**Civil and Environmental Engineering**
**Vol. 22, Issue 1, 105-118, DOI: 10.2478/cee-2026-0008**

# Enhancing Daily Rainfall Data Completeness Using Satellite Rainfall Estimates

Rafika ANDARI[1,2], Nurhamidah NURHAMIDAH[2*], Darwizal DAOED[2], Marzuki MARZUKI[3]

[1] Civil Engineering Department, Universitas Andalas, Padang 25163, Indonesia

[2] Electrical Engineering Department, Institut Teknologi Padang, Padang 25173, Indonesia

[3] Physics Department, Universitas Andalas, Padang 25163, Indonesia

* corresponding author: nurhamidah@eng.unand.ac.id

Date of Submission: 22 June 2025
Revision Date: 28 July 2025
Date of Acceptance: 29 July 2025

Civil and Environmental Engineering
Journal of the Faculty of Civil Engineering | University of Žilina

**Abstract**
The availability of complete daily rainfall data is crucial for various hydrological, climatological, and meteorological studies. However, observed data often suffer from gaps due to equipment failures or station relocations. This study aims to improve the completeness of daily rainfall records by utilizing satellite rainfall estimates from TRMM, GPM-IMERG, and GSMaP products. To estimate the missing data, both linear regression and regional weighting methods were employed. A specific case study was conducted in the Kuranji watershed, Padang, Indonesia, using data spanning 2014 to 2016. The results indicate that the linear regression method based on satellite data, particularly GPM-IMERG, exhibits a higher correlation with observed data compared to the regional weighting method at most locations and across the study years. The results show that the regional weighting method was superior in 2014, while linear regression with GPM-IMERG and GSMaP was better in 2015 and 2016 in terms of accuracy, although regional weighting showed a lower relative bias. Nevertheless, the regional weighting method demonstrated a lower relative bias. This study highlights the potential of satellite-based rainfall data as a viable alternative for filling missing rainfall records. However, further calibration and validation are necessary to address spatial variability and inaccuracies during high-intensity rainfall events.

**Keywords**
Missing rainfall data; Linear regression; Regional weighting; Satellite.

## 1. Introduction

Rainfall data availability is the primary basis for hydrological, climatological, and metrological investigations (Qutbudin et al., 2019). The complex hydrological mechanisms that explain how rainfall and flow interact are demonstrated by the various paths that rainwater takes, including surface runoff and subterranean flow (Halecki et al., 2023). Hydrological characteristics, such as water flow direction, river networks, and basin morphology, are closely related to the availability of accurate rainfall and hydrological data (Abd Elrahman & Ataalmanan, 2023). The hydrological cycle and precipitation trends (rainfall) are one of the periodic seasonal processes that influence the amount of water resources in a region, which is part of the interconnected subsystems of the Earth (Smrčková et al., 2024). However, rainfall data collection often faces significant challenges, such as missing data. This phenomenon can be caused by various factors, ranging from station relocation due to urbanization (Liu & Niyogi, 2019), flaws in rainfall measurement methodologies, or equipment failure over time (Jeřábek & Kavka, 2024), particularly in flood-prone locations (Armanuos et al., 2020), (Wang et al., 2023). Missing

© Author(s) 2026. This work is distributed under the Creative Commons BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).

values in rainfall data series can compromise the analytical output of hydrological and meteorological models (Duarte et al., 2022). Therefore, estimating missing values and filling these gaps has become crucial in contemporary hydrological research (Armanuos et al., 2020), (Hamzah et al., 2021).

Previous research employed rudimentary statistical methodologies, such as arithmetic averaging, to address the issue of missing data. The validity of the arithmetic mean method hinges on the stipulation that the average annual rainfall at adjacent stations must be within 10% of the target station's standard annual rainfall. In the event that this criterion is not met, the arithmetic mean approach is deemed unfeasible to calculate the missing rainfall at the target station (Caldera et al., 2016), (Adilah & Hannani, 2021). However, algorithms, such as those used in artificial intelligence and data-driven machine learning, have emerged as more accurate approaches. These advanced methods utilize rainfall data obtained from neighboring stations within the same region and at the same time, taking into account stations with similar climatic conditions, irrespective of precipitation events (Sattari et al., 2017).

To create complete datasets, methods for filling missing rainfall values are used to impute model-based rainfall estimates into the incomplete time series. Popular approaches include regional vectors, linear regression models, and the regional weighting method (Wangwongchai et al., 2023). The inverse distance weighting approach (Teegavarapu & Chandramouli, 2005), the gauge mean estimator (Strachan et al., 2016), and ordinary Kriging (Borga & Vizzaccaro, 1997), (Plouffe et al., 2015), (Xu et al., 2018) have also become more popular due to their simplicity and convenience. These methods, albeit having different mathematical formulations, are based on a common logic. Using data from nearby locations facilitates the estimation of missing values at the target location by "averaging" these values (Rohma, 2022). According to (Duarte et al., 2022), these techniques are better suited for estimating monthly and annual rainfall levels because of the significant spatial variability and intermittent character of the rainfall process at fine temporal scales. The aggregation process lessens small timeframe extremes and dry spells, which efficiently smoothest out the higher-resolution time series. Consequently, there may be a significant decrease in estimation errors (Beran et al., 2020). Nonetheless, hydrological modeling and rainfall frequency analysis depend on daily and sub-daily precipitation data. Models that fill in missing information at such short intervals are typically limited and have poor predictive powers because of the high degrees of unpredictability in the process (Hasan & Croke, 2013).

A common practice to address this problem is to remove years (or other time periods) with incomplete data. Consequently, the random variable distribution's upper tail might not be adequately described, or extremely tiny samples might be produced (Hristopulos & Baxevani, 2020). If the modeling mistakes are acceptable—which is rarely the case with daily data—an additional choice is to approximate the missing values using regression-based techniques. Multiple linear regression (Nathans et al., 2012), simple substitution (Pappas et al., 1955), and machine-learning algorithms (Shin et al., 2021) have all been employed for this purpose. These approaches rely on the ability of the measuring network to reproduce the observed rainfall fields, especially those with intricate spatial patterns and the more intense events. Therefore, further research is needed to address these deficiencies in daily rainfall data (Longman et al., 2020).

Rainfall estimates obtained from remote-sensing data sources, including satellites, have been more and more popular in recent years among researchers and practitioners worldwide, particularly in areas with inadequate density gauging (Brocca et al., 2014). In actuality, satellites offer end users inexpensive costs, high sample frequencies, enhanced spatial resolution at huge catchments or areas, and products free from missing data corruption (Haloho & Supriyadi, 2024). This feature has led to the question of whether using satellite rainfall estimates is a more accurate and efficient way to fill in the gaps in rainfall gauging stations that are surrounded by satellite pixels than using data from nearby locations.

The rationale for completing missing rainfall values over longer periods, such as monthly or seasonal, has been addressed in recent studies (Siabi et al., 2020), (de Moraes Cordeiro & Blanco, 2021), (Abu Romman et al., 2021). However, to the best of our knowledge, the literature has not thoroughly examined the use of satellite retrievals to supplement daily rainfall data. The primary objective of this study is to determine the feasibility of directly imputing remote sensing data from TRMM, GPM-IMERG, and GSMaP satellites to address the research gap caused by the absence of rain gauge data.

## 2. Methodology

### 2.1. Study Site

This study was conducted in the Kuranji watershed in Padang, West Sumatera, Indonesia, located at 00°48'-00°56' N and 100°20'-100°34' E along Sumatra's western coast. The upper portion of the Kuranji watershed borders Padang City and Solok Regency. The Kuranji watershed's elevation ranges from 1858 meters above sea level to an average altitude of 215.615 square kilometers (Andari et al., 2024b). Figure 1 shows a map of the Kuranji Watershed.

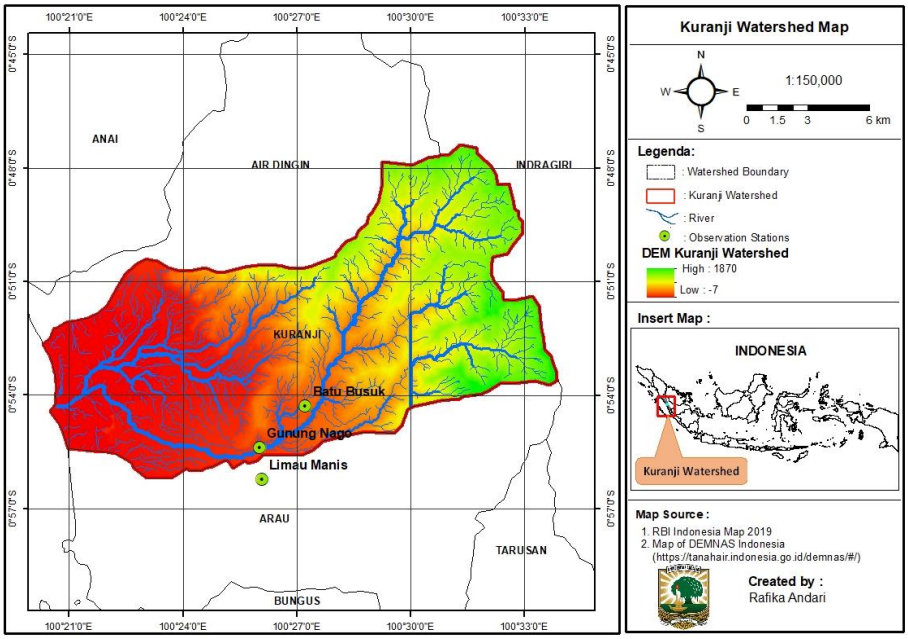



**Figure 1:** Map of The Kuranji Watershed (Source : Authors)

### 2.2. Data Collection

Daily rainfall data was collected from gauging stations within the Kuranji watershed for this study. Specifically, data were collected from three observation sites: Limau Manis, Batu Busuk, and Gunung Nago. These daily precipitation records, obtained between 2014 and 2016, were provided by the Dinas Pengelolaan Sumber Daya Air (PSDA)/Water Resources Management Department, West Sumatera.

This study assesses satellite rainfall products from the TRMM, GPM-IMERG, and GSMaP datasets. On November 27, 1997, the NASA-JAXA collaborative project TRMM was initiated. Its main goals were to support research on global climate change and to track and analyze tropical precipitation (de Moraes Cordeiro & Blanco, 2021), (Rozante et al., 2018). Specifically, TRMM rainfall data was retrieved from product 3B42, version 7, using the Giovanni platform (GESDISC Online Visualization and Analysis Infrastructure). This platform is accessible via https://giovanni.gsfc.nasa.gov/. A geographic resolution of 0.25° x 0.25° was used to collect and evaluate the TRMM's daily results (Elnashar et al., 2020).

The GPM-IMERG satellite was developed by NASA and JAXA to replace the TRMM satellite, which only monitored tropical rainfall, and launched in 2014. GPM-IMERG produces precipitation data with temporal resolution up to 30 min and a spatial resolution of 0.1◦ × 0.1◦ (Ramadhan, et al., 2022). The GPM-IMERG satellite produces three main products (IMERG Early, IMERG Late, and IMERG Final) with varying degrees of precision and temporal delay, which are very useful for hydrological analysis, especially in areas without ground-based precipitation measuring equipment (Zhou et al., 2020). The

primary distinction between the three data kinds is the amount of time it takes for the data to become available; early and late data have a 4 and 12 hour delay, respectively, while final data has a 2.5–3.5 month wait (Ramadhan, et al., 2022).

GSMaP was created in partnership with PMM, the Japan Aerospace Exploration Agency's Precipitation Measuring Mission. For the GSMaP project, the PMM team produced three precipitation datasets. There are numerous products available, such as GSMaP-MVK, GSMaP-Gauge, and GSMaP-NRT (Ramadhan et al., 2023), (Nepal et al., 2021).

## 2.3. Methods for Estimating Missing Rainfall Data

Daily rainfall data from the rain gauges were estimated using two methods for addressing missing data: linear regression and regional weighting (dos Santos et al., 2021). The data from the rain gauges was used to apply the regional weighting approach. The data from the rain gauges and the satellite pixels containing the analysis gauges were utilized independently for the linear regression, allowing the estimations to be compared.

## 2.4. Linear Regression Method

In hydrology, simple and multiple linear regression techniques are frequently employed to estimate monthly and annual hydrological data (Elshorbagy et al., 2000). At each measurement site with comparable weather conditions, the missing weather data is estimated using this statistical method (Armanuos et al., 2020). There are two parts involved in this method: first, the relationship between the predictor and the missing value is estimated, and then the gap is filled using a trend equation (Yozgatligil et al., 2013). The following equation is used to estimate data (Armanuos et al., 2020).

$$y = a + b.x \quad (1)$$

Where:

*a - the linear coefficient,*

*b - the angular coefficient,*

*y - the projected rainfall data,*

*x - the observed rainfall data.*

## 2.5. Regional Weighting Method

The regional weighting method is a widely used approach for estimating missing or average rainfall values by leveraging data from nearby rain gauges. This method improves the completeness and reliability of rainfall datasets, which is crucial for hydrological and climate studies. The missing or average rainfall value is calculated as a weighted average of the available data from neighboring stations, giving more influence to stations that are closer or more similar to the target location (Byun et al., 2024; Gomes et al., 2018; Nathans et al., 2012). The estimation by regional weighting was made by equation:

$$p(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{N_y}{N_x}\,x \quad (2)$$

Where:

*p(x)* - *the estimated monthly rainfall,*

*x* - *the rainfall measurement that was taken during the same time period as the missing data in p(x),*

$N_x$ *and* $N_y$ - *stand for the average annual rainfall at nearby gauges and the rain gauge with missing data,*

*N* - *the number of surrounding gauges that were used in the weighting process.*

### 2.6. Comparison of Methods

To estimate missing rainfall data, two methods were employed: the linear regression method and the regional weighting method. For statistical comparison, four coefficients were utilized: correlation coefficient (R), root mean square error (RMSE), Nash-Sutcliffe Efficiency (NSE), and relative bias (RB). The evaluation will proceed according to the formula provided below.

$$R = \frac{n \sum S_i G_i - \sum S_i \sum G_i}{\sqrt{[n \sum {S_i}^2 - (\sum S_i)^2][n \sum {G_i}^2 - (\sum G_i)^2]}} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (S_i - G_i)^2}{n}} \quad (4)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (Si - Gi)^2}{\sum_{i=1}^{n} (Si - \overline{Gi})^2} \quad (5)$$

$$RB = \frac{\sum_{i=1}^{n} (S_i - G_i)}{\sum_{i=1}^{n} G_i} \quad (6)$$

Where:
*$G_i$ - the data from the observed rain gauge,*
*$S_i$ - the amount of rainfall as determined by the satellite,*
*n - the number of samples.*

The correlation coefficient (R) quantifies the relationship between measured and estimated ground precipitation data. The closer the data aligns with gauge measurements, the higher the R-value, ranging from 0 to 1, with 1 being optimal. The accuracy of simulated and observed data calculations is assessed using the RMSE, which ranges from 0 to positive infinity. Higher estimation accuracy is indicated by RMSE values closer to zero (de Moraes Cordeiro & Blanco, 2021), (dos Santos et al., 2021).

NSE is used because it complements the evaluation of model fit. NSE can accurately determine the presence and extent of bias in a model, indicating whether it underestimates or overestimates (McCuen et al., 2006). The estimation result is considered good if the Nash-Sutcliffe Efficiency coefficient (NSE) exceeds 0.75, and adequate if the NSE is less than 0.75 (Motovilov et al., 1999), (Van Liew et al., 2007). Conversely, the disparity in relative volume between the estimated and observed volumes is determined by the relative bias (RB). A negative value indicates underestimation, while a positive value indicates overestimation (Ramadhan et al., 2023).

## 3. Results

### 3.1. Analysis of Missing Rainfall Filling Method

This analysis uses daily rainfall measurements from 2014, 2015, and 2016. Figure 2 compares the distribution of daily rainfall data from satellite data with 2014 data from the Batu Busuk observation station.

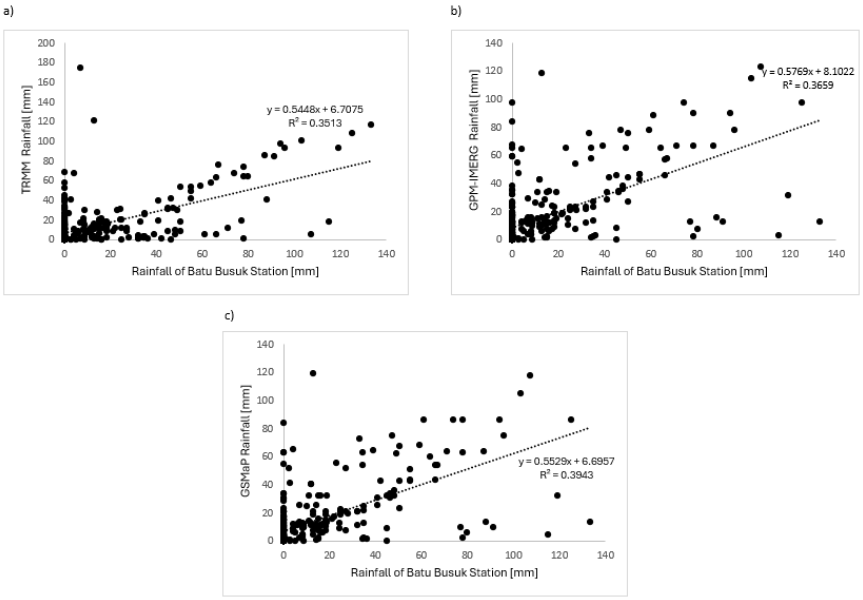



**Figure 2:** Distribution graph between rainfall at Batu Busuk Station and rainfall predicted by satellite data in 2014, a) TRMM satellite, b) GPM-IMERG satellite, and c) GSMaP satellite

The linear regression graphs in Figure 2 show the relationship between rainfall measurements at Batu Busuk Station and rainfall estimates from satellite data (TRMM, GPM-IMERG, and GSMaP). These graphs show varying levels of accuracy. While all three indicate a positive relationship, the coefficients of determination ($R^2$) are low: 0.351 for TRMM, 0.366 for GPM-IMERG, and 0.362 for GSMaP. These $R^2$ values suggest that the linear regression method explains only 35-36% of the data's variation. This suggests that satellite estimates account for only a small portion of the observed rainfall variability, indicating significant inaccuracies, especially for high-intensity events. These inaccuracies may be due to factors such as limited spatial resolution, topographic influences, and variable meteorological conditions in the study area.

Figures 3 and 4 show the linear regression graphs and coefficients of determination ($R^2$) obtained during the estimation of missing rainfall data for the Limau Manis and Gunung Nago observation stations, respectively.

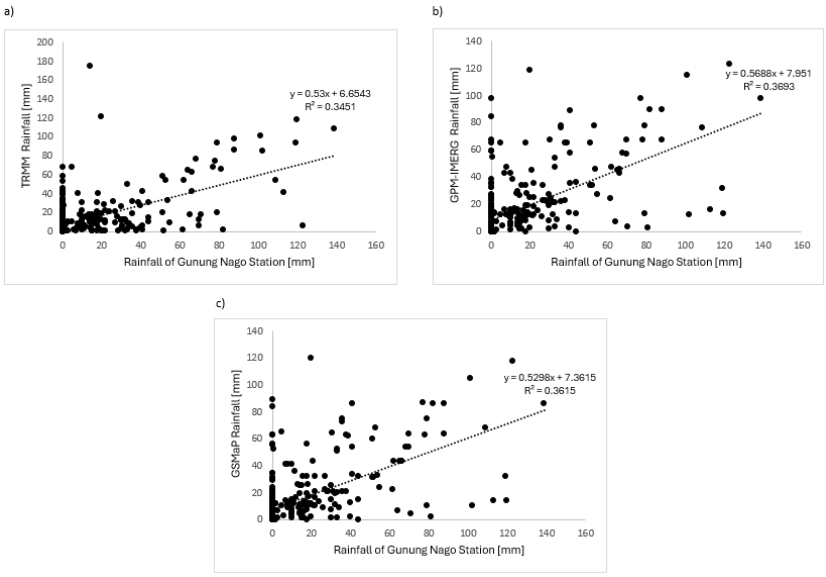



**Figure 3:** Distribution graph between rainfall at Gunung Nago Station and rainfall predicted by satellite data in 2014, a) TRMM satellite, b) GPM-IMERG satellite, and c) GSMaP satellite

The regression analysis in Figure 3 reveals a positive linear relationship between the three satellite data sources and the rainfall data from Gunung Nago Station. The $R^2$ values vary, indicating that the linear regression method explains 34-36% of the data's variation. The GPM-IMERG satellite shows the strongest correlation, suggesting its rainfall estimates are more accurate than those of TRMM and GSMaP for this station.

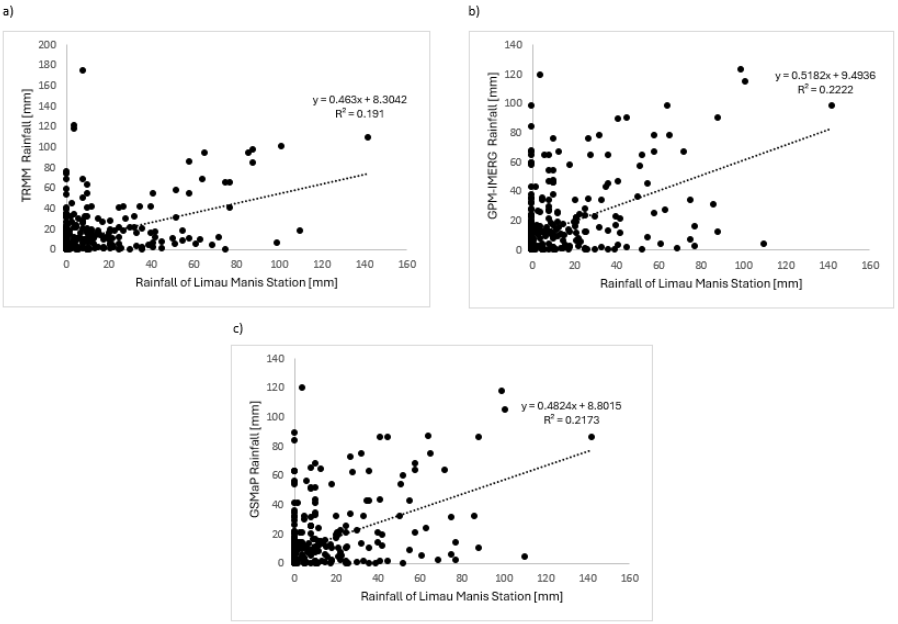



**Figure 4:** Distribution graph between rainfall at Limau Manis Station and rainfall predicted by satellite data in 2014, a) TRMM satellite, b) GPM-IMERG satellite, and c) GSMaP satellite

As shown in Figure 4, the regression graphs for the three satellites demonstrate that the rainfall data has a lower correlation than the data from the Batu Busuk and Gunung Nago stations. At the Limau Manis station, this linear model explains 19-22% of the rainfall data variability.

Similarly, the estimation procedure was applied to daily rainfall data from 2015 and 2016. The recalculated regression equations and $R^2$ values for each year are presented in Tables 1 and 2, respectively.

Table 1 shows that GPM-IMERG and GSMaP beat TRMM in estimating lost rainfall in 2015, with higher $R^2$ values. The Batu Busuk station consistently demonstrated higher $R^2$ values, indicating more accurate rainfall estimations. These disparities are due to variances in satellite algorithms, resolution, and sensitivity to environmental factors.

**Table 1:** Recapitulation of Linear Regression Equations in Estimating Missing Rainfall Data in 2015

| Observation Station | Satellite data | Linear Regression Equation | $R^2$ [-] |
|---|---|---|---|
| Batu Busuk | TRMM | y = 0.4335x + 5.6047 | 0.3089 |
| | GPM-IMERG | y = 0.4357x + 7.8819 | 0.2985 |
| | GSMaP | y = 0.4505x + 8.6002 | 0.2961 |
| Gunung Nago | TRMM | y = 0.206x + 7.8342 | 0.0788 |
| | GPM-IMERG | y = 0.2304x + 9.9243 | 0.0944 |
| | GSMaP | y = 0.2411x + 10.688 | 0.0959 |
| Limau Manis | TRMM | y = 0.2778x + 7.6487 | 0.0784 |
| | GPM-IMERG | y = 0.3133x + 9.6995 | 0.0954 |
| | GSMaP | y = 0.3262x + 10.464 | 0.0960 |

As shown in Table 2, the linear regression equation for estimating missing rainfall data in 2016 reveals that the Batu Busuk station has a higher $R^2$ value than the other two stations. The GPM-IMERG satellite has the highest $R^2$ value (0.5021). In contrast, the linear regression model for the Gunung Nago station has a relatively low $R^2$ value (below 0.2), suggesting it explains less of the rainfall data variation. The $R^2$ value for this station is also low, with the highest recorded value (0.2630) measured by the GPM-IMERG satellite.

**Table 2:** Recapitulation of Linear Regression Equations in Estimating Missing Rainfall Data in 2016

| Observation Station | Satellite data | Linear Regression Equation | $R^2$ [-] |
|---|---|---|---|
| Batu Busuk | TRMM | y = 0.4595x + 4.6877 | 0.4659 |
| | GPM-IMERG | y = 0.7002x + 7.8673 | 0.5021 |
| | GSMaP | y = 0.5752x + 6.8665 | 0.4637 |
| Gunung Nago | TRMM | y = 0.141x + 8.1554 | 0.2421 |
| | GPM-IMERG | y = 0.1679x + 13.905 | 0.1592 |
| | GSMaP | y = 0.1456x + 11.703 | 0.1639 |
| Limau Manis | TRMM | y = 0.3968x + 5.9521 | 0.2292 |
| | GPM-IMERG | y = 0.5874x + 9.9886 | 0.2331 |
| | GSMaP | y = 0.5334x + 8.0376 | 0.2630 |

## 3.2. Comparative Analysis of Missing Rainfall Method

Tables 3-5 present a comparison of the missing rainfall data filling approaches, showing the adjusted regression models and estimations produced by the linear regression and regional weighting methods.

**Table 3:** Performance Assessment Results of the Missing Rainfall Filling Method in 2014

| Station Analyzed | Parameter Statistics | Linear Regression Method | | | Regional Weighting Method |
|---|---|---|---|---|---|
| | | TRMM | GPM-IMERG | GSMaP | |
| Batu Busuk | R [-] | 0.75 | 0.76 | 0.75 | 0.93 |
| | RMSE [-] | 16.94 | 16.17 | 16.77 | 8.89 |
| | NSE [-] | 0.50 | 0.54 | 0.51 | 0.86 |
| | RB [%] | 40.16 | 24.47 | 32.52 | 0.01 |
| Gunung Nago | R [-] | 0.74 | 0.76 | 0.76 | 0.92 |
| | RMSE [-] | 17.57 | 16.61 | 17.27 | 9.90 |
| | NSE [-] | 0.48 | 0.54 | 0.50 | 0.84 |
| | RB [%] | 43.54 | 27.77 | 36.04 | -0.03 |
| Limau Manis | R [-] | 0.62 | 0.65 | 0.66 | 0.78 |
| | RMSE [-] | 16.53 | 15.76 | 15.80 | 13.72 |
| | NSE [-] | 0.37 | 0.43 | 0.42 | 0.57 |
| | RB [%] | 25.49 | 9.32 | 20.12 | 0.02 |

Table 3 demonstrates the analysis of the performance of the missing rainfall filling method in 2014. The Regional Weighting method consistently demonstrated superior performance across three stations, exhibiting the highest correlation (R), efficiency (NSE), and lowest relative bias (RB). This indicates its strong ability to accurately capture rainfall patterns. While the Linear Regression method showed comparable performance at most stations, it generally had lower accuracy and higher bias. Specifically, at Batu Busuk, Regional Weighting achieved an R-value of 0.93 and an NSE of 0.86. Similarly, at Gunung Nago, it excelled with an R-value of 0.92, an NSE of 0.84, and the lowest RB. At Limau Manis, the Regional Weighting method continued its strong performance with an R-value of 0.78 and an NSE of 0.57. Overall, the Regional Weighting method proved superior to the Linear Regression method in capturing rainfall patterns at all three observed stations.

**Table 4:** Performance Assessment Results of the Missing Rainfall Filling Method in 2015

| Station Analyzed | Parameter Statistics | Linear Regression Method | | | Regional Weighting Method |
|---|---|---|---|---|---|
| | | TRMM | GPM-IMERG | GSMaP | |
| Batu Busuk | R [-] | 0.78 | 0.78 | 0.77 | 0.63 |
| | RMSE [-] | 16.20 | 15.25 | 14.86 | 18.39 |
| | NSE [-] | 0.40 | 0.46 | 0.49 | 0.23 |
| | RB [%] | 53.89 | 42.87 | 36.61 | -0.03 |
| Gunung Nago | R [-] | 0.63 | 0.65 | 0.64 | 0.74 |
| | RMSE [-] | 19.93 | 19.04 | 18.72 | 22.23 |
| | NSE [-] | 0.20 | 0.27 | 0.29 | -0.01 |
| | RB [%] | 62.96 | 52.97 | 48.36 | -0.01 |
| Limau Manis | R [-] | 0.77 | 0.80 | 0.75 | 0.67 |
| | RMSE [-] | 13.21 | 11.98 | 11.92 | 19.04 |
| | NSE [-] | 0.35 | 0.47 | 0.47 | -0.34 |
| | RB [%] | 58.89 | 47.21 | 40.01 | 0.04 |

Based on Table 4, the Linear Regression method, utilizing both GPM-IMERG and GSMaP satellite data, generally exhibited superior performance compared to the Regional Weighting method in terms of correlation (R), RMSE, and NSE. However, the Regional Weighting method excelled in relative bias (RB), showing values closer to zero, particularly at the Batu Busuk and Limau Manis stations. The highest R-value (0.80) was observed for the Linear Regression method using GPM-IMERG satellite data at Limau Manis station, while the lowest (0.63) came from the Regional Weighting method. Although Regional Weighting recorded the highest R-value (0.74) at Gunung Nago station, RMSE analysis indicated that Linear Regression with GPM-IMERG and GSMaP satellites had smaller prediction errors across all stations. Furthermore, the highest NSE values were achieved by the Linear Regression method, with GSMaP reaching 0.49 at Batu Busuk, whereas the Regional Weighting method showed negative NSE values at Gunung Nago (-0.01) and Limau Manis (-0.34). Conversely, the Regional Weighting method demonstrated a very small relative bias compared to the Linear Regression method, which tended to overestimate with significantly positive RB values, such as TRMM reaching 53.89%. These results confirm that while the Linear Regression method is more accurate in representing rainfall patterns, the Regional Weighting method is more reliable in terms of bias estimation.

**Table 5:** Performance Assessment Results of the Missing Rainfall Filling Method in 2016

| Station Analyzed | Parameter Statistics | Linear Regression Method | | | Regional Weighting Method |
|---|---|---|---|---|---|
| | | TRMM | GPM-IMERG | GSMaP | |
| Batu Busuk | R [-] | 0.84 | 0.85 | 0..83 | 0.66 |
| | RMSE [-] | 19.60 | 13.27 | 15.94 | 22.83 |
| | NSE [-] | 0.38 | 0.71 | 0.59 | 0.16 |
| | RB [%] | 62.51 | 14.90 | 36.95 | 0.05 |
| Gunung Nago | R [-] | 0.50 | 0.45 | 0.47 | 0.55 |
| | RMSE [-] | 57.24 | 55.57 | 56.37 | 49.08 |
| | NSE [-] | 0.04 | 0.10 | 0.07 | 0.30 |
| | RB [%] | 75.40 | 56.40 | 65.39 | 0.04 |
| Limau Manis | R [-] | 0.68 | 0.67 | 0.70 | 0.56 |
| | RMSE [-] | 16.82 | 15.33 | 14.67 | 23.80 |
| | NSE [-] | 0.31 | 0.42 | 0.47 | -0.39 |
| | RB [%] | 50.95 | -2.25 | 19.50 | -0.05 |

As shown in Table 5, the linear regression method, particularly when using satellite precipitation products such as GPM-IMERG and GSMaP, consistently outperforms the regional weighting method. At Batu Busuk Station, linear regression with GPM-IMERG had the highest correlation (R) of 0.85 and the lowest root mean square error (RMSE) of 13.27. This method also had good Nash-Sutcliffe efficiency (NSE) of 0.71, though it had moderate relative bias (RB) of 14.90% overestimation. At Gunung Nago Station, all methods showed a decline in performance. However, Linear Regression with GSMaP was slightly

better in terms of R (0.55) and RMSE (56.37). It should be noted that the RB was very high (up to 75.40%), indicating significant overestimation. This is in contrast to the minimal bias in the Regional Weighting Method (0.004%). At Limau Manis Station, Linear Regression with GSMaP was the best method, showing the highest R (0.70) and relatively low RMSE (14.67). GPM-IMERG Linear Regression showed slightly more desirable underestimation (RB = -2.25%), while Regional Weighting Method showed poor predictive performance, with negative NSE. Thus, although satellite-based linear regression methods offer higher accuracy, spatial variations must be considered, and careful validation and calibration are necessary to address potential location-based biases.

An evaluation of the overall statistical parameters for the 2014–2016 data revealed a fairly strong correlation in R values across all methods and locations, though variations occurred between years and locations. The linear regression method using the GPM-IMERG satellite consistently produced the highest R values in nearly all locations and years, suggesting greater predictive accuracy than other methods. Regional weighting produced relatively lower R values, suggesting that it may be less effective than linear regression in this context. Variation in R values among locations (Limau Manis, Gunung Nago, and Batu Busuk) highlights the influence of spatial characteristics on satellite product performance.

## 4. Discussion

### 4.1. Interpretation of Results

A comprehensive statistical evaluation shows relatively strong correlations (R values) for all methods and locations, although variations exist between years and locations. The linear regression method using the GPM-IMERG satellite consistently produces the highest R values in most locations and years, indicating better prediction accuracy compared to other methods. In contrast, regional weighting shows comparatively lower R values, suggesting it may be less effective than linear regression in this context. The variation in R-values between locations (Batu Busuk, Limau Manis, and Gunung Nago) suggests that spatial characteristics influence the performance of the satellite data (Andari et al., 2024b).

Overestimation or underestimation of satellite estimations may be caused by the spatial variability of weather systems and potential errors in rain gauge data (de Moraes Cordeiro & Blanco, 2021). The frequency of short-lived convective complexes, which occur within a few hours and are not always fully captured by satellites, contributes to precipitation forecast inaccuracy (Soares et al., 2016). Thus, discrepancies between rainfall data and satellite estimates arise from various factors, including errors in precipitation estimation algorithms, errors in algorithms used to aggregate estimates, satellite sampling limitations, and errors in bias correction algorithms for rain gauges (dos Santos et al., 2021), (Andari et al., 2024a).

Compared to prior research using statistical methods for filling missing rainfall data, such as (Portuguez ‐ maurtua et al., 2022) and (Camuffo et al., 2022), this study found slightly lower Root Mean Square Error (RMSE) and RB values, with sufficiently high correlation values for both methods. This discrepancy may be due to variations in climatic conditions, such as alternating wet and dry periods, and the occurrence of significant rainfall events, which can affect the direct comparison of results across different regions is discrepancy could be attributed to variations in climatic conditions, characterized by alternating periods of wet and dry weather, as well as the occurrence of substantial rainfall events. These factors potentially influence the direct comparison of results across different geographical regions (Duarte et al., 2022).

Completing missing data is often a complex and time-consuming task, but it is essential for effective water resource planning and management. Satellite data has shown similar behavior to data from nearby rain gauges in estimating missing hydrological data within the study area, indicating its potential as a data source for analysis (Bárdossy & Pegram, 2014).

The utilization of satellite data to address the deficit in observational rainfall data carries profound ramifications for hydrological studies, weather forecasting, and climate monitoring. Satellite data like GPM-IMERG and GSMaP offer high spatial resolution, improving the accuracy of rainfall estimates across geographic areas (de Moraes Cordeiro & Blanco,

2021). This is particularly useful in areas with complex terrain or inadequate rain gauge networks (Chua et al., 2023). Then, satellite data is an effective solution for addressing deficiencies in observational datasets, especially when conventional methods are ineffective due to constraints such as instrument failures, maintenance issues, or natural barriers like mountains that impede radar observations (Lutfiah et al., 2024).

### 4.2. Limitations

While this study successfully demonstrates the potential of both linear regression and regional weighting methods, particularly with GPM-IMERG satellite data, for addressing missing daily rainfall data in the Kuranji watershed, several limitations affect the broader interpretation of its findings. The inherent spatial resolution limitations of satellite data, coupled with topographic influences and variable meteorological conditions in the study area, can lead to inaccuracies, especially for high-intensity rainfall events. Furthermore, potential errors within the satellite's precipitation estimation and aggregation algorithms, as well as bias correction algorithms for rain gauges, contribute to discrepancies between observed and estimated rainfall. The challenge of accurately filling daily and sub-daily rainfall data is also acknowledged due to its high unpredictability. Finally, the generalizability of these results to other regions may be limited given the influence of varying climatic conditions.

## 5. Conclusion

This study demonstrates that satellite-based rainfall estimation, especially using the GPM-IMERG product, is effective for filling gaps in daily rainfall data in the Kuranji watershed, Padang. The linear regression method utilizing satellite data achieved higher correlation and Nash-Sutcliffe efficiency across most observation stations compared to the regional weighting approach, although the latter showed superior performance in minimizing relative bias. Variations in method performance across locations underscore the importance of considering spatial characteristics and local meteorological conditions. The findings imply that integrating satellite rainfall data can enhance the quality and completeness of rainfall records, which is critical for water resource management, disaster mitigation, and hydrological research, particularly in regions with limited ground-based observations. Future research should focus on developing more adaptive calibration and validation techniques to improve accuracy during intense rainfall and diverse climatic conditions.

## Acknowledgements

Sincere gratitude and admiration to the dissertation and paper's supervisors, as well as the Universitas Andalas faculty and personnel.

## Author Contributions

**R.A.** designed the research, collected and analyzed the data, and wrote the initial manuscript. **N.N.** performed data curation, formal analysis, and contributed to manuscript writing – review & editing. **D.D**. assisted with data analysis, validated the results. **M.M.** contributed to data interpretation and critically reviewed the manuscript. All authors critically reviewed and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

## Disclosure of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

## References


Qutbudin, I., Shiru, M. S., Sharafati, A., Ahmed, K., Al-Ansari, N., Yaseen, Z. M., Shahid, S., & Wang, X. (2019). Seasonal drought pattern changes due to climate variability: Case study in Afghanistan. *Water (Switzerland)*, *11*(5). https://doi.org/10.3390/w11051096.

Halecki, W., Młyńska, A., Sionkowski, T., & Chmielowski, K. (2023). Linking elevated rainfall with sewage discharge volume. *Ochrona Srodowiska i Zasobow Naturalnych*, *34*(4), 135–146. https://doi.org/10.2478/oszn-2023-0020.

Abd Elrahman, S. I. M., & Ataalmanan, I. M. I. (2023). Determination of the Hydrological and Morphometric Characteristics Using GIS. *Civil and Environmental Engineering*, *19*(1), 39–47. https://doi.org/10.2478/cee-2023-0004.

Smrčková, D., Chromčák, J., Mužík, J., Ižvoltová, J., & Kostelecký, J. (2024). Space Geodesy Data Implementation for Earth System Geodynamics Monitoring: a Case Study of the Aegean Microplate. *Civil and Environmental Engineering*, *20*(2), 1203–1220. https://doi.org/10.2478/cee-2024-0088.

Liu, J., & Niyogi, D. (2019). Meta-analysis of urbanization impact on rainfall modification. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-019-42494-2.

Jeřábek, J., & Kavka, P. (2024). Sensitivity and uncertainty analysis of a surface runoff model using ensemble of artificial rainfall experiments. *Journal of Hydrology and Hydromechanics*, *72*(4), 466–485. https://doi.org/10.2478/johh-2024-0021.

Armanuos, A. M., Al-Ansari, N., & Yaseen, Z. M. (2020). Cross assessment of twenty-one different methods for missing precipitation data estimation. *Atmosphere*, *11*(4), 1–35. https://doi.org/10.3390/ATMOS11040389.

Wang, X., Shi, S., Zhu, L., Nie, Y., & Lai, G. (2023). Traditional and Novel Methods of Rainfall Observation and Measurement: A Review. *Journal of Hydrometeorology*, *24*(12), 2153–2176. https://doi.org/10.1175/JHM-D-22-0122.1.

Duarte, L. V., Formiga, K. T. M., & Costa, V. A. F. (2022). Comparison of Methods for Filling Daily and Monthly Rainfall Missing Data: Statistical Models or Imputation of Satellite Retrievals? *Water (Switzerland)*, *14*(19). https://doi.org/10.3390/w14193144.

Hamzah, F. B., Hamzah, F. M., Razali, S. F. M., & Samad, H. (2021). A comparison of multiple imputation methods for recovering missing data in hydrological studies. *Civil Engineering Journal (Iran)*, *7*(9), 1608–1619. https://doi.org/10.28991/cej-2021-03091747.

Caldera, H. P. G. M., Piyathisse, V. R. P. C., & Nandalal, K. D. W. (2016). A Comparison of Methods of Estimating Missing Daily Rainfall Data. *Engineer: Journal of the Institution of Engineers, Sri Lanka*, *49*(4), 1. https://doi.org/10.4038/engineer.v49i4.7232.

Adilah, N. A. A. G., & Hannani, H. (2021). Comparison of Methods to Estimate Missing Rainfall Data for Short Term Period at UMP Gambang. *IOP Conference Series: Earth and Environmental Science*, *682*(1). https://doi.org/10.1088/1755-1315/682/1/012027.

Sattari, M. T., Rezazadeh-Joudi, A., & Kusiak, A. (2017). Assessment of different methods for estimation of missing data in precipitation studies. *Hydrology Research*, *48*(4), 1032–1044. https://doi.org/10.2166/nh.2016.364.

Wangwongchai, A., Waqas, M., Dechpichai, P., Hlaing, P. T., Ahmad, S., & Humphries, U. W. (2023). Imputation of missing daily rainfall data; A comparison between artificial intelligence and statistical techniques. *MethodsX*, *11*(October), 102459. https://doi.org/10.1016/j.mex.2023.102459.

Teegavarapu, R. S. V., & Chandramouli, V. (2005). Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, *312*(1–4), 191–206. https://doi.org/10.1016/j.jhydrol.2005.02.015.

Strachan, S., Kelsey, E. P., Brown, R. F., Dascalu, S., Harris, F., Kent, G., Lyles, B., McCurdy, G., Slater, D., & Smith, K. (2016). Filling the Data Gaps in Mountain Climate Observatories Through Advanced Technology, Refined Instrument Siting, and a Focus on Gradients. *Mountain Research and Development*, *36*(4), 518–527. https://doi.org/10.1659/MRD-JOURNAL-D-16-00028.1.

Borga, M., & Vizzaccaro, A. (1997). On the interpolation of hydrologic variables: Formal equivalence of multiquadratic surface fitting and kriging. *Journal of Hydrology*, *195*(1–4), 160–171. https://doi.org/10.1016/S0022-1694(96)03250-7.

Plouffe, C. C. F., Robertson, C., & Chandrapala, L. (2015). Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka. *Environmental Modelling and Software*, *67*, 57–71. https://doi.org/10.1016/j.envsoft.2015.01.011.

Xu, P., Wang, D., Singh, V. P., Wang, Y., Wu, J., Wang, L., Zou, X., Liu, J., Zou, Y., & He, R. (2018). A kriging and entropy-based approach to raingauge network design. *Environmental Research*, *161*(August 2017), 61–75. https://doi.org/10.1016/j.envres.2017.10.038.

Rohma, N. N. (2022). Estimation of Ordinary Kriging Method with Jackknife Technique on Rainfall Data in Malang Raya.

*International Journal on Information and Communication Technology (IJoICT)*, *8*(2), 22–39. https://doi.org/10.21108/ijoict.v8i2.678.

Beran, J., Liu, H., & Ghosh, S. (2020). On aggregation of strongly dependent time series. *Scandinavian Journal of Statistics*, *47*(3), 690–710. https://doi.org/10.1111/sjos.12421.

Hasan, M. M., & Croke, B. F. W. (2013). Filling gaps in daily rainfall data: A statistical approach. *Proceedings - 20th International Congress on Modelling and Simulation, MODSIM 2013*, *December*, 380–386. https://doi.org/10.36334/modsim.2013.a9.hasan.

Hristopulos, D. T., & Baxevani, A. (2020). Effective probability distribution approximation for the reconstruction of missing data. *Stochastic Environmental Research and Risk Assessment*, *34*(2), 235–249. https://doi.org/10.1007/s00477-020-01765-5.

Nathans, L. L., Oswald, F. L., & Nimon, K. (2012). Interpreting multiple linear regression: A guidebook of variable importance. *Practical Assessment, Research and Evaluation*, *17*(9), 1–19.

Pappas, C., Papalexiou, S. M., & Koutsoyiannis, D. (1955). Journal of geophysical research. *Nature*, *175*(4449), 238. https://doi.org/10.1038/175238c0.

Shin, K., Song, J. J., Bang, W., & Lee, G. (2021). Approaches with Operational Dual-Polarization Radar Data. *Remote Sensing*, *13*(694), 1–21.

Longman, R. J., Newman, A. J., Giambelluca, T. W., & Lucas, M. (2020). Characterizing the uncertainty and assessing the value of gap-filled daily rainfall data in hawaii. *Journal of Applied Meteorology and Climatology*, *59*(7), 1261–1276. https://doi.org/10.1175/JAMC-D-20-0007.1.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., & Levizzani, V. (2014). Journal of Geophysical Research : Atmospheres rainfall from satellite soil moisture data. *Journal of Geophysical Research: Atmospheres*, *119*, 5128–5141. https://doi.org/10.1002/2014JD021489.

Haloho, L. S., & Supriyadi, A. A. (2024). Utilization of satellite technology in communication systems, disaster monitoring, border surveillance, and military intelligence: a literature review. *Remote Sensing Technology in Defense and Environment*, *1*(1), 36–44. https://doi.org/10.61511/rstde.v1i1.2024.842.

Siabi, N., Sanaeinejad, S. H., & Ghahraman, B. (2020). Comprehensive evaluation of a spatio-temporal gap filling algorithm: Using remotely sensed precipitation, LST and ET data. *Journal of Environmental Management*, *261*(March), 110228. https://doi.org/10.1016/j.jenvman.2020.110228.

de Moraes Cordeiro, A. L., & Blanco, C. J. C. (2021). Assessment of satellite products for filling rainfall data gaps in the Amazon region. *Natural Resource Modeling*, *34*(2), 1–21. https://doi.org/10.1111/nrm.12298.

Abu Romman, Z., Al-Bakri, J., & Al Kuisi, M. (2021). Comparison of methods for filling in gaps in monthly rainfall series in arid regions. *International Journal of Climatology*, *41*(15), 6674–6689. https://doi.org/10.1002/joc.7219.

Andari, R., Nurhamidah, N., Daoed, D., & Marzuki. (2024a). Evaluation of bias correction methods for multi-satellite rainfall estimation products. *IOP Conference Series: Earth and Environmental Science*, *1317*(1). https://doi.org/10.1088/1755-1315/1317/1/012008.

Rozante, J. R., Vila, D. A., Chiquetto, J. B., Fernandes, A. de A., & Alvim, D. S. (2018). Evaluation of TRMM/GPM blended daily products over Brazil. *Remote Sensing*, *10*(6), 1–17. https://doi.org/10.3390/rs10060882.

Elnashar, A., Zeng, H., Wu, B., Zhang, N., Tian, F., Zhang, M., Zhu, W., Yan, N., Chen, Z., Sun, Z., Wu, X., & Li, Y. (2020). Downscaling TRMM monthly precipitation using google earth engine and google cloud computing. *Remote Sensing*, *12*(23), 1–22. https://doi.org/10.3390/rs12233860.

Ramadhan, R., Yusnaini, H., Marzuki, M., Muharsyah, R., Suryanto, W., Sholihun, S., Vonnisa, M., Harmadi, H., Ningsih, A. P., Battaglia, A., Hashiguchi, H., & Tokay, A. (2022). Evaluation of GPM IMERG Performance Using Gauge Data over Indonesian Maritime Continent at Different Time Scales. *Remote Sensing*, *14*(5), 1–24. https://doi.org/10.3390/rs14051172.

Zhou, Z., Guo, B., Xing, W., Zhou, J., Xu, F., & Xu, Y. (2020). Comprehensive evaluation of latest GPM era IMERG and GSMaP precipitation products over mainland China. *Atmospheric Research*, *246*. https://doi.org/10.1016/j.atmosres.2020.105132.

Ramadhan, R., Marzuki, M., Yusnaini, H., Muharsyah, R., Suryanto, W., Sholihun, S., Vonnisa, M., Battaglia, A., & Hashiguchi, H. (2022). Capability of GPM IMERG Products for Extreme Precipitation Analysis over the Indonesian Maritime Continent. *Remote Sensing*, *14*(2). https://doi.org/10.3390/rs14020412.

Ramadhan, R., Marzuki, M., Yusnaini, H., Muharsyah, R., Tangang, F., Vonnisa, M., & Harmadi, H. (2023). A Preliminary Assessment of the GSMaP Version 08 Products over Indonesian Maritime Continent against Gauge Data. *Remote Sensing*, *15*(4), 1115. https://doi.org/10.3390/rs15041115.

Nepal, B., Shrestha, D., Sharma, S., Shrestha, M. S., Aryal, D., & Shrestha, N. (2021). Assessment of GPM-Era satellite

products' (IMERG and GSMaP) ability to detect precipitation extremes over mountainous country nepal. *Atmosphere*, *12*(2), 254. https://doi.org/10.3390/atmos12020254.

dos Santos, E. P., Dias, R. L. S., Maciel, I. P., Kolling Neto, A., & da Silva, D. D. (2021). Estimation of missing hydrological data in monthly rainfall series using meteorological satellite data. *Environmental Earth Sciences*, *80*(3), 1–9. https://doi.org/10.1007/s12665-021-09409-9.

Elshorbagy, A. A., Panu, U. S., & Simonovic, S. P. (2000). Group-based estimation of missing hydrological data: I. Approach and general methodology. *Hydrological Sciences Journal*, *45*(6), 849–866. https://doi.org/10.1080/02626660009492388.

Yozgatligil, C., Aslan, S., Iyigun, C., & Batmaz, I. (2013). Comparison of missing value imputation methods in time series: The case of Turkish meteorological data. *Theoretical and Applied Climatology*, *112*(1–2), 143–167. https://doi.org/10.1007/s00704-012-0723-x.

Gomes, E. P., Blanco, C. J. C., & Pessoa, F. C. L. (2018). Regionalization of precipitation with determination of homogeneous regions via fuzzy c-means. *Rbrh*, *23*(0). https://doi.org/10.1590/2318-0331.231820180079.

Byun, J., Kim, H. J., Kang, N., Yoon, J., Hwang, S., & Jun, C. (2024). Optimizing Temporal Weighting Functions to Improve Rainfall Prediction Accuracy in Merged Numerical Weather Prediction Models for the Korean Peninsula. *Remote Sensing*, *16*(16). https://doi.org/10.3390/rs16162904.

McCuen, R. H., Knight, Z., & Cutter, A. G. (2006). Evaluation of the Nash–Sutcliffe Efficiency Index. *Journal of Hydrologic Engineering*, *11*(6), 597–602. https://doi.org/10.1061/(asce)1084-0699(2006)11:6(597).

Motovilov, Y. G., Gottschalk, L., Engeland, K., & Rodhe, A. (1999). Validation of a distributed hydrological model against spatial observations. *Agricultural and Forest Meteorology*, *98–99*, 257–277. https://doi.org/10.1016/S0168-1923(99)00102-1.

Van Liew, M. W., Veith, T. L., Bosch, D. D., & Arnold, J. G. (2007). Suitability of SWAT for the Conservation Effects Assessment Project: Comparison on USDA Agricultural Research Service Watersheds. *Journal of Hydrologic Engineering*, *12*(2), 173–189. https://doi.org/10.1061/(asce)1084-0699(2007)12:2(173).

Soares, A., Paz, A., & Piccilli, D. (2016). Avaliação das estimativas de chuva do satélite TRMM no Estado da Paraíba / Assessment of rainfall estimates of TRMM satellite on Paraíba state. *Revista Brasileira de Recursos Hídricos*, *21*(2), 288–299. https://doi.org/10.21168/rbrh.v21n2.p288-299.

Andari, R., Nurhamidah, N., Daoed, D., & Marzuki, M. (2024b). Validation of TRMM and GPM Satellite Data Using Daily Precipitation Observations. *International Journal on Advanced Science, Engineering and Information Technology*, *14*(2), 555–562. https://doi.org/10.18517/ijaseit.14.2.18980.

Portuguez-maurtua, M., Arumi, J. L., Lagos, O., Stehr, A., & Arquiñigo, N. M. (2022). Filling Gaps in Daily Precipitation Series Using Regression and Machine Learning in Inter-Andean Watersheds. *Water (Switzerland)*, *14*(11). https://doi.org/10.3390/w14111799.

Camuffo, D., Becherini, F., della Valle, A., & Zanini, V. (2022). A comparison between different methods to fill gaps in early precipitation series. *Environmental Earth Sciences*, *81*(13), 1–14. https://doi.org/10.1007/s12665-022-10467-w.

Bárdossy, A., & Pegram, G. (2014). Infilling missing precipitation records - A comparison of a new copula-based method with other techniques. *Journal of Hydrology*, *519*(PA), 1162–1170. https://doi.org/10.1016/j.jhydrol.2014.08.025.

Chua, Z. W., Kuleshov, Y., Watkins, A. B., Choy, S., & Sun, C. (2023). A Statistical Interpolation of Satellite Data with Rain Gauge Data over Papua New Guinea. *Journal of Hydrometeorology*, *24*(12), 2369–2387. https://doi.org/10.1175/JHM-D-23-0035.1.

Lutfiah, Q., Purnaditya, N. P., Priyambodho, B. A., & Wigati, R. (2024). *Evaluation Of Pdir-Now Satellite Rainfall Data On Observational Rainfall Data ( Case Study : Taktakan District , Serang City , Banten )*. *13*(2).

## How to Cite This Article

Andari, R., Nurhamidah, N., Daoed, D., & Marzuki, M. (2025). Enhancing Daily Rainfall Data Completeness Using Satellite Rainfall Estimates. Civil and Environmental Engineering, Vol. 22, Issue 1, 105-118. https://doi.org/10.2478/cee-2026-0008