

Satellite-Based Estimation of Monthly Rainfall Across the Indonesian Archipelago: A Month-Specific Multiple Linear Regression Framework Using Bias-Corrected NASA POWER Data over Java, Bali, Nusa Tenggara, and Maluku

Christian Cahyono

`christian.cahyono@binus.ac.id`

Bina Nusantara University

Alfaldo Branoyasensa Baria

Bina Nusantara University

Karina Fakhriyah Desvayana

Undergraduate Student, Bina Nusantara University

Shafa Naulia

Undergraduate Student, Bina Nusantara University

Denik Sri Krisnayanti

University of Nusa Cendana

Dyah Erny Herwindiati

Lecture, Tarumanegara University

Andrew John Pierre

Bina Nusantara University

Kania Cahyadi

Undergraduate Student, Bina Nusantara University

Maria Putri Anggraini Ticoalu

Undergraduate Student, Bina Nusantara University

Research Article

Keywords: monthly rainfall estimation, NASA POWER, multiple linear regression, bias correction, BMKG, tropical hydrometeorology, Indonesian Maritime Continent

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10491770/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Reliable long-term rainfall records remain scarce across Indonesia because the freely accessible BMKG Data Online portal typically distributes only daily observations for roughly the two most recent years, which is insufficient for robust trend analysis and model calibration. This study develops and evaluates a transparent, month-specific Multiple Linear Regression (MLR) framework for estimating monthly rainfall across four Indonesian regions Java, Bali, Nusa Tenggara, and Maluku using bias-corrected NASA Prediction of Worldwide Energy Resources (NASA POWER) satellite data (2011–2026) as predictors and BMKG station observations as the reference. Air temperature (T2M), relative humidity (RH2M), and wind speed (WS2M) were used as predictors, and the satellite precipitation field was first bias-corrected against gauge data using a probability-curve (duration-curve) approach. Twelve month-specific regression equations were fitted per region and evaluated at 66 BMKG stations using Root Mean Square Error (RMSE), the Pearson correlation coefficient (R), Nash–Sutcliffe Efficiency (NSE), and Relative Bias (RB). Aggregated regional performance was good in every region: Java (RMSE = 139.18 mm/month, R = 0.65, NSE = 0.50, RB = 5.67%), Bali (183.41, 0.73, 0.70, 3.76%), Nusa Tenggara (148.93, 0.70, 0.63, 5.41%), and Maluku (190.81, 0.64, 0.50, 12.35%). Relative humidity and air temperature were the dominant predictors, whereas wind speed was mostly negative during the dry monsoon. The results confirm that bias-corrected NASA POWER data provide a viable, low-cost alternative for monthly rainfall estimation in data-sparse regions of central and eastern Indonesia, supporting water-resource planning, hydrometeorological disaster mitigation, and evidence-based public policy.

Introduction

Rainfall is the most spatially and temporally variable component of the tropical hydrological cycle, and it directly governs agriculture, water-resource management, transportation, and disaster mitigation. In Indonesia, shifting rainfall patterns driven by climate variability have amplified the risk of floods, landslides, droughts, and crop failure, so accurate rainfall information is a prerequisite for sustainable water management and precise public policy [1], [2]. Reliable estimation of monthly rainfall is particularly valuable for reservoir operation, irrigation scheduling, flood-control design, and raw-water supply planning. A persistent obstacle, however, is the limited availability of long observational records. The freely accessible BMKG Data Online portal generally provides only daily rainfall for approximately the two most recent years at any given station, which is inadequate for characterising long-term trends or for calibrating statistically robust prediction models. This short record length constrains the ability of any model to capture the full annual cycle and interannual variability of monsoonal rainfall, reducing predictive skill. Missing-value and record-length problems are recurrent themes in Indonesian rainfall studies, motivating both careful imputation and the use of alternative long-record datasets [13], [14].

To overcome the scarcity of long ground records, satellite- and reanalysis-based precipitation products have been widely adopted for hydrological analysis. The Tropical Rainfall Measuring Mission (TRMM) has been validated against gauge networks for watershed hydrology and slope-stability applications in Indonesia [4], [5], [6]; its successor, the Global Precipitation Measurement (GPM) IMERG product, has

been evaluated extensively over the Indonesian Maritime Continent for extreme-precipitation and multi-timescale hydrological analysis [3], [7]. Reanalysis precipitation from ERA5 and ERA5-Land has been assessed as a reference dataset for hydrological modelling in several climatic settings [8], [10]. The Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) product has been compared with GPM for generating daily rainfall information in Java [9], while gauge-adjusted Global Satellite Mapping of Precipitation (GSMaP) estimates have also been validated for Indonesian applications [6]. Collectively, these studies confirm that satellite and reanalysis rainfall products are a practical alternative wherever gauge coverage is sparse, provided that bias correction against local observations is applied.

Among these products, NASA POWER is especially attractive for the present application because it delivers a continuous, openly accessible, and spatially homogeneous record that spans the entire Indonesian archipelago from 1981 to near-real time, together with the meteorological predictors (temperature, humidity, and wind speed) needed for regression-based rainfall modelling from a single integrated platform. NASA POWER has been benchmarked for estimating tropical precipitation and temperature extremes [10], used as an alternative rainfall resource for hydrology and planning in Bali [11], validated against BMKG observations in Sumatra [12], and coupled with machine-learning models for tropical rainfall prediction. Because NASA POWER provides uniform coverage over Java, Bali, Nusa Tenggara, and Maluku, it enables a single, consistent modelling framework to be applied across regions with very different observational densities.

The objective of this study is therefore to develop a transparent linear-regression framework that estimates and predicts monthly rainfall across four Indonesian regions using a new, corrected dataset NASA POWER meteorological data that have been bias-corrected against BMKG ground-station observations. Building on a recent Multiple Linear Regression (MLR) baseline established for Jakarta, South Tangerang, and Bandung [20], the present work extends the approach from western Java to the central and eastern archipelago (Bali, Nusa Tenggara, and Maluku), where long gauge records are scarcest and where the operational value of a satellite-driven estimator is greatest. The specific aims are: (i) to bias-correct NASA POWER rainfall against BMKG observations; (ii) to fit month-specific MLR equations for each region; (iii) to evaluate model performance at 66 stations using four objective functions; and (iv) to compare predictive skill between Java and the central–eastern regions.

Basic Theory

Multiple Linear Regression

Multiple Linear Regression (MLR) models the linear dependence of a response variable on several predictor variables. It is a transparent, reproducible, and computationally light baseline that has been widely used for rainfall prediction from routinely observed meteorological variables such as temperature, humidity, and wind speed [15], [16], [17]. In this study monthly rainfall is the dependent variable (Y) and the three NASA POWER meteorological variables are the independent variables. The general model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

1

where Y is monthly rainfall (mm); β_0 is the regression constant; β_1 , β_2 , and β_3 are the regression coefficients; X_1 is air temperature (°C); X_2 is relative humidity (%); X_3 is wind speed (m/s); and ϵ is the residual error. Coefficients are estimated by Ordinary Least Squares (OLS), which minimises the residual sum of squares:

$$\min \sum (Y_i - \hat{Y}_i)^2$$

2

Because monsoonal Indonesia exhibits strong seasonality, a single annual equation cannot capture the changing influence of each meteorological driver. Following the month-specific strategy of [20], a separate equation was fitted for each calendar month, allowing the constant and the three coefficients to vary through the year and thereby representing the seasonal dynamics of the atmosphere [18].

Model Performance Evaluation

Model skill was quantified with four complementary objective functions whose interpretive thresholds (very good, good, satisfactory, and unsatisfactory) follow the criteria compiled for global-data rainfall evaluation by [19]. The Root Mean Square Error (RMSE) measures the average magnitude of prediction error in mm/month:

$$RMSE = \sqrt{\frac{1}{n} \sum (O_i - P_i)^2}$$

3

The Pearson correlation coefficient (R) measures the strength and direction of the linear relationship between observed (O) and predicted (P) rainfall:

$$R = \frac{\sum (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{(\sum (O_i - \bar{O})^2) \sum (P_i - \bar{P})^2}}$$

4

The Nash–Sutcliffe Efficiency (NSE) expresses the proportion of observed variance reproduced by the model; $NSE = 1$ denotes a perfect fit and $NSE \leq 0$ indicates that the model is no better than the observed mean:

$$NSE = 1 - \frac{\sum (O_i - P_i)^2}{\sum (O_i - \bar{O})^2}$$

5

The Relative Bias (RB, %) measures the average systematic over- or under-estimation of the model relative to observations:

$$RB = \frac{\sum (P_i - O_i)}{\sum O_i} \times 100\%$$

6

where O_i and P_i are the observed and predicted monthly rainfall at time step i , $\bar{O}$ and $\bar{P}$ their means, and n the number of observations. Together these four metrics characterise error magnitude (RMSE), linear association (R), variance reproduction (NSE), and systematic bias (RB), giving a balanced picture of model performance [19].

Research Locations and Data

Study Area

The study spans four regions of the Indonesian archipelago: the island of Java (including Madura), Bali, Nusa Tenggara (West and East Nusa Tenggara), and Maluku (Maluku and North Maluku). This transect crosses the Indonesian Maritime Continent from west to east and captures a wide range of hydroclimatic settings densely gauged urban-coastal corridors in Java, small tourism-dominated catchments in Bali, the strongly seasonal monsoon-arid regime of Nusa Tenggara, and the convective, island-fragmented maritime climate of Maluku. Rainfall across all four regions is governed primarily by the Asian-Australian monsoon system, with the west monsoon (roughly October–April) delivering moist air from the Indian Ocean and the east monsoon (roughly April–October) advecting drier air from the Australian continent; large-scale drivers such as the Madden–Julian Oscillation further modulate rainfall variability, particularly over the eastern islands [21]. The extension of the western-Java MLR baseline [20] to these central and eastern regions is the core novelty of this study, in line with the growing need for statistical rainfall estimators in data-sparse Indonesian settings [22], [23], [24].

BMKG Observation Data

Observed monthly rainfall from the Meteorological, Climatological, and Geophysical Agency (BMKG) served as the reference for bias correction and validation. Rainfall is the only BMKG variable used, because it is the target variable to be predicted. Freely accessible BMKG records were obtained for 2024–2026 through the official Data Online portal; although these gauge observations are highly reliable

as direct field measurements, their short accessible length is precisely the limitation that motivates the use of a long satellite record. A total of 66 BMKG stations were used across the four regions, distributed as summarised in Table 1.

Table 1
Number of BMKG observation stations used in each region (66 stations in total).

Region	Island group	No. of stations
Java	Java + Madura	34
Bali	Bali	4
Nusa Tenggara	NTB + NTT	15
Maluku	Maluku + North Maluku	13
Total	-	66

NASA POWER Satellite Data

The NASA Prediction of Worldwide Energy Resources (NASA POWER) platform was the primary data source for building the prediction models, owing to its long temporal coverage (2011–2026 was used here), continuous availability, spatially homogeneous global grid, and open access. NASA POWER derives its meteorological fields from the MERRA-2 atmospheric reanalysis and provides all required variables from a single platform, ensuring a homogeneous predictor set across every study region [10], [12]. Four variables were extracted: Temperature at 2 m (T2M), Relative Humidity at 2 m (RH2M), Wind Speed at 2 m (WS2M), and Corrected Precipitation (PRECTOTCORR). These variables were selected for their well-established physical linkage to cloud formation and precipitation, making them suitable predictors for regional rainfall modelling. The key advantage of NASA POWER for this study is that it offers uniform, long-record coverage over Java, Bali, Nusa Tenggara, and Maluku alike, including the sparsely gauged eastern islands where terrestrial observations are most limited [11].

Research Methodology

The methodological framework spans data acquisition, quality control, satellite bias correction, month-specific MLR calibration, prediction, and performance evaluation across the four regions. The complete workflow is shown in Fig. 1 and described step by step below.

Start and research-objective formulation. The workflow begins by defining the objective estimating monthly rainfall across the four regions and the study area (66 BMKG stations spanning Java, Bali, Nusa Tenggara, and Maluku). Data collection. Two data streams are collected in parallel. BMKG rainfall serves as the observation/reference series, while NASA POWER supplies the predictor variables (T2M, RH2M, WS2M) together with PRECTOTCORR for the calibration period 2011–2026. Data quality control. Both series undergo quality control comprising missing-value imputation and consistency verification

between daily and monthly aggregations. Records that fail the check are routed to a data-checking-and-cleaning step and re-entered into the workflow; only quality-controlled records proceed. Careful imputation is essential because short and incomplete records otherwise bias downstream statistics [13], [14]. Probability-curve bias correction. Satellite rainfall and BMKG observation are matched over a common period, the paired values are ranked into probability (duration) curves, and interval-wise correction relationships are derived so that the corrected satellite series reproduces the observed distribution. Following established practice, satellite data are corrected only where they show good agreement with observations (correlation above roughly 0.6). The station-level correlations between the original NASA POWER and BMKG monthly series are strong for the majority of stations, confirming that the correction is well founded.

Month-specific MLR calibration. Using the corrected data, twelve month-specific regression equations [Eq. (1)] are fitted per region by OLS, aggregating all stations within a region. Each equation yields a distinct constant and set of coefficients, allowing the influence of temperature, humidity, and wind to vary seasonally.

Prediction and evaluation. The calibrated equations are applied to the NASA POWER predictors to produce monthly rainfall estimates, which are verified against independent BMKG observations. Model performance is evaluated with RMSE, R, NSE, and RB [Eqs. (3)–(6)] at each station and aggregated per region. A final cross-island comparison and discussion precede the conclusions and recommendations.

Analysis and Findings

This section reports the month-specific regression equations, the per-station objective-function results, and the actual-versus-predicted scatter for each region, followed by an aggregated cross-island comparison. In each coefficient table the equation is read as $Y = \beta_0 + \beta_1 \cdot T_{avg} + \beta_2 \cdot RH_{avg} + \beta_3 \cdot FF_{avg}$, where Y is monthly rainfall. The overall performance is good in every region, confirming that the bias-corrected NASA POWER data support skilful monthly rainfall estimation across the archipelago [19].

Java

The monthly regression coefficients for Java are given in Table 2. The constant and coefficients differ month to month, reflecting the monsoonal contrast between wet and dry seasons. The temperature coefficient (T_{avg}) is largely positive peaking at 125.93 in February consistent with enhanced evaporation and moisture availability under warmer conditions, while relative humidity (RH_{avg}) is positive in every month, underscoring its central role in condensation and cloud formation. Wind speed (FF_{avg}) is predominantly negative from April to December, consistent with the dry east-monsoon advection of relatively dry Australian air, but positive in January when the wet-season circulation still supports moisture transport.

Table 2
Month-specific MLR coefficients for Java.

Month	β_0	Tavg (β_1)	RHavg (β_2)	FFavg (β_3)
January	-5393.43	55.52	49.78	10.30
February	-11551.75	125.93	100.44	-4.89
March	-1306.09	0.56	18.10	-3.61
April	-2900.89	32.29	27.08	-32.48
May	-3315.05	36.78	29.13	-4.17
June	-1654.39	23.94	13.93	-23.58
July	-1098.84	19.46	8.82	-21.77
August	-292.05	6.45	2.36	-12.82
September	-706.67	15.31	5.96	-32.58
October	-792.57	11.20	9.37	-46.44
November	-666.04	-1.15	12.15	-39.40
December	-4624.42	68.23	39.21	-74.11

The scatter of predicted against observed rainfall (Fig. 2) shows that most points cluster between about 100 and 500 mm/month and follow the reference line, so the model reproduces the low, medium, and high rainfall categories of the BMKG classification and also captures extreme events above 300 mm/month. Station-level metrics (Table 3) place most Java stations in the good-to-very-good category for correlation, with the highest R at Penggung (0.85), Perak 1 (0.81), Pasuruan (0.79), and Maritim Tanjung Perak (0.77). RMSE mostly falls in the 100–200 mm/month band, with the lowest value at Maritim Serang (97.50 mm/month) and the highest at Jawa Barat (257.14 mm/month). NSE reaches very good values at Kemayoran (1.00) and Penggung (0.69), and RB is smallest at Tunggul Wulung (0.15), Soekarno-Hatta (1.10), and Juanda (1.13).

Table 3
Per-station evaluation for Java (34 BMKG stations).

BMKG station	RMSE (mm/mo)	R	NSE	RB (%)
Halim Perdanakusuma	197.07	0.31	-0.57	20.69
Kemayoran	135.33	0.69	1.00	7.39
Tanjung Priok	121.67	0.76	0.58	3.78
Banten	151.71	0.38	-0.41	1.94
Budiarto	167.50	0.26	-0.87	11.83
Geofisika Tangerang	112.37	0.52	0.53	6.73
Soekarno-Hatta	120.63	0.71	0.49	1.10
Maritim Serang	97.50	0.74	0.52	6.45
Tunggul Wulung	201.53	0.20	-0.39	0.15
Banjar Negara	149.46	0.63	0.35	13.34
Jawa Tengah	130.83	0.72	0.37	19.24
Ahmad Yani	111.35	0.74	0.43	9.62
Maritim Tanjung Emas	125.20	0.66	0.25	7.25
Maritim Tegal	126.18	0.71	-0.13	39.17
Bandung	106.04	0.56	-0.07	6.29
Jawa Barat	257.14	0.13	-1.19	38.01
Citeko	142.07	0.70	0.31	21.27
Kertajati	146.69	0.66	0.35	16.53
Penggung	105.12	0.85	0.69	14.36
Sleman	157.71	0.57	0.32	2.70
Klimatologi DIY	140.39	0.69	0.45	4.04
Meteorologi DIY	123.32	0.70	0.46	10.81
Dhoho	150.15	0.66	0.41	15.57
Klimatologi Jawa Timur	127.97	0.68	0.11	15.64
Malang	187.89	0.50	0.15	2.85
Banyuwangi	117.30	0.74	-0.25	58.01
Pasuruan	178.16	0.79	0.43	33.15

BMKG station	RMSE (mm/mo)	R	NSE	RB (%)
Juanda	139.35	0.75	0.50	1.13
Nganjuk	204.09	0.64	0.38	14.62
Tuban	181.49	0.56	-2.47	76.73
Sangkapura	168.69	0.69	0.20	41.04
Trunojoyo	139.66	0.51	-0.34	10.04
Perak 1	105.89	0.81	0.46	23.91
Maritim Tanjung Perak	117.00	0.77	0.28	31.21

Maluku

The Maluku coefficients (Table 4) again vary seasonally. Air temperature is generally positive but turns negative in May, June, August, and December, while relative humidity is positive in every month, confirming its dominant role in precipitation formation. Wind speed alternates in sign, reflecting the changing atmospheric regime through the year over this fragmented, strongly convective maritime domain.

Table 4
Month-specific MLR coefficients for Maluku.

Month	β_0	Tavg (β_1)	RHavg (β_2)	FFavg (β_3)
January	-3529.82	57.80	22.45	108.73
February	-2105.49	34.19	14.12	59.76
March	-2300.67	47.37	12.69	60.12
April	-7189.56	147.87	40.28	18.52
May	-2158.75	-69.47	51.14	83.84
June	-2863.90	-14.71	42.62	8.93
July	-13356.60	155.14	114.12	-1.44
August	-2311.95	-11.48	34.02	-16.88
September	-2596.46	30.81	24.34	-25.28
October	-1884.80	36.24	12.22	-4.61
November	-3476.37	64.21	21.79	28.81
December	-1044.40	-3.76	12.83	120.62

Figure 3 shows a positive correspondence between predicted and observed rainfall, with most events in the 100–300 mm/month range and some exceeding 500 mm/month. Although the point spread is wider than in Java and some points lie outside the $\pm 20\%$ envelope, the model still reproduces the overall rainfall trend, including high-intensity events relevant to flood and surface-runoff risk. Per-station results (Table 5) show the strongest correlation at Amahai (0.80, NSE = 0.57) and the weakest skill at Namlea and Bandaneira, where negative NSE values reveal that the local rainfall regime is only partly explained by the predictor set.

Table 5
Per-station evaluation for Maluku (13 BMKG stations).

BMKG station	RMSE (mm/mo)	R	NSE	RB (%)
Amahai	157.41	0.80	0.57	9.70
Bandaneira	210.79	0.21	-0.21	16.38
Karel Sadsuitubun	137.41	0.68	0.37	14.27
Mathilda Batlayeri	167.70	0.62	-0.09	43.54
Namlea	204.27	0.16	-3.52	108.42
Kuffar	139.90	0.56	0.20	17.08
Klimatologi Maluku	171.79	0.65	0.21	5.86
Pattimura	278.91	0.68	0.38	22.57
Maritim Ambon	312.63	0.56	0.19	29.01
Gamar Malamo	93.13	0.60	0.32	7.38
Oesman Sadik	119.68	0.35	0.05	8.60
Emalamo	125.98	0.21	-1.39	30.75
Sultan Babullah	208.29	0.41	-0.45	30.75

Nusa Tenggara

For Nusa Tenggara (Table 6) the temperature coefficient is predominantly negative, indicating that in most months higher temperatures coincide with lower rainfall consistent with the pronounced dry season of this region while relative humidity is generally positive (negative only in November). Wind speed is positive early and late in the year and negative in mid-year, mirroring the monsoon reversal.

Table 6
Month-specific MLR coefficients for Nusa Tenggara.

Month	β_0	Tavg (β_1)	RHavg (β_2)	FFavg (β_3)
January	-2654.34	6.20	34.26	26.01
February	-1229.29	-35.53	28.28	62.35
March	1521.09	-98.94	15.75	59.02
April	2354.07	-106.41	7.71	50.29
May	-592.39	-5.48	10.69	-7.60
June	-622.07	10.80	5.42	-10.02
July	-88.65	-0.55	1.89	-4.61
August	132.49	-7.05	1.08	-3.66
September	199.44	-13.48	3.66	-19.68
October	1465.19	-54.67	3.13	-31.45
November	4600.88	-150.13	-1.45	-13.38
December	3759.09	-130.26	0.91	50.69

The scatter (Fig. 4) is concentrated in the 0–300 mm/month range with occasional events above 700 mm/month, and predicted increases generally track observed increases. Nusa Tenggara returns the strongest station-level correlations of any region (Table 7): Komodo (0.93), Gewayantana (0.92), Mali and Eltari (0.91), and Klimatologi NTT (0.90). Several stations also achieve very good NSE (Gewayantana 0.87, Eltari 0.83, Klimatologi NTT 0.82), although a few coastal stations show large RB, indicating localised systematic bias that would benefit from station-specific correction.

Table 7
Per-station evaluation for Nusa Tenggara (15 BMKG stations).

BMKG station	RMSE (mm/mo)	R	NSE	RB (%)
Komodo	128.05	0.93	-0.41	71.35
Frans Sales Lega	182.04	0.67	0.77	0.67
Fransiskus Xaverius Seda	101.52	0.83	0.57	38.78
Gewayantana	90.01	0.92	0.87	21.95
Mali	125.84	0.91	0.07	66.10
Umbu Mehang Kunda	181.57	0.82	-1.04	140.46
Klimatologi NTT	104.85	0.90	0.82	24.78
Eltari	100.07	0.91	0.83	25.12
David Constatijn Saudale	135.85	0.80	0.75	31.59
Tardamu	109.29	0.70	0.59	8.36
Maritim Tenau	202.86	0.69	-0.83	143.29
Klimatologi NTB	140.77	0.62	0.12	79.93
Sultan Muhammad Kaharuddin	143.48	0.85	0.12	78.83
Sultan Muhammad Salahuddin	144.57	0.81	0.31	26.72
Zainuddin Abdul Madjid	175.48	0.42	0.27	53.71

Bali

The Bali coefficients (Table 8) display the largest month-to-month swings most notably the very large January and February constants and temperature coefficients reflecting the strong wet-season signal on this small, topographically varied island. Relative humidity remains positive throughout the year and wind speed switches sign seasonally, again pointing to a monsoon-controlled interaction among the climate variables.

Table 8
Month-specific MLR coefficients for Bali.

Month	β_0	Tavg (β_1)	RHavg (β_2)	FFavg (β_3)
January	-35523.80	693.71	212.44	-29.58
February	-14658.20	252.02	98.01	98.19
March	-3363.62	17.09	34.84	184.16
April	-1116.94	-34.62	22.51	152.88
May	-4156.71	47.08	34.59	45.42
June	-2755.35	50.86	18.59	-14.05
July	-1458.71	6.05	15.52	26.09
August	-320.71	9.65	1.43	-8.87
September	-2715.74	75.81	12.07	-43.42
October	-1956.47	-2.96	25.41	54.66
November	-3892.78	43.81	35.06	62.80
December	-7028.36	98.46	55.50	110.93

Figure 5 shows most Bali data in the 0–300 mm/month range with some events approaching 900 mm/month; predicted and observed rainfall increase together, though the spread widens at the highest intensities. Per-station results (Table 9) are excellent at Ngurah Rai ($R = 0.92$, $NSE = 0.85$, $RB = 0.62\%$) and Denpasar ($R = 0.91$, $NSE = 0.82$), with weaker skill at the outlying Bali station, giving a strong aggregated regional performance.

Table 9
Per-station evaluation for Bali (4 BMKG stations).

BMKG station	RMSE (mm/mo)	R	NSE	RB (%)
Bali	281.83	0.45	-2.61	31.92
Ngurah Rai	92.99	0.92	0.85	0.62
Kahang-Kahang	193.35	0.77	0.54	8.09
Denpasar	95.37	0.91	0.82	1.99

Cross-Island Comparison

Aggregating the station results to the regional scale (Table 10) shows that the framework performs in the good category everywhere, with RB in the very-good range for Java, Bali, and Nusa Tenggara. Bali attains the highest correlation (0.73) and NSE (0.70) and the smallest bias (3.76%), followed closely by

Nusa Tenggara ($R = 0.70$, $NSE = 0.63$, $RB = 5.41\%$). Java, despite being the most densely gauged region, returns a slightly lower correlation (0.65) and NSE (0.50); this reflects the strong influence of local convective and orographic processes and intense urbanisation in Java, which introduce sub-monthly variability that a linear, station-aggregated model cannot fully resolve [20], [22]. Maluku shows the highest $RMSE$ (190.81 mm/month) and RB (12.35%), consistent with its fragmented island geography, sparse gauge network, and highly convective maritime rainfall that is harder to represent with surface meteorological predictors [3].

Table 10
Aggregated regional performance across the four regions.

Region	RMSE (mm/mo)	R	NSE	RB (%)
Java	139.18	0.65	0.50	5.67
Bali	183.41	0.73	0.70	3.76
Nusa Tenggara	148.93	0.70	0.63	5.41
Maluku	190.81	0.64	0.50	12.35

Comparing Java with the central and eastern regions is instructive. Bali and Nusa Tenggara actually outperform Java on correlation and NSE , which at first appears counter-intuitive given their sparser networks. The explanation lies in rainfall regime rather than data density: Bali and Nusa Tenggara have a sharper, more deterministic monsoonal seasonality that a month-specific linear model captures well, whereas Java's rainfall is more strongly modulated by mesoscale convection, complex topography, and urban heat-island effects that add non-linear noise. Maluku sits at the difficult end of this spectrum, where dispersed islands and deep convection reduce the explanatory power of large-scale surface variables. Overall, the consistency of good-category scores from the densely gauged west to the data-sparse east demonstrates that bias-corrected NASA POWER data can substitute effectively for long gauge records in monthly rainfall estimation, echoing prior evidence that global datasets, once bias-corrected, are fit for hydrological use in Indonesia [10], [19], [23].

Conclusions

This study developed and evaluated a transparent, month-specific Multiple Linear Regression framework for estimating monthly rainfall across four Indonesian regions Java, Bali, Nusa Tenggara, and Maluku using bias-corrected NASA POWER satellite data as predictors and BMKG station observations as the reference. The limited length of freely accessible BMKG records is a genuine obstacle to long-term rainfall modelling in Indonesia; the results here show that the long, spatially homogeneous NASA POWER record, once corrected against gauge data through a probability-curve approach, provides an effective and low-cost alternative.

Twelve month-specific equations per region reproduced the seasonal shift in the influence of temperature, humidity, and wind, with relative humidity emerging as the most consistent positive

predictor and wind speed generally negative during the dry east monsoon. Aggregated regional performance was good throughout: Java (RMSE = 139.18 mm/month, $R = 0.65$, NSE = 0.50, RB = 5.67%), Bali (183.41, 0.73, 0.70, 3.76%), Nusa Tenggara (148.93, 0.70, 0.63, 5.41%), and Maluku (190.81, 0.64, 0.50, 12.35%). Notably, the central and eastern regions of Bali and Nusa Tenggara matched or exceeded the more densely gauged Java in correlation and NSE, because their sharper monsoonal seasonality is well captured by a linear month-specific model, whereas Java's convective, orographic, and urban variability and Maluku's fragmented maritime convection introduce non-linear signals that limit skill.

These findings confirm that integrating bias-corrected NASA POWER data with BMKG observations is a viable strategy for monthly rainfall estimation in data-sparse regions, supporting water-resource management, hydrometeorological disaster mitigation, and adaptive, data-driven public policy. Future work should extend the calibration to longer historical periods, incorporate additional meteorological and large-scale climate predictors (for example ENSO and MJO indices), apply station-specific bias correction where regional models leave high residual bias, and benchmark the linear framework against machine-learning methods to better represent the non-linear rainfall processes of the Indonesian Maritime Continent.

Declarations

Declaration of Generative AI Usage

During the preparation of this manuscript, the authors used a generative artificial-intelligence tool solely to assist with language editing and the structural design and formatting of the manuscript. The AI tool was not used to generate, compute, or interpret any of the data, analyses, regression models, or results reported in this study; all data processing, bias correction, regression modelling, evaluation, and scientific interpretation were performed and verified by the authors. The authors reviewed and edited all content and take full responsibility for the integrity and accuracy of the published work.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

C .C.: conceptualisation, overall modelling process, and manuscript writing. A. B. B.: result visualisation and overall manuscript writing. A. J. P. : data collection, modelling process, and manuscript writing. D. S. K., D. E. H.: conceptualisation, supervision, and review of the modelling process and the overall manuscript. K. F. D.; S. N.; K. C.; M. P. A. Ti.: Calculating and Modelling, Manuscript Writing

All authors have read and agreed to the published version of the manuscript.

Acknowledgements

The authors express their gratitude to the Indonesian Meteorological, Climatological, and Geophysical Agency (BMKG) for providing the ground-station meteorological data used in this research, and to the NASA Langley Research Center for the freely available NASA POWER meteorological dataset.

Data Availability

The meteorological data used in this study are publicly available. BMKG observation data can be obtained from the BMKG Data Online portal (<https://dataonline.bmkg.go.id/data-harian>), and NASA POWER data are available from the NASA POWER Data Access Viewer (<https://power.larc.nasa.gov/data-access-viewer/>).

References

1. Kurniadi A, Weller E, Salmond J, Aldrian E (2024) Future projections of extreme rainfall events in Indonesia. *Int J Climatol* 44(1):160–182. 10.1002/joc.8321
2. Juliastuti Y, Wijayanti AAS, Gunawan E, Irwansyah, Wulandari S (2024) Nature-based solutions for flood mitigation in metropolitan areas. *Eng Technol Appl Sci Res* 14(6):18896–18901
3. Ramadhan R et al (2022) Capability of GPM IMERG products for extreme precipitation analysis over the Indonesian Maritime Continent, *Remote Sens.*, vol. 14, no. 2, art. 412. 10.3390/rs14020412
4. Krisnayanti DS, Welkis DFB, Hepy FM, Legono D (2020) Evaluasi kesesuaian data Tropical Rainfall Measuring Mission (TRMM) dengan data pos hujan pada DAS Temef di Kabupaten Timor Tengah Selatan. *J Sumber Daya Air* 16(1):51–62. 10.32679/jsda.v16i1.646
5. Lufi S, Ery S, Rispiningtati (2020) Hydrological analysis of TRMM (Tropical Rainfall Measuring Mission) data in Lesti sub-watershed. *Civ Environ Sci* 3(1):18–30. 10.21776/ub.civense.2020.00301.3
6. Simare Mare DR, Hadiani RR Validation of TRMM rainfall data on slope stability in Karanganyar, Indonesia, in *Proc. 5th Int. Conf. Rehabilitation and Maintenance in Civil Engineering (ICRMCE 2021)*, 2022, pp. 101–112. 10.1007/978-981-16-9348-9_10
7. Ramadhan R et al (2022) Evaluation of GPM IMERG performance using gauge data over Indonesian Maritime Continent at different time scales, *Remote Sens.*, vol. 14, no. 5, art. 1172.

8. Tarek M, Brissette FP, Arsenault R (2020) Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrol Earth Syst Sci* 24(5):2527–2544. 10.5194/hess-24-2527-2020
9. Faisol A, Indarto E, Novita, Budiyo (2020) Komparasi antara Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) dan Global Precipitation Measurement (GPM) dalam membangkitkan informasi curah hujan harian di Provinsi Jawa Timur. *J Teknol Pertan Andalas* 24(2):148. 10.25077/jtpa.24.2.148-155
10. Tan ML et al (2023) Evaluation of NASA POWER and ERA5-Land for estimating tropical precipitation and temperature extremes, *J. Hydrol.*, vol. 624, art. 129940. 10.1016/j.jhydrol.2023.129940
11. Dharmayasa IG et al (2022) NASA POWER's: an alternative rainfall data resources for hydrology research and planning activities in Bali Island, Indonesia. *J Infrastruct Plan Eng* 1(1):1–7
12. Aprilia A, Wahidin AB, Abdurrahman AF, Prihanto S, Al-Fath Y (2025) Comparative validation of NASA POWER and ERA5 satellite-based meteorological data using BMKG observations in Bandar Lampung, Indonesia. *J Info Sains: Inf dan Sains* 15(2):360–369
13. Wangwongchai A, Waqas M, Dechpichai P, Hlaing PT, Ahmad S, Humphries UW (2023) Imputation of missing daily rainfall data: A comparison between artificial intelligence and statistical techniques, *MethodsX*, vol. 11, art. 102459. 10.1016/j.mex.2023.102459
14. Widyawati AS et al (2025) Performance of multivariate missing data imputation methods on climate data. *J Appl Inf Comput* 9(6):3953–3963
15. Sulistiyono M, Satria B, Sidauruk A, Wardhana R (2023) Rainfall prediction using multiple linear regression algorithm. *JITK (J Ilmu Pengetah dan Teknol Komput)*, 9, 1
16. Ahmed HAY, Mohamed SWA (2021) Rainfall prediction using multiple linear regression model, in 2020 Int. Conf. Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), pp. 1–5
17. Sedghnejad N, Nozari H, Marofi S (2024) Comparative analysis of classification techniques and input–output patterns for monthly rainfall prediction. *Water Sci* 38(1):192–208
18. Anuradha T, Aruna Sri PSG, RamaDevi J (2024) Hybrid model for rainfall prediction with statistical and technical indicator feature set, *Expert Syst. Appl.*, vol. 249, art. 123260. 10.1016/j.eswa.2024.123260
19. Sabu MKH, Yudianto D, Wijaya OT (2025) Evaluasi curah hujan berbasis data global pada DAS Wae Mese, Labuan Bajo. *J Sumber Daya Air* 21(1):15–30. 10.32679/jsda.v21i1.906
20. Cahyono C, Baria AB, Pierre AJ, Irawan AP (2026) Statistical estimation of monthly rainfall in Jakarta, South Tangerang, and Bandung: A multiple linear regression baseline using BMKG station data. *Eng Technol Appl Sci Res* 16(3):36979–36985. 10.48084/etasr.19328
21. Lubis SW et al (2022) Record-breaking precipitation in Indonesia's capital of Jakarta in early January 2020 linked to the northerly surge, equatorial waves, and MJO, *Geophys. Res. Lett.*, vol. 49, no. 22, art. e2022GL101513. 10.1029/2022GL101513

22. Nasution BI, Saputra FM, Kurniawan R, Ridwan AN, Fudholi A, Sumargo B (2022) Urban vulnerability to floods investigation in Jakarta, Indonesia: A hybrid optimized fuzzy spatial clustering and news media analysis approach, *Int. J. Disaster Risk Reduct.*, vol. 83, art. 103407. 10.1016/j.ijdrr.2022.103407
23. Rizal HM, Hariadi M, Arif YM, Warni E (2025) Enhancing flood prediction in urban areas: A machine learning approach for Makassar City. *Eng Technol Appl Sci Res* 15(2):21678–21684
24. Sapkota N, Khattri KB, Aryal D (2025) Modeling precipitation: A statistical and machine learning approach. *Int J Eng Technol* 2(2):188–203

Figures

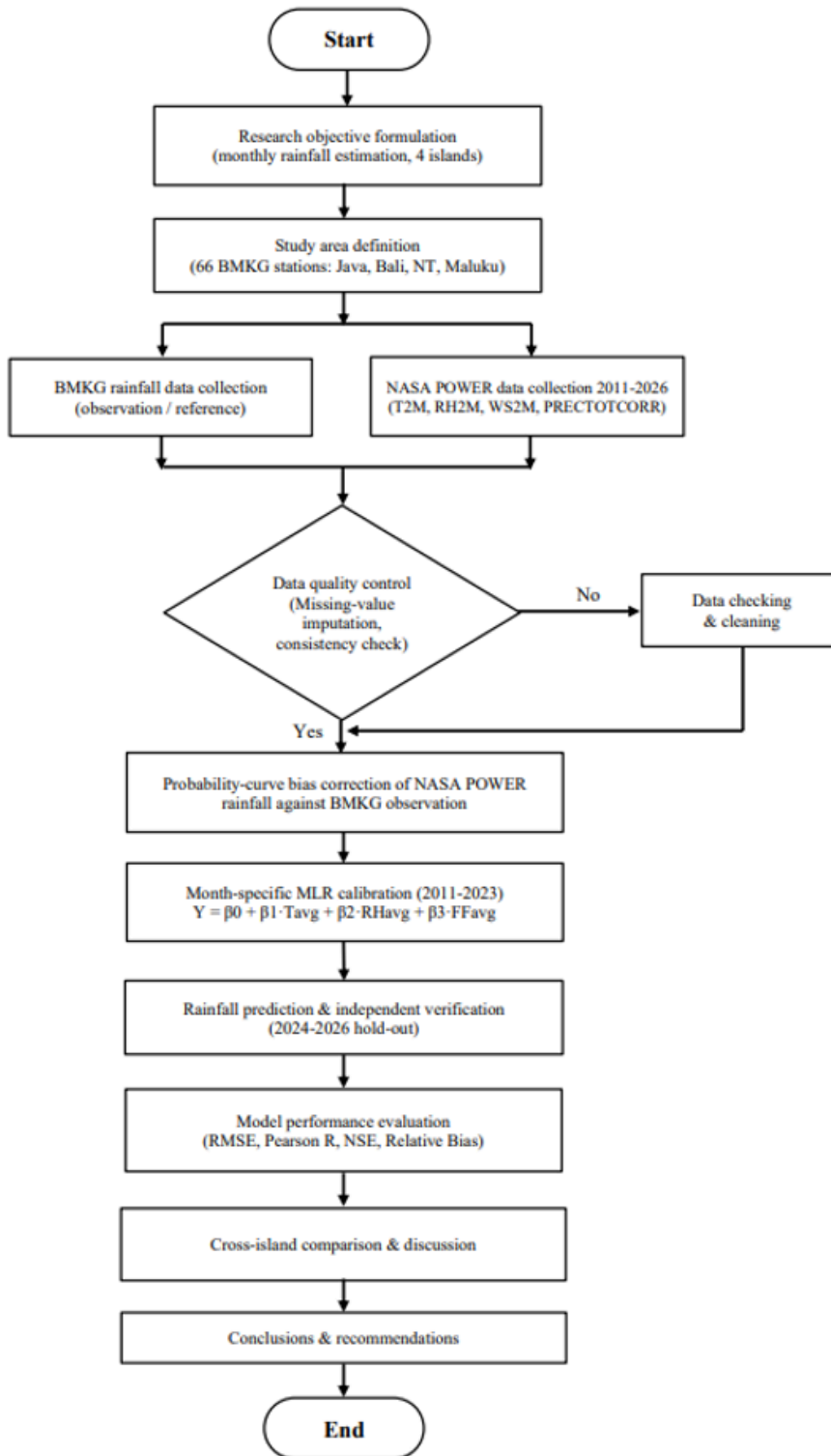


Figure 1

Flowchart of the month-specific MLR rainfall-estimation methodology.

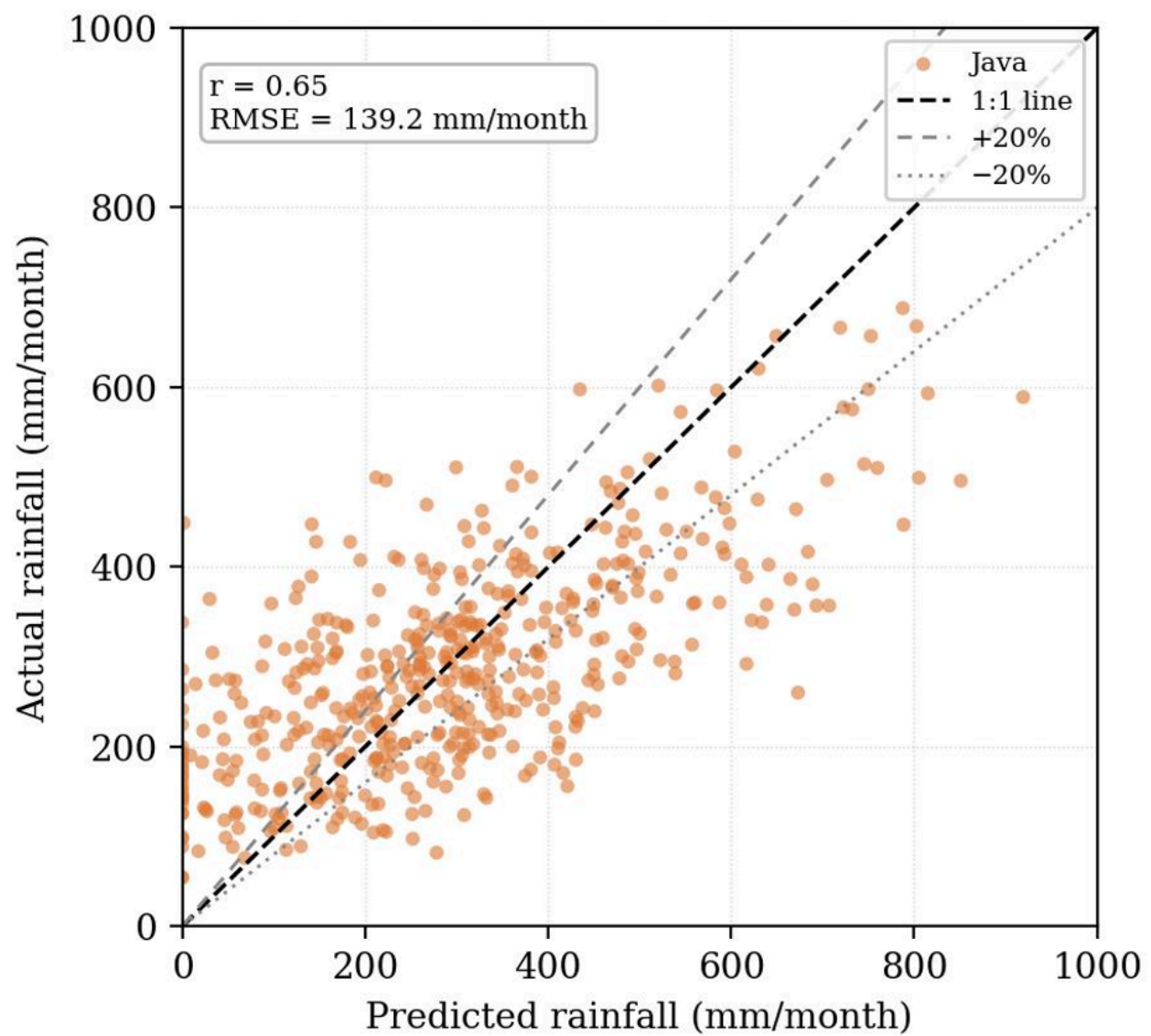


Figure 2

Actual versus predicted monthly rainfall for Java (1:1 line with $\pm 20\%$ bounds).

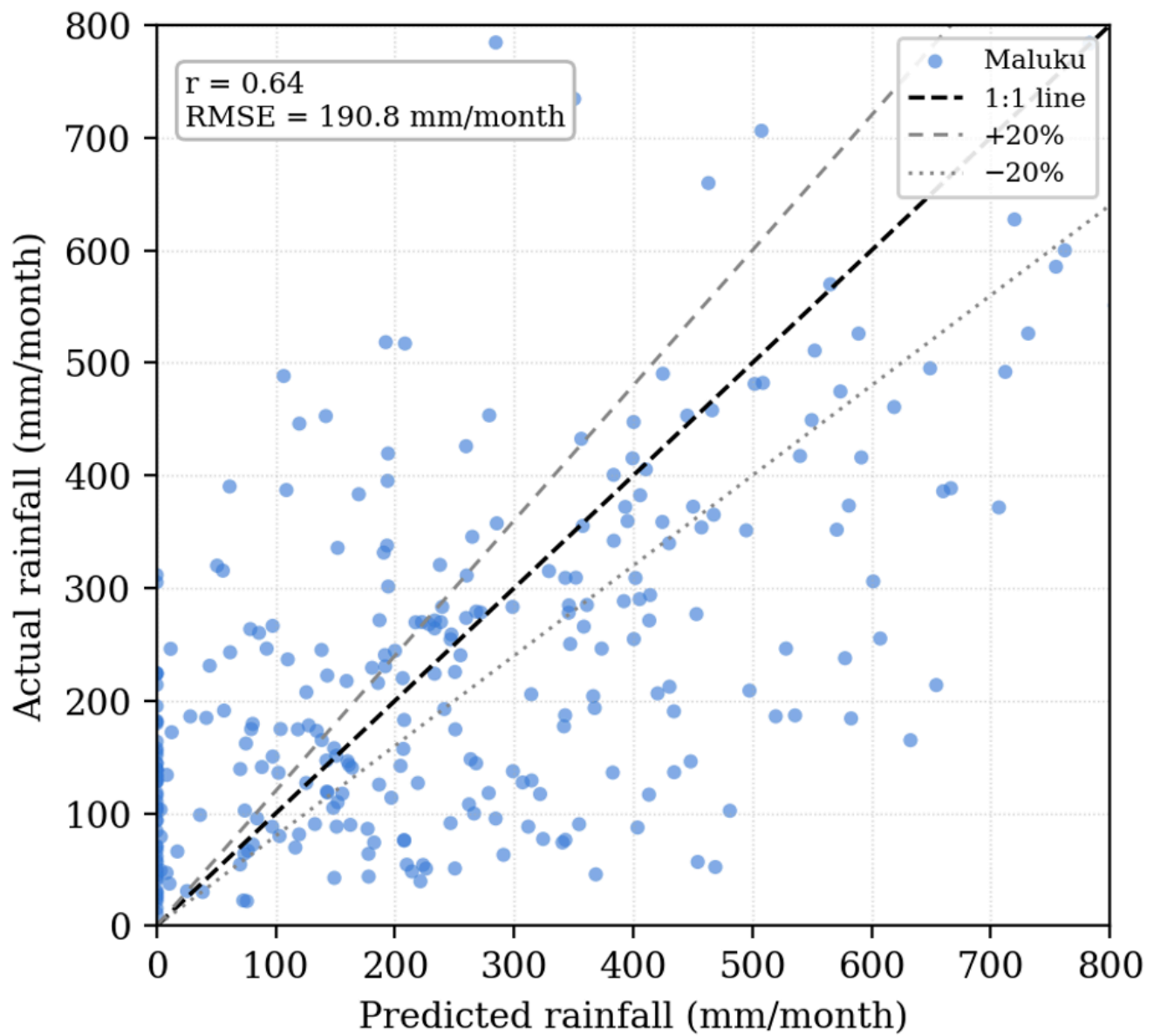


Figure 3

Actual versus predicted monthly rainfall for Maluku (1:1 line with $\pm 20\%$ bounds).

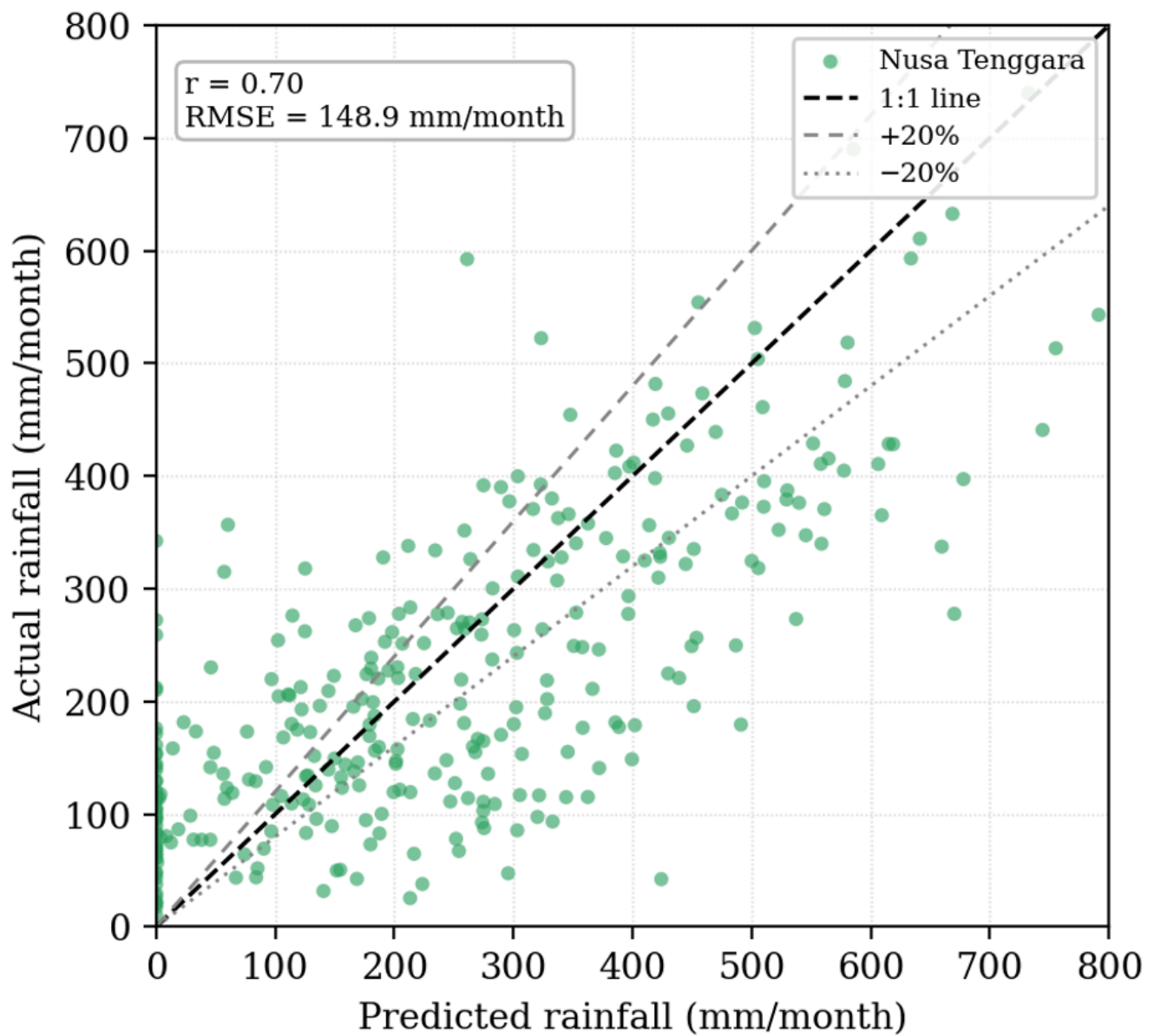


Figure 4

Actual versus predicted monthly rainfall for Nusa Tenggara (1:1 line with $\pm 20\%$ bounds).

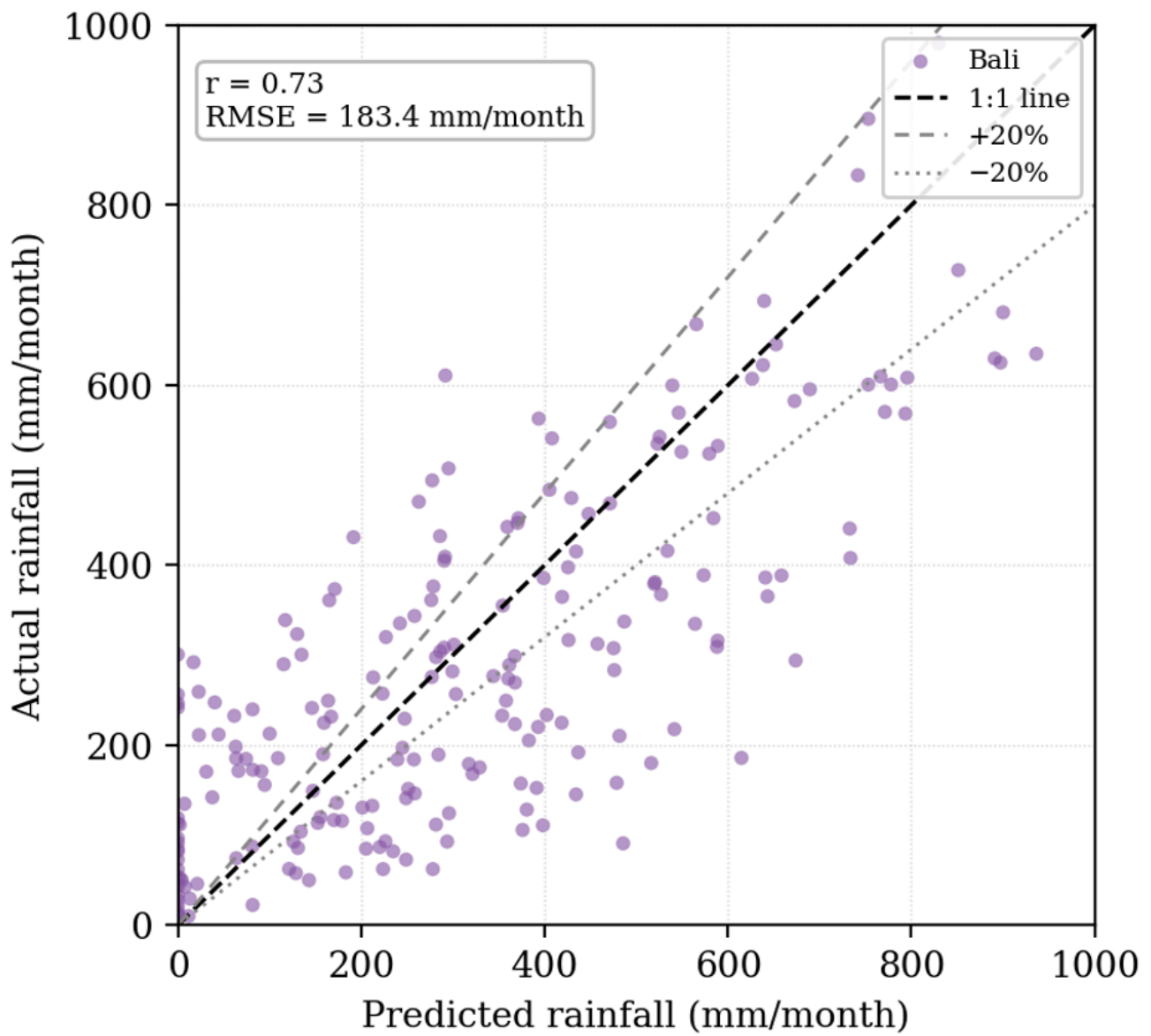


Figure 5

Actual versus predicted monthly rainfall for Bali (1:1 line with $\pm 20\%$ bounds).