the satellite product provides poor rainfall estimations at the daily and annual time scales. However, the accuracy of GPM IMERG Final datasets is improved when temporally average to monthly timescale (R2 of 0.728, RMSE of 68.318 mm and NSE of 0.725), wet seasonal time scale (R2 of 0.673, RMSE of 79.287 mm and NSE of 0.658) and dry seasonal time scale (R2 of 0.947, RMSE of 20.356 mm and NSE of 0.924).

1. Introduction

The quantification of water resources is a significant point globally [1]. In Indonesia, water resource management is one of the main issues still under discussion by the government. The reason is that many parts of Indonesia still have not received water. Utama [2] reports that the Indonesian government has predicted a shortage of clean water in Java, one of Indonesia's major islands, by 2040. Kompas.com [3] reports that some Indonesian territories will face water shortages in recent years. This situation is due to accessibility, quality, quantity, continuity, climate change, and water capacity. For example, North Sumatra, the 4th highest population in Indonesia, requires potable water every year. According to PDAM Tirtanadi (2017), 30% of the Medan area, one of the northern areas of Sumatra, suffers from a lack of clean water due to the absence of water resources.

To calculate the water availability in a watershed, rainfall recorders or rainfall station must provide the actual ground area data. One of the problems still followed in the rainfall recorder is a lack of data from the rainfall station because of data collection failure and many watersheds in Indonesia still do not have a rainfall station at all. So it is essential to provide a series of rainfall data that can be used to calculate a water balance. Gourley and Vieux [4] emphasize that precipitation is the most crucial spatial input for distributed hydrological models, so accurate rainfall estimates over a catchment or a region are critical.

In order to address this situation, the number of remotely sensed precipitation products with high spatial and temporal resolution has been developed recently [5]–[7]. These products become the alternative solution for forcing data for hydrological models when the observation data are not readily available or not available enough. Moreover, they also pose new issues to be solved by the hydrologists applying this information to their studies. Global Precipitation Measurement (GPM) is a worldwide satellite mission to support next-production observations of rain and snow every three hours.

*Corresponding author.
E-mail address: malthus.bachtiar@gmail.com (B. M. Hutagaol)
Doi: 10.21776/ub.civense.2023.00601.5
Received 13 December 2022; Accepted 27 April 2023
E-ISSN: 2620-6218/

There are three types of GPM IMERG products, namely, GPM IMERG Early Run (IMERG-E), GPM IMERG Late Run (IMERG-L), and GPM IMERG Final Run (IMERG-F). These products have spatial resolution 0.1∘ × 0.1∘ and 30 min temporal resolution. IMERG-E is a GPM product close to a real-time product with a latency of ∼4 hours. IMERG-L is a product of GPM, which is close to a real-time product with a latency of ∼14 hours. Those GPM products are suitable for monitoring natural disasters like floods and cyclones. IMERG-F is a GPM product close to the measured-adjusted product with a latency of ∼3.5 months. According to Huffman [8], GPM IMERG-F is a recommended product for research because it has been calibrated.

In water resources management, GPM IMERG has often been used globally as an alternative to predicting ground station rainfall data, such as by [1], [9], [10], [11], and [12]. In Indonesia, [13], [14], [15], [16], [17], [18], [19], and [20] have also conducted the same study. Especially for the island of Sumatra, [21], [22], [23], and [24] showed that GPM IMERG is one of the possible rainfall satellite data that can be used in imitating the ground station data.

Eventually, this study will discuss the relationship between rainfall data on the ground station and at the satellite and the reasonable formula to interpret the relationship. Moreover, the three datasets of GPM IMERG products are used to define reasonable correlation for the ground station data.

2. Study area, data collection, and methodology

2.1. Study area and data collection

Lau Simeme Reservoir is one of the Indonesian government projects located in the Deli Serdang region of North Sumatera Province. In detail, the watershed of Lau Simeme Reservoir is a part of the Percut watershed, which is at the upstream part of the Belawan Ular Padang River region (Figure 1). This watershed covers about + 10.062 Ha area, and the coordinate location of the reservoir lies between 3°21’0.39 “N and 98°38’59.39 “E (Figure 2).

Tropical climates influence Lau Simeme Watershed. This climate has two seasons, namely, the rainy season and the dry season. The two seasons are estimated to have the same weight at the research location. The rainy season starts in October-March, while the dry season starts in April-September. Hydrology condition in the Percut Watershed is influenced by several rainfall stations, such as Tongkoh Karo rainfall station, Kuta Jurung rainfall station, and Sibolangit rainfall station (Figure 3). Therefore, this study will cover those stations’ data from 2005 to 2019.

IMERG is the GPM level 3 multi-satellite precipitation algorithm, which combines alternate precipitation estimates from all microwave configuration sensors, IR observations using geosynchronous satellites, and monthly precipitation data. The system is run more than once for each shelf life. The system runs multiple times for each observing time. The IMERG Early Run (IMERG_E) is a quick estimate (near-time with l latency of 6 h), and the Late Run (IMERG_L) successively provides better estimates as more data arrive (reprocesses real-time with a latency of 18 h). Early Run and Late Run products are available in near real-time. The Final Run (IMERG_E) uses monthly measurement data to create search products (adjusted with a delay of several months).

In this study, the IMERG V06 products (IMERG_E, IMERG_L, and IMERG_F) at daily and half-hourly scales, published in 2019, are used to analyze the available rainfall events, heavy rainfall events, and extreme precipitation events from 2005 to 2019. The IMERG data could be obtained from https://giovanni.gsfc.nasa.gov/.

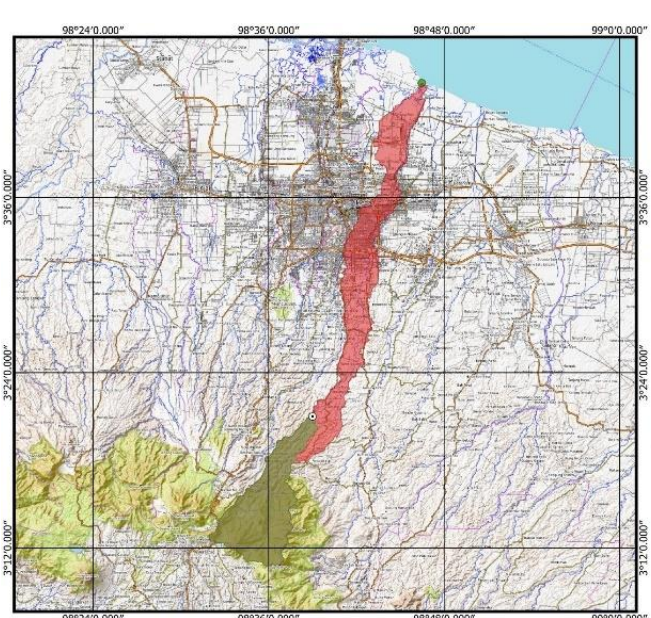

Fig. 1. Percut Watershed.

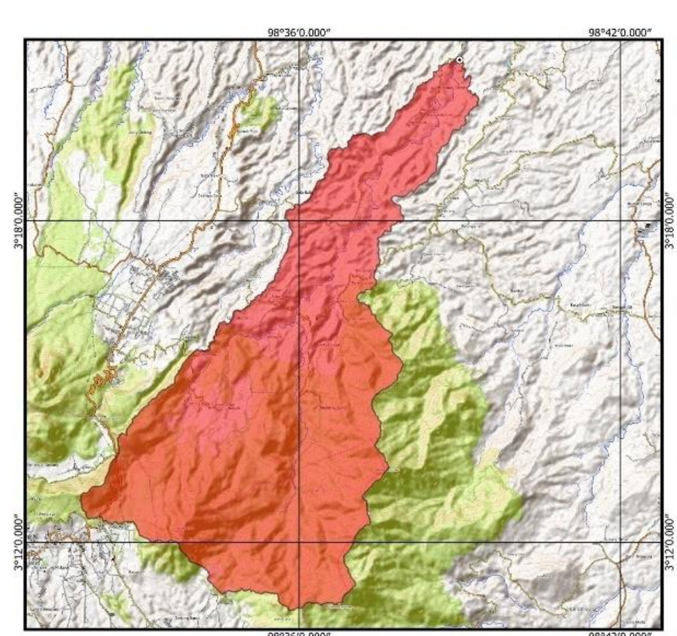

Fig. 1. Lausimeme Watershed.

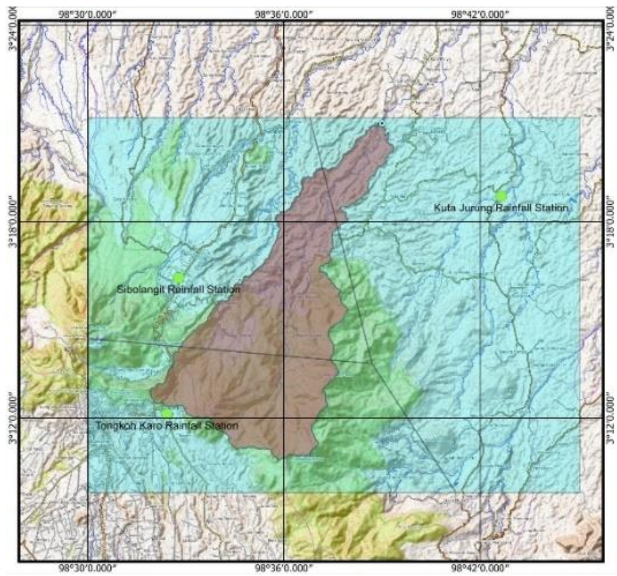

Fig. 3. Rainfall Station.

2.2 Methodology

This study will compare GPM IMERG data with ground station data to determine the relationship between the two data. The first thing to do is to do a statistical check for both data. The results of this examination will determine whether the two data are consistent, stable, and persistent. The statistical tests used are consistency tests, trend tests, stationary tests, and persistency tests.

Furthermore, to equalize the comparison, the two data are converted into the mean rainfall area using Thiessen polygon analysis. It is used because Goodrich et al. [25] and Woods et al. [26] state that although the spatial variability of rainfall plays an essential role in the process of runoff formation, rainfall modeled in Polygon Thiessen is usually assumed to be uniform for smaller common catchments.

The best correlation between the two data will be determined based on statistical metrics, which contain Correlation Coefficient (r), Root Mean Square Error (RMSE), Nash Sutcliffe Error (NSE), and Percent Bias (PBias) (Table 1). r describes the agreement between IMERG data and rain gauge data. Moreover, r is also used to determine the best variable order in multiple regression. PBias is used to describe the systematic bias of IMERG products. A smaller absolute value of bias indicates a minor deviation. Positive values indicate an overestimation of the amount of precipitation, while negative values mean an underestimation. RMSE measures the average absolute error magnitude of the IMERG products. The smaller RMSE, the closer the IMERG data is to the rain gauge data. The Nash-Sutcliffe Efficiency is a standardized statistic that defines the relative amount of the residual difference compared to the calculated data variance [27]. NSE values range from -∞ to 1. When the values are negative, the model performance is unacceptable. When the values are equal to 1, it means the models are perfectly matched with the observed data, and when the values are 0, the models are as accurate as the empirical data [28], [29].

In order to get the best formula for the two data relationships, calibration and validation were carried out. The purpose of this equation is to further research in predicting the value of rainfall in the following years so that it can support the calculation of inflow at the Lau Simeme dam. In detail, the data are separated into two types of data, namely calibration data and validation data.

Table 1.
Statistical metrics.

Name/ Symbol	Formula	Perfect Value
Correlation Coefficient (r)	$\frac{\sum_1^n (O_i - \bar{O})(O_i - \bar{P})}{\sqrt{\sum_1^n (O_i - \bar{O})^2}\sqrt{\sum_1^n (P_i - \bar{P})^2}}$	1
Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{n}\sum_1^n (O_i - P_i)^2}$	0
Nash Sutcliffe Error (NSE)	$1 - \frac{\sum_1^n (O_i - P_i)^2}{\sum_1^n (O_i - \bar{O})^2}$	1
Percent Bias (PBias)	$\left[\frac{\sum_1^n (O_i - P_i)\ x\ 100}{\sum_1^n O_i}\right]$	0

Table 2.
Calibration and validation data types.

Test	Calibration	Validation
Scenario 1	10 years	5 years
Scenario 2	12 years	3 years
Scenario 3	14 years	1 year

Calibration for single variable

The method uses the Excel program for calibration processes that use single variables, such as linear, exponential, logarithmic, polynomial and power regression. In this process, data will be collected in the Excel program and plotted in a scatter plot diagram. The plot results will show the best coefficient of determination with the approach formula as a reflection of the relationship between the two data.

Calibration for multiple variables

For calibration processes that use multiple variables, such as multiple linear, the process uses a Genetic Algorithm (GA) method. This method will be applied using Matlab R2021a. GA is a haphazard search and optimization method based on the evolutionary process of natural collection and genetics [30]. The idea of GA is that every point in a search section is assigned a qualification according to an objective purpose. Then the finest can be approached by starting with a pool of arbitrary solutions and sprouting these solutions continuously with genetic hands; the series generated from routes of solutions that derive closer and closer to the optimum [31].

In the genetic algorithm, each nominee's result is called a chromosome. The collection of all nominee results is called the population. The population size is a constraint that defines the number of nominee results. As the population develops from invention to invention, weak solutions tend to vanish, and decent solutions tend to generate well solutions. So first, the preliminary population is randomly produced, considering problematic constraints. The value that defines how excellent a chromosome is for an optimum solution is called a fitness value. In the GA, the fitness value of the result improves its prospects of being chosen for selection, mutation, crossover, and elitism processes to create a new population.

Defining the interactions between the various parameters of GA impacts the quality of the solution, and keeping the values of the parameters "balanced" enhances the solution of the GA. GA uses four basic and important parameters: crossover rate, mutation rate, population size, elitism ratio, and the number of generations. In this study, the GA parameters can be shown in Table 3.

3. Result and discussion

3.1 Choosing the best correlation

This section compares ground station data and satellite station data to determine the largest correlation that reflects the relationship between the two data. This comparison is applied to daily, monthly, annual, and seasonal data. In order to determine the highest correlation, the coefficient correlation equation is applied. The result can be seen in Table 4.

In Table 4, Figure 4, Figure 5, and Figure 6, the correlation between the ground station and the GPM IMERG shown at the daily and annual stages is weak, with a value below 0.5. It indicates that the IMERG GPM will be unable to represent the value of the ground station at this stage. This low satellite station value is probably due to the rapid displacement of clouds and space and the temporal distribution of precipitation influenced by storm events within the study area [1].

In contrast, the correlation between those data regarding the monthly, wet, and dry seasons shows excellent relationships where the values are between 0.6 and 0.8. It means that the GPM IMERG can interpret the ground station value well. It can be seen that the relationship between the ground station and GPM IMERG F is the highest correlation. In terms of the monthly stage and the wet stage, they stand at 0.79. In terms of the dry stage, it lies at 0.77. It means that those correlations are strong. Overall, it can be seen that the movement of the data from those graphs fluctuates in Figure 4, Figure 5, and Figure 6. It means that those data are reasonably distributed in monthly, wet, and dry steps.

Therefore, the best GPM IMERG data that can interpret the calibration and validation of the ground station in Lau Simeme Watershed is the GPM IMERG Final dataset from monthly step, wet, and dry seasonal steps.

Table 3.
GA parameters.

Control	Test 1	Test 2	Test 3
Number of Chromosomes (M)	20	30	50
Maximum Generation	100	500	1000
Crossover ratio	0.8	0.8	0.8
Mutation ratio	0.2	0.2	0.2
Elitism ratio	0.2	0.2	0.2

Table 4.
Correlation coefficient for every time step.

Comparison	Pearson Coefficient Correlation (r)				
	Daily	Monthly	Annual	Wet Season	Dry Season
Groundstation – GPM IMERG E	0.25	0.64	0.25	0.69	0.63
Groundstation – GPM IMERG L	0.25	0.65	0.31	0.70	0.64
Groundstation – GPM IMERG F	0.26	0.79	0.26	0.79	0.77

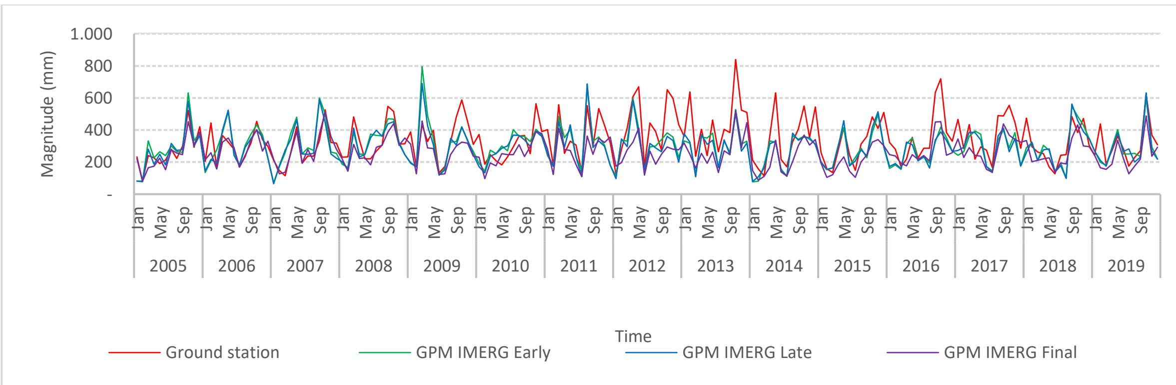

Fig. 2. Monthly step correlation.

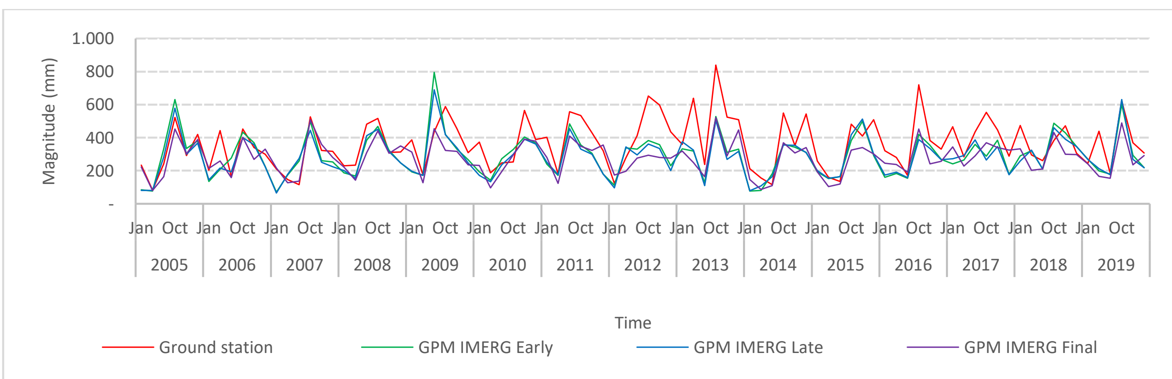

Fig. 3. Wet season step correlation.

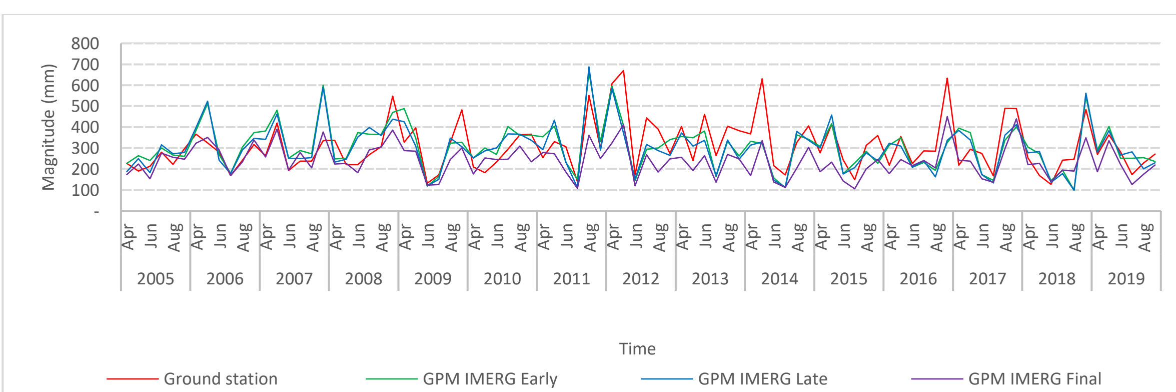

Fig. 4. Dry season step correlation.

3.2 Calibration and validation result

In order to define the relationship between the ground station and GPM IMERG data, it is important to model the relationship using mathematical regression. Several simple regression analysis alternatives that are commonly used in hydrological data analysis are single regression and multiple regression [32]. This study applies simple regression, such as linear, exponential, logarithmic, polynomial, and power regressions, using the mathematical software Excel using monthly rainfall data at the location under review (x). The multiple regressions for independent variables are tested using a genetic algorithm method. The GA parameters are shown in Table 3. In addition, the order of the variables tested for each regression in multiple regression are monthly rainfall data at the location under review (x_1), monthly data on the average rainfall at the location under review (x_2), monthly minimum rainfall data at the location under review (x_3), monthly data on maximum rainfall at the location under review (x_4). Hence, Table 5, Table 6, and Table 7 reflect the results of calibration and validation calculations for the monthly step, dry step, and wet steps, respectively, of the relationship between the ground station and the GPM IMERG satellite station.

Based on Tables 5, Table 6, and Table 7, most of the relationships between the two data, the GPM IMERG and the ground station, can be reflected in equations with single and multiple variables. The results found that 99% of the trials carried out produced satisfactory NSE and R2 values with values above 0.5. It shows that the resulting equation is reasonably good for interpreting the data on the ground station. In Table 5, the validation process for the equation with one variable for the monthly stage, it is found that the linear regression in monthly scenario 2, with 12 years of calibration data and three years of validation data, gave a very satisfactory interpretation of NSE and R2 with an NSE value of 0.725. In Table 6 and Table 7, the wet and dry seasons, it is found that the linear regression in wet scenario one and the power regression in dry scenario 3 gave very satisfactory NSE and R2 values with NSE values of 0.662 and 0.924, respectively.

In Table 5, the validation process for equations with multiple variables for monthly stages, it is found that the equation in monthly scenario one with three variables produces the best NSE and R2 values with NSE values of 0.694. In Table 6, the wet season, it is found that the regression with variable 3 in wet scenario 1 gave very satisfactory NSE and R2 values with an NSE value of 0.658. In Table 7, the dry season, it is found that the regression with variable 4 in dry scenario 2 gave very satisfactory NSE and R2 values with an NSE value of 0.836.

Overall, the results reflect that a good NSE value in the calibration process does not yet mean a good NSE value in the validation process. In other words, the results obtained in the calibration process are not directly proportional to those obtained in the validation process. Even though those tables show that the ground station value (the corrected GPM IMERG) on the Lau Simeme Watershed can be obtained using the GPM IMERG final dataset on a monthly and seasonal scale with one or multiple variables.

Table 5.
Calibration and validation results for monthly data.

Variable	Regression types	Equation	Calibration result				Validation result			
			R^2	RMSE	NSE	PBIAS	R^2	RMSE	NSE	PBIAS
Monthly Scenario 1 (10 years calibration vs 5 years validation)										
Single Variable	Linear	$y = 1.1802x + 28.841$	0.59	91.982	0.589	-0.002	0.645	73.465	0.633	1.823
	Logarithmic	$y = 273.05\ln(x) - 1162.9$	0.563	94.887	0.563	0.011	0.626	75.12	0.616	0.702
	Exponential	$y = 115.36e^{0.0038x}$	0.553	98.26	0.531	1.462	0.612	80.966	0.554	3.804
	Polynomial	$y = -0.0002x^2 + 1.2845x + 16.209$	0.589	91.961	0.589	0.227	0.645	73.429	0.633	2.019
	Power	$y = 2.0682x^{0.9103}$	0.59	92.664	0.583	3.246	0.645	72.335	0.644	4.97
Multiple Variable (2 Variable)	Test 1	$y = 32.12 + 1.41\ x_1 - 7.14\ x_2$	0.589	91.972	0.589	-0.033	0.722	73.442	0.681	7.423
	Test 2	$y = 31.26 + 1.26\ x_1 - 2.73\ x_2$	0.589	91.970	0.589	0.000	0.722	73.442	0.681	7.423
	Test 3	$y = 21.16 + 2.09\ x_1 - 23.63\ x_2$	0.589	91.966	0.589	0.000	0.722	73.442	0.681	7.423
Multiple Variable (3 Variable)	Test 1	$y = 15.94 + 0.30\ x_1 + 26.57\ x_2 + 70.14\ x_3$	0.596	91.226	0.596	0.093	0.699	71.961	0.694	0.954
	Test 2	$y = 38.71 + 1.17\ x_1 - 2.28\ x_2 + 95.07\ x_3$	0.603	90.436	0.603	-0.003	0.681	74.396	0.673	-0.392
	Test 3	$y = 36.90 + 1.26\ x_1 - 4.92\ x_2 + 97.56\ x_3$	0.603	90.421	0.603	0.000	0.680	74.881	0.668	-0.451
Multiple Variable (4 Variable)	Test 1	$y = 40.35 + 0.73\ x_1 + 15.99\ x_2 + 44.79\ x_3 - 0.89\ x_4$	0.615	89.111	0.614	0.106	0.694	72.041	0.693	-0.116
	Test 2	$y = 54.71 + 0.44\ x_1 + 25.92\ x_2 + 82.39\ x_3 - 1.60\ x_4$	0.616	88.937	0.616	0.037	0.646	78.617	0.635	-1.978
	Test 3	$y = 60.68 + 1.47\ x_1 - 5.66\ x_2 + 62.62\ x_3 - 1.54\ x_4$	0.62	88.458	0.62	0.005	0.657	77.089	0.649	-2.009
Monthly Scenario 2 (12 years calibration vs 3 years validation)										
Single Variable	Linear	$y = 1.2222x + 21.786$	0.628	87.449	0.628	0.003	0.728	68.318	0.725	-2.741
	Logarithmic	$y = 281.67\ln(x) - 1206.8$	0.595	91.262	0.595	-0.011	0.689	73.342	0.683	-3.885
	Exponential	$y = 115.07e^{0.0038x}$	0.594	93.145	0.578	2.921	0.7	73.765	0.68	0.648
	Polynomial	$y = -3E\text{-}05x^2 + 1.238x + 19.874$	0.628	87.448	0.628	0.03	0.728	68.519	0.724	-2.72
	Power	$y = 1.9407x^{0.9241}$	0.628	88.145	0.622	3.011	0.728	70.455	0.708	0.289
Multiple Variable (2 Variable)	Test 1	$y = 7.14 + 0.50\ x_1 - 23.10\ x_2$	0.585	88.318	0.621	0.385	0.659	71.822	0.647	-1.761
	Test 2	$y = 34.97 + 1.76\ x_1 - 17.61\ x_2$	0.588	87.376	0.629	-0.111	0.631	75.101	0.615	-3.304
	Test 3	$y = 28.51 + 1.58\ x_1 - 11.57\ x_2$	0.589	87.341	0.629	0.000	0.636	74.533	0.620	-3.040
Multiple Variable (3 Variable)	Test 1	$y = 38.01 + 2.14\ x_1 - 31.39\ x_2 + 84.13\ x_3$	0.599	85.992	0.640	1.169	0.577	86.581	0.488	-5.368
	Test 2	$y = 35.02 + 1.44\ x_1 - 9.64\ x_2 + 98.84\ x_3$	0.603	85.479	0.645	-0.018	0.582	87.576	0.476	-6.547
	Test 3	$y = 37.44 + 1.46\ x_1 - 10.75\ x_2 + 101.18\ x_3$	0.603	85.475	0.645	0.000	0.579	87.980	0.471	-6.639

Variable	Regression types	Equation	Calibration result				Validation result			
			R^2	RMSE	NSE	PBIAS	R^2	RMSE	NSE	PBIAS
Multiple Variable (4 Variable)	Test 1	$y = 30.22 + 1.62\,x_1 - 8.53\,x_2 + 64.46\,x_3 - 1.13\,x_4$	0.619	84.442	0.653	0.446	0.588	87.34	0.479	-6.779
	Test 2	$y = 30.18 + 0.77\,x_1 + 20.65\,x_2 + 63.69\,x_3 - 1.83\,x_4$	0.647	85.364	0.646	-0.052	0.592	87.949	0.471	-7.703
	Test 3	$y = 30.83 + 0.94\,x_1 + 12.03\,x_2 + 78.54\,x_3 - 1.16\,x_4$	0.686	84.467	0.653	0.016	0.592	87.379	0.478	-7.232
Monthly Scenario 3 (14 years calibration vs 1 years validation)										
Single Variable	Linear	$y = 1.2138x + 21.935$	0.629	85.507	0.629	0	0.638	74.221	0.626	1.071
	Logarithmic	$y = 283.13\ln(x) - 1217.5$	0.6	88.796	0.6	0.006	0.557	79.954	0.566	0.379
	Exponential	$y = 114.97e^{0.0038x}$	0.592	91.525	0.575	2.568	0.685	80.727	0.558	2.428
	Polynomial	$y = -0.0002x^2 + 1.3222x + 8.7282$	0.63	85.481	0.629	0.067	0.632	74.686	0.622	1.242
	Power	$y = 1.8722x^{0.9296}$	0.63	86.109	0.624	2.905	0.634	74.661	0.622	3.965
Multiple Variable (2 Variable)	Test 1	$y = 27.55 + 1.52\,x_1 - 9.74\,x_2$	0.629	85.484	0.629	-0.246	0.074	144.384	-0.474	-17.353
	Test 2	$y = 27.53 + 1.43\,x_1 - 7.12\,x_2$	0.629	85.474	0.629	-0.001	0.075	143.516	-0.457	-16.979
	Test 3	$y = 25.56 + 1.42\,x_1 - 6.56\,x_2$	0.629	85.472	0.629	0.000	0.075	143.796	-0.462	-17.016
Multiple Variable (3 Variable)	Test 1	$y = 19.41 + 1.22\,x_1 - 0.65\,x_2 + 49.83\,x_3$	0.638	84.589	0.637	-0.282	0.071	147.642	-0.542	-17.797
	Test 2	$y = 26.48 + 1.11\,x_1 + 1.45\,x_2 + 66.75\,x_3$	0.638	84.477	0.638	0.050	0.069	146.547	-0.519	-17.295
	Test 3	$y = 32.53 + 1.34\,x_1 - 6.36\,x_2 + 67.26\,x_3$	0.638	84.424	0.638	0.000	0.066	146.717	-0.522	-17.443
Multiple Variable (4 Variable)	Test 1	$y = 27.06 + 3.14\,x_1 - 59.22\,x_2 + 50.03\,x_3 - 0.08\,x_4$	0.623	87.421	0.612	0.668	0.046	154.959	-0.698	-18.607
	Test 2	$y = 40.66 + 1.76\,x_1 - 14.61\,x_2 + 32.97\,x_3 - 0.96\,x_4$	0.646	83.591	0.646	0.236	0.052	152.739	-0.65	-19.251
	Test 3	$y = 38.97 + 1.21\,x_1 + 1.25\,x_2 + 55.35\,x_3 - 0.83\,x_4$	0.647	83.381	0.647	0.005	0.057	151.298	-0.619	-18.922

Table 6.

Calibration and Validation results for wet season data.

Variable	Regression types	Equation	Calibration result				Validation result			
			R^2	RMSE	NSE	PBIAS	R^2	RMSE	NSE	PBIAS
Wet Season Scenario 1 (10 years calibration vs 5 years validation)										
Single Variable	Linear	$y = 1.1712x + 36.531$	0.615	99.594	0.615	0.001	0.666	78.952	0.662	2.027
	Logarithmic	$y = 278.96\ln(x) - 1182.8$	0.61	100.34	0.609	0.018	0.627	83.309	0.623	-0.439
	Exponential	$y = 112.69e^{0.0038x}$	0.556	111.288	0.519	2.683	0.638	87.324	0.586	6.939
	Polynomial	$y = -0.0009x^2 + 1.6924x - 26.454$	0.621	98.83	0.621	-0.54	0.658	79.504	0.657	0.674
	Power	$y = 1.718x^{0.9459}$	0.617	100.079	0.611	3.12	0.665	80.641	0.647	5.009
Multiple Variable (2 Variable)	Test 1	$y = 26.72 + 0.97\,x_1 + 7.03\,x_2$	0.613	99.817	0.613	-0.285	0.667	78.741	0.663	1.943
	Test 2	$y = 50.13 + 1.66\,x_1 - 15.87\,x_2$	0.616	99.412	0.616	0.002	0.658	79.569	0.656	1.606
	Test 3	$y = 53.17 + 1.64\,x_1 - 15.61\,x_2$	0.616	99.404	0.616	0.000	0.658	79.613	0.656	1.602
Multiple Variable (3 Variable)	Test 1	$y = 16.41 + 1.08\,x_1 + 3.65\,x_2 + 93.18\,x_3$	0.634	97.374	0.632	-0.507	0.673	79.287	0.658	0.669
	Test 2	$y = 58.12 + 1.69\,x_1 - 19.15\,x_2 + 93.89\,x_3$	0.637	96.678	0.637	0.033	0.664	79.435	0.657	0.524
	Test 3	$y = 59.31 + 1.56\,x_1 - 16.09\,x_2 + 129.87\,x_3$	0.639	96.433	0.639	0.000	0.660	80.810	0.645	0.174
Multiple Variable (4 Variable)	Test 1	$y = 13.70 + 1.93\,x_1 - 13.80\,x_2 + 24.53\,x_3 - 1.33\,x_4$	0.635	100.049	0.611	-0.235	0.639	85.113	0.606	0.434
	Test 2	$y = 15.10 + 0.64\,x_1 + 25.92\,x_2 + 82.39\,x_3 - 0.53\,x_4$	0.637	96.894	0.636	0.024	0.665	80.14	0.651	1.267
	Test 3	$y = 58.27 + 0.96\,x_1 - 7.05\,x_2 + 99.33\,x_3 - 0.94\,x_4$	0.643	95.9	0.643	-0.121	0.653	80.716	0.646	0.363
Wet Scenario 2 (12 years calibration vs 3 years validation)										
Single Variable	Linear	$y = 1.2286x + 23.506$	0.651	95	0.651	0.004	0.478	86.093	0.441	-0.556
	Logarithmic	$y = 289.69\ln(x) - 1240.6$	0.636	97.025	0.636	0.013	0.438	86.542	0.435	-2.946
	Exponential	$y = 110e^{0.0039x}$	0.591	106.777	0.559	3.137	0.481	98.903	0.262	4.285
	Polynomial	$y = -0.0008x^2 + 1.6549x - 27.937$	0.655	94.568	0.654	1.007	0.466	85.639	0.446	-0.301
	Power	$y = 1.5193x^{0.9693}$	0.652	95.496	0.647	2.941	0.477	86.706	0.433	2.091
Multiple Variable (2 Variable)	Test 1	$y = 31.55 + 2.28\,x_1 - 31.91\,x_2$	0.652	95.460	0.647		0.451	92.529	0.334	-1.939
	Test 2	$y = 31.18 + 1.47\,x_1 - 8.05\,x_2$	0.652	94.799	0.652	-0.027	0.471	86.906	0.412	-0.780
	Test 3	$y = 40.00 + 1.72\,x_1 - 16.39\,x_2$	0.653	94.728	0.653	0.000	0.464	87.745	0.401	-0.936

Variable	Regression types	Equation	Calibration result				Validation result			
			R^2	RMSE	NSE	PBIAS	R^2	RMSE	NSE	PBIAS
Multiple Variable (3 Variable)	Test 1	$y = 69.59 + 1.88 x_1 - 24.82 x_2 + 47.53 x_3$	0.669	92.759	0.667	-1.026	0.457	89.197	0.381	-2.163
	Test 2	$y = 48.32 + 1.31 x_1 - 6.91 x_2 + 96.69 x_3$	0.678	91.347	0.677	0.137	0.467	90.075	0.369	-1.008
	Test 3	$y = 53.31 + 1.59 x_1 - 16.21 x_2 + 130.04 x_3$	0.680	90.944	0.680	0.000	0.454	94.646	0.303	-1.688
Multiple Variable (4 Variable)	Test 1	$y = 12.75 + 1.04 x_1 + 4.08 x_2 + 87.75 x_3 + 0.30 x_4$	0.669	92.623	0.668	-0.631	0.481	90.596	0.362	-1.708
	Test 2	$y = 33.68 + 1.44 x_1 - 6.79 x_2 + 93.04 x_3 - 0.50 x_4$	0.681	90.88	0.68	0.271	0.46	94.14	0.311	-1.388
	Test 3	$y = 63.98 + 1.56 x_1 - 12.07 x_2 + 99.47 x_3 - 0.81 x_4$	0.683	90.498	0.683	-0.178	0.447	94.162	0.31	-1.766
Wet Scenario 3 (14 years calibration vs 1 years validation)										
Single Variable	Linear	$y = 1.2057x + 28.405$	0.635	92.729	0.634	0.002	0.56	99.69	0.512	4.29
	Logarithmic	$y = 286.58\ln(x) - 1226.4$	0.623	94.148	0.623	0.001	0.463	105.931	0.448	3.511
	Exponential	$y = 112.74e^{0.0038x}$	0.574	103.448	0.545	3.933	0.628	110.165	0.403	5.995
	Polynomial	$y = -0.0009x^2 + 1.6897x - 30.839$	0.64	92.178	0.639	0.961	0.518	103.2	0.477	5.759
	Power	$y = 1.5989x^{0.9599}$	0.636	93.141	0.631	2.707	0.556	102.608	0.482	7.124
Multiple Variable (2 Variable)	Test 1	$y = 32.34 + 1.35 x_1 - 4.55 x_2$	0.635	92.642	0.635	-0.086	0.553	100.456	0.504	3.992
	Test 2	$y = 39.01 + 1.50 x_1 - 9.75 x_2$	0.636	92.593	0.636	0.000	0.544	101.118	0.497	3.797
	Test 3	$y = 39.01 + 1.50 x_1 - 9.75 x_2$	0.636	92.589	0.636	0.000	0.541	101.458	0.494	3.709
Multiple Variable (3 Variable)	Test 1	$y = 41.84 + 1.29 x_1 - 4.74 x_2 + 83.01 x_3$	0.657	90.054	0.655	-1.566	0.556	108.989	0.416	-1.607
	Test 2	$y = 61.05 + 1.74 x_1 - 21.12 x_2 + 99.73 x_3$	0.658	89.674	0.658	-0.062	0.535	112.504	0.378	-1.834
	Test 3	$y = 53.20 + 1.43 x_1 - 11.46 x_2 + 123.54 x_3$	0.660	89.457	0.660	0.000	0.547	114.908	0.351	-2.487
Multiple Variable (4 Variable)	Test 1	$y = 65.84 + 2.34 x_1 - 34.43 x_2 + 47.07 x_3 - 1.21 x_4$	0.652	92.113	0.639	4.685	0.427	124.066	0.243	5.627
	Test 2	$y = 51.92 + 0.84 x_1 + 13.38 x_2 + 90.89 x_3 - 1.38 x_4$	0.66	89.487	0.66	0.026	0.489	119.619	0.296	1.019
	Test 3	$y = 34.24 + 0.82 x_1 + 13.72 x_2 + 99.08 x_3 - 0.93 x_4$	0.661	89.303	0.661	0.158	0.52	117.947	0.316	0.688

Table 7.
Calibration and validation results for dry season data.

Variable	Regression types	Equation	Calibration result				Validation result			
			R^2	RMSE	NSE	PBIAS	R^2	RMSE	NSE	PBIAS
Dry Scenario 1 (10 years calibration vs 5 years validation)										
Single Variable	Linear	$y = 1.1693x + 26.32$	0.514	83.347	0.514	0.002	0.766	55.052	0.771	1.487
	Logarithmic	$y = 252.31\ln(x) - 1062.5$	0.472	86.921	0.472	0.008	0.719	61.187	0.717	1.367
	Exponential	$y = 117.52e^{0.0037x}$	0.53	82.834	0.52	3.502	0.742	58.574	0.741	3.305
	Polynomial	$y = 0.0021x^2 + 0.1254x + 144.81$	0.528	82.188	0.528	-0.019	0.746	57.319	0.752	0.306
	Power	$y = 3.0988x^{0.8342}$	0.51	84.842	0.497	3.198	0.764	58.464	0.742	4.078
Multiple Variable (2 Variable)	Test 1	$y = 9.17 + 1.11 x_1 + 3.60 x_2$	0.514	83.514	0.512	1.026	0.765	55.448	0.759	3.026
	Test 2	$y = 26.80 + 1.29 x_1 - 3.71 x_2$	0.514	83.344	0.514	0.000	0.767	54.983	0.763	1.461
	Test 3	$y = 26.47 + 1.30 x_1 - 3.95 x_2$	0.514	83.344	0.514	0.003	0.767	54.971	0.763	1.474
Multiple Variable (3 Variable)	Test 1	$y = 23.42 - 0.10 x_1 + 37.78 x_2 + 32.49 x_3$	0.515	83.432	0.513	1.359	0.732	58.851	0.728	1.854
	Test 2	$y = 35.41 + 2.14 x_1 - 31.53 x_2 + 53.84 x_3$	0.518	83.031	0.518	-0.260	0.716	60.549	0.713	-1.008
	Test 3	$y = 31.65 + 1.23 x_1 - 3.59 x_2 + 57.55 x_3$	0.520	82.859	0.520	0.000	0.712	61.289	0.705	-0.702
Multiple Variable (4 Variable)	Test 1	$y = 13.70 + 1.93 x_1 - 13.80 x_2 + 24.53 x_3 - 1.33 x_4$	0.544	81.664	0.534	3.611	0.649	68.72	0.63	-0.851
	Test 2	$y = 15.10 + 0.64 x_1 + 25.92 x_2 + 82.39 x_3 - 0.53 x_4$	0.557	79.654	0.557	-0.141	0.664	67.835	0.639	-5.275
	Test 3	$y = 58.27 + 0.96 x_1 - 7.05 x_2 + 99.33 x_3 - 0.94 x_4$	0.559	79.482	0.558	0.06	0.689	64.515	0.674	-3.863
Dry Scenario 2 (12 years calibration vs 3 years validation)										
Single Variable	Linear	$y = 1.1906x + 25.99$	0.563	78.99	0.563	0	0.738	57.294	0.741	-5.922
	Logarithmic	$y = 259.09\ln(x) - 1094.6$	0.513	83.373	0.513	0.005	0.712	60.053	0.715	-5.984
	Exponential	$y = 120.56e^{0.0037x}$	0.577	78.016	0.573	2.214	0.702	61.326	0.703	-4.783
	Polynomial	$y = 0.0018x^2 + 0.2907x + 130.46$	0.575	77.876	0.575	-0.723	0.718	61.619	0.7	-7.502
	Power	$y = 3.0669x^{0.8391}$	0.557	80.547	0.545	2.973	0.738	55.995	0.753	-3.173
Multiple Variable (2 Variable)	Test 1	$y = 33.27 + 1.45 x_1 - 8.76 x_2$	0.563	78.963	0.563	0.019	0.738	57.397	0.712	-6.169
	Test 2	$y = 27.46 + 1.81 x_1 - 18.91 x_2$	0.563	78.917	0.563	0.005	0.738	57.459	0.712	-6.168
	Test 3	$y = 27.68 + 1.79 x_1 - 18.28 x_2$	0.563	78.917	0.563	0.000	0.738	57.458	0.712	-6.172

Variable	Regression types	Equation	Calibration result				Validation result			
			R^2	RMSE	NSE	PBIAS	R^2	RMSE	NSE	PBIAS
Multiple Variable (3 Variable)	Test 1	$y = 19.48 + 0.69\ x_1 + 53.98\ x_2 + 15.00\ x_3$	0.564	78.907	0.563	0.207	0.717	67.007	0.608	-9.390
	Test 2	$y = 31.26 + 2.17\ x_1 - 31.00\ x_2 + 41.90\ x_3$	0.566	78.661	0.566	0.040	0.721	64.400	0.638	-9.467
	Test 3	$y = 31.50 + 1.72\ x_1 - 17.61\ x_2 + 46.94\ x_3$	0.567	78.618	0.567	0.000	0.719	65.325	0.627	-9.735
Multiple Variable (4 Variable)	Test 1	$y = 11.42 + 0.17\ x_1 + 39.49\ x_2 - 14.21\ x_3 - 1.28\ x_4$	0.583	77.335	0.581	0.118	0.74	60.547	0.68	-8.096
	Test 2	$y = 35.66 - 1.62\ x_1 + 90.71\ x_2 + 38.12\ x_3 - 1.45\ x_4$	0.567	78.625	0.567	0.06	0.73	69.834	0.574	-12.547
	Test 3	$y = 40.81 + 0.96\ x_1 + 12.73\ x_2 + 31.65\ x_3 - 1.56\ x_4$	0.592	76.238	0.592	-0.003	0.739	69.07	0.583	-13.493
Dry Scenario 3 (14 years calibration vs 1 years validation)										
Single Variable	Linear	$y = 1.2017x + 20.646$	0.5874	77.377	0.587	-0.003	0.94	27.49	0.861	-3.515
	Logarithmic	$y = 266.91\ln(x) - 1139.5$	0.5406	81.642	0.541	-0.017	0.972	27.181	0.864	-4.652
	Exponential	$y = 116.91e^{0.0038x}$	0.5893	77.512	0.586	1.561	0.884	27.27	0.863	-1.59
	Polynomial	$y = 0.0012x^2 + 0.5605x + 96.005$	0.594	76.811	0.593	0.778	0.91	26.168	0.874	-2.943
	Power	$y = 2.5596x^{0.8703}$	0.5838	78.682	0.573	2.955	0.947	20.356	0.924	-1.342
Multiple Variable (2 Variable)	Test 1	$y = 19.04 + 0.29\ x_1 + 27.99\ x_2$	0.583	77.791	0.583	-0.424	0.948	25.945	0.794	-3.684
	Test 2	$y = 23.16 + 1.64\ x_1 - 13.57\ x_2$	0.588	77.303	0.588	-0.080	0.936	28.479	0.751	-3.805
	Test 3	$y = 22.46 + 2.02\ x_1 - 24.91\ x_2$	0.588	77.296	0.588	0.000	0.932	29.570	0.732	-3.764
Multiple Variable (3 Variable)	Test 1	$y = 16.35 + 0.70\ x_1 + 14.41\ x_2 + 36.57\ x_3$	0.587	77.577	0.585	1.183	0.899	44.855	0.383	-4.829
	Test 2	$y = 19.82 + 0.95\ x_1 + 16.99\ x_2 + 7.46\ x_3$	0.588	77.345	0.588	-0.239	0.922	36.371	0.594	-4.984
	Test 3	$y = 25.24 + 1.86\ x_1 - 20.93\ x_2 + 23.12\ x_3$	0.590	77.155	0.590	0.000	0.903	41.199	0.480	-5.664
Multiple Variable (4 Variable)	Test 1	$y = 3.73 + 2.51\ x_1 - 30.96\ x_2 - 20.73\ x_3 - 1.25\ x_4$	0.605	76.16	0.6	-0.043	0.958	23.15	0.836	-3.072
	Test 2	$y = 28.06 + 1.45\ x_1 - 3.32\ x_2 + 8.58\ x_3 - 1.10\ x_4$	0.606	75.632	0.606	0.666	0.928	31.364	0.698	-5.401
	Test 3	$y = 26.71 + 1.85\ x_1 - 13.77\ x_2 - 0.28\ x_3 - 1.34\ x_4$	0.607	75.578	0.606	0.023	0.936	29.444	0.734	-5.718

4. Conclusion

In the next few years, the Lau Simeme reservoir will become a vital dam for water supply and flood control for North Sumatra Province, especially in the Medan area and its surroundings. For this reason, it is necessary to carry out a very in-depth study regarding the operation of the hydrological system in the watershed. In order to make the hydrological function effective, rainfall is fundamental data to ensure the operation of the reservoir. By using satellite station data, deficiencies and loss of rainfall data can be reduced and handled. The method carried out in this study provides excellent benefits in overcoming deficiencies and loss of data and predicting future data for the use of water resources management.

The results found that the GPM IMERG final datasets are the best data that can be used to reflect the ground station data in Lau Simeme Watershed in terms of monthly and wet and dry seasonal steps. The calibration and validation processes show that 99% of the trials passed satisfactory NSE and R2 values above 0.5. Furthermore, it shows that the resulting equation is reasonably good for interpreting the data on the ground station. In monthly stages, scenario 2 in simple regression for a single variable and scenario 1 in there variables for multiple variables are the best formula to describe the corrected GPM IMERG data in Lau Simeme Watershed with NSE values of 0.725 and 0.694, respectively. In the wet season stages, scenario 1 in simple regression for a single variable and scenario 1 in three variables for multiple variables reflect the reasonable formula with NSE 0.662 and 0.658, respectively. Whereas, in the dry season stages, scenario 3 in power regression for a single variable and scenario 2 in four variables for multiple variables reflect the excellent formula with NSE 0.924 and 0.836, respectively.

The study results also found that a good NSE value in the calibration process did not provide a good value yet in the validation process. Therefore, the form of the equation that can be used to interpret the ground station data through the GPM IMERG final dataset on the Lau Simeme watershed is as follows:

- Monthly period, scenario 2, Single variable with linear Regression type
- Monthly period, scenario 1, Multiple variables 3 with test 1 Regression type
- The wet season, scenario 1, Single variable with linear Regression type
- The wet season, scenario 1, Multiple variables 3 with test 1 Regression type
- Dry period, scenario 3, Single variable with power Regression type
- Dry period, scenario 3, Multiple variables 4 with test 1 Regression type

To further study: (1) It is necessary to research other satellite rainfall data to find out the best provider that can produce even better results; (2) Use other equations to determine the relationship between rainfall stations.

References

[1] H. Ouatiki *et al.*, "Evaluation of TRMM 3B42 V7 rainfall product over the Oum Er Rbia watershed in Morocco," *Climate*, vol. 5, no. 1, 2017, doi: 10.3390/cli5010001.

[2] A. Utama, "Jawa 'kehabisan air' tahun 2040: Ratusan juta orang terancam bencana yang 'tak pernah terbayangkan' (Java 'runs out of water' in 2040: Hundreds of million people at risk of 'unimaginable' disaster) - BBC News Indonesia," *BBC Indonesia*, 2019. https://www.bbc.com/indonesia/indonesia-49190635 (accessed Nov. 09, 2021).

[3] Kompas.com, "Jawa, Bali dan Nusa Tenggara, Wilayah yang Sering Alami Krisis Air," (Java, Bali, and Nusa Tenggara, Areas Frequently experiencing water crisis) *Kompas.com*, 2021. https://www.kompas.com/properti/read/2021/03/22/193836621/jawa-bali-dan-nusa-tenggara-wilayah-yang-sering-alami-krisis-air (accessed Nov. 09, 2021).

[4] J. J. Gourley and B. E. Vieux, "A method for evaluating the accuracy of quantitative precipitation estimates from a hydrologic modeling perspective," *J. Hydrometeorol.*, vol. 6, no. 2, pp. 115–133, Apr. 2005, doi: 10.1175/JHM408.1.

[5] R. J. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie, "CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution," *J. Hydrometeorol.*, vol. 5, no. 3, pp. 487–503, 2004, doi: 10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

[6] C. Kidd, D. R. Kniveton, M. C. Todd, and T. J. Bellerby, "Satellite rainfall estimation using combined passive microwave and infrared algorithms," *J. Hydrometeorol.*, vol. 4, no. 6, pp. 1088–1104, 2003, doi: 10.1175/1525-7541(2003)004<1088:SREUCP>2.0.CO;2.

[7] S. Sorooshian, K. L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, "Evaluation of PERSIANN system satellite-based estimates of tropical rainfall," *Bull. Am. Meteorol. Soc.*, vol. 81, no. 9, pp. 2035–2046, 2000, doi: 10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2.

[8] G. Huffman *et al.*, "NASA GPM Integrated Multi-satellitE Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD) Version 06," *Nasa/Gsfc*, no. January, p. 29, 2020, [Online]. Available: https://pmm.nasa.gov/sites/default/files/imce/times_allsat.jpg%0Ahttps://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06.pdf.

[9] U. Foelsche, G. Kirchengast, J. Fuchsberger, J. Tan, and W. A. Petersen, "Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria," *Hydrol. Earth Syst. Sci*, vol. 21, pp. 6559–6572, 2017, doi: 10.5194/hess-21-6559-2017.

[10] M. Yang, G. Liu, T. Chen, Y. Chen, and C. Xia, "Evaluation of GPM IMERG precipitation products with the point rain gauge records over Sichuan, China," *Atmos. Res.*, vol. 246, no. June, p. 105101, 2020, doi: 10.1016/j.atmosres.2020.105101.

[11] M. Arshad *et al.*, "Evaluation of GPM-IMERG and TRMM-3B42 precipitation products over Pakistan," *Atmos. Res.*, vol. 249, 2021, doi: 10.1016/j.atmosres.2020.105341.

[12] S. Moazami and M. R. Najafi, "A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada," *J. Hydrol.*, vol. 594, p. 125929, 2021, doi: 10.1016/j.jhydrol.2020.125929.

[13] L. Suryaningtyas, E. Suhartanto, and R. Rispiningtati, "Hydrological analysis of TRMM (Tropical Rainfall Measuring Mission) data in Lesti Sub watershed," *Civ. Environ. Sci.*, vol. 003, no. 01, pp. 018–030, 2020, doi: 10.21776/ub.civense.2020.00301.3.

[14] S. Wahyuni, R. Hambali, and F. Fahriani, "Evaluasi Ketelitian Data Curah Hujan Satelit GPM Terhadap Data Curah Hujan Permukaan," (Evaluation of the Accuracy of GPM Satellite Rainfall Data on Surface Rainfall Data) *Pertemuan Ilmiah Tahunan (PIT) HATHI - 37*. Universitas Bangka Belitung, pp. 1083–1094, 2020.

[15] N. A. Arrokhman, S. Wahyuni, and E. Suhartanto, "Evaluasi Kesesuaian Data Satelit untuk Curah Hujan dan Evaporasi Terhadap Data Pengukuran di Kawasan Waduk Sutami," (Evaluation of Suitability of Satellite Data for Rainfall and Evaporation Against Measurement Data in the Sutami Reservoir Area) *J. Teknol. dan Rekayasa Sumber Daya Air*, vol. 1, no. 2, pp. 904–916, 2021, doi: 10.21776/ub.jtresda.2021.001.02.46.

[16] A. Nomleni, E. Suhartanto, and D. Harisuseno, "Estimation of flow discharge model at Temef watershed-East Nusa Tenggara using TRMM satellite data," *civense*, vol. 4, no. 2, pp. 115–126, May 2021.

[17] R. Ramadhan *et al.*, "Evaluation of GPM IMERG Performance Using Gauge Data over Indonesian Maritime Continent at Different Time Scales," *Remote Sens.*, vol. 14, no. 5, p. 1172, 2022, doi: 10.3390/rs14051172.

[18] F. Simanjuntak, I. Jamaluddin, T. H. Lin, H. A. W. Siahaan, and Y. N. Chen, "Rainfall Forecast Using Machine Learning with High Spatiotemporal Satellite Imagery Every 10 Minutes," *Remote Sens.*, vol. 14, no. 23, p. 5950, Dec. 2022, doi: 10.3390/rs14235950.

[19] I. W. A. Yuda, T. Osawa, M. Nagai, and R. Prasetia, "The Regionalization of Indonesian Maritime Continent Rainfall based on Integrated Multi-satellite Retrievals for GPM (IMERG)," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 893, no. 1, 2021, doi: 10.1088/1755-1315/893/1/012065.

[20] D. F. B. Welkis *et al.*, "Evaluasi Pemanfaatan Data Pos Hujan Dan Data Satelit Trmm Dalam Pemodelan," (Evaluation of Utilization of Rain Post Data and Trmm Satellite Data in Modeling) *J. Tek. Sumber Daya Air*, vol. 2, no. 1, pp. 35–46, 2022.

[21] S. Y. Iryani, M. B. Al Amin, and A. Muhtarom, "Simulation of Rainfall Data by the GPM Satellite (Case Study at Sriwijaya University, Indralaya)," in *IOP Conference Series: Earth and Environmental Science*, 2021, vol. 832, no. 1, p. 12051, doi: 10.1088/1755-1315/832/1/012051.

[22] R. Ramadhan *et al.*, "Ground Validation of GPM IMERG-F Precipitation Products with the Point Rain Gauge Records on the Extreme Rainfall Over a Mountainous Area of Sumatra Island," *J. Penelit. Pendidik. IPA*, vol. 8, no. 1, pp. 163–170, Jan. 2022, doi: 10.29303/jppipa.v8i1.1155.

[23] I. J. A. Saragih *et al.*, "Improvement in WRF model prediction for heavy rain events over North Sumatra region using satellite data assimilation," in *IOP Conference Series: Earth and Environmental Science*, Nov. 2021, vol. 893, no. 1, p. 012040, doi: 10.1088/1755-1315/893/1/012040.

[24] B. Zhu, Z. Pu, B. Zhu, and Z. Pu, "Numerical Simulations of Diurnal Variations of Precipitation, Temperature, and Wind over the Coastal Area of Western Sumatra during a Madden-Julian Oscillation with Radar Data Assimilation," *AGUFM*, vol. 2019, pp. A43I-3057, 2019, Accessed: Dec. 22, 2022. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2019AGUFM.A43I3057Z/abstract.

[25] D. C. Goodrich, J. M. Faurès, D. A. Woolhiser, L. J. Lane, and S. Sorooshian, "Measurement and analysis of small-scale convective storm rainfall variability," *J. Hydrol.*, vol. 173, no. 1–4, pp. 283–308, 1995, doi: 10.1016/0022-1694(95)02703-R.

[26] R. Woods and M. Sivapaln, "A synthesis of space-time variability in storm response Rainfall runoff.pdf," *Water Resour. Res.*, vol. 35, no. 8, pp. 2469–2485, 1999, doi: 0043-1397/99/1999WR9000145.

[27] J. E. Nash and J. V Sutcliffe, "River flow forecasting through conceptual models part I - A discussion of principles," *J. Hydrol.*, vol. 10, no. 3, pp. 282–290, 1970, doi: 10.1016/0022-1694(70)90255-6.

[28] M. Bal, A. K. Dandpat, and B. Naik, "Hydrological modeling with respect to impact of land-use and land-cover change on the runoff dynamics in Budhabalanga river basing using ArcGIS and SWAT model," *Remote Sens. Appl. Soc. Environ.*, vol. 23, 2021, doi: 10.1016/j.rsase.2021.100527.

[29] D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations," *Trans. ASABE*, vol. 50, no. 3, pp. 885–900, 2007.

[30] K. V. SeethaRam, "Three Level Rule Curve for Optimum Operation of a Multipurpose Reservoir using Genetic Algorithms," *Water Resour. Manag.*, vol. 35, no. 1, pp. 353–368, 2021, doi: 10.1007/s11269-020-02738-7.

[31] D. . Goldberg and J. H. Holland, "Genetic Algorithms and Machine Learning," in *Machine Learning 3*, 1988, pp. 95–99.

[32] Soewarno, "Hidrologi - Aplikasi Metode Statistik untuk Analisa Data - Jilid II" (Hydrology - Application of Statistical Methods for Data Analysis - Series 2). Bandung, Indonesia: NOVA. 1995.