Sumatra Journal of Disaster, Geography and Geography Education: June 2026. Vol.10. No.1. pp. 187-200
Disaster, Geography, Geography Education http://sjdgge.ppj.unp.ac.id/index.php/Sjdgge
ISSN: 2580 - 4030 (Print) 2580 - 1775 (Online), Indonesia

IMPROVING RAINFALL PREDICTION IN LAMPUNG, SOUTHERN SUMATRA USING WRF: THE ROLE OF BIAS CORRECTION AND ENSEMBLE TECHNIQUES FOR DISASTER MITIGATION

Andreas Boni Baik Simamora, *Alvin Pratama, Ridlo Wahyudi Wibowo, Farras Ghaly

Department of Atmospheric and Planetary Science, Faculty of Science, Institut Teknologi Sumatera, Jalan Terusan Ryacudu, South Lampung, Lampung, Indonesia
email: alvin.pratama@sap.itera.ac.id

*Corresponding Author, Received: April, 23, 2026. Revised: May 19, 2026. Accepted: June 17, 2026

ABSTRACT: Rainfall prediction in the tropics is highly challenging due to strong convective variability and limited observations. This study evaluates the Weather Research and Forecasting (WRF) model over southern Sumatra, using ERA5 and GFS-FNL as boundary datasets. Ten combinations of microphysics and cumulus parameterizations were tested, followed by bias correction, ensemble construction, and verification against AWS observations. Two bias correction methods were compared: Linear Scaling (LS) and Quantile Mapping (QM). LS consistently improved correlation and reduced RMSE, while QM often degraded performance. For rainfall intensity, ERA5-driven simulations outperformed GFS, with ERA5 Member 9 (WSM6 + BMJ) showing the highest skill (r = 0.91; RMSE ≈ 20 mm). Rainfall occurrence verification showed that ERA5 ensembles, particularly Ensemble 3 with POD–FAR weighting, achieved the most balanced skill (CSI ≈ 0.60), while GFS Member 5 occasionally surpassed the ensemble mean. ROC analysis revealed complementary strengths: ERA5 was more effective for moderate to heavy rainfall, whereas GFS showed greater sensitivity to light events. These results underscore ERA5's advantage for tropical downscaling and highlight the potential of multisource ensembles to improve rainfall prediction in data-scarce regions.

Keywords: WRF model, ERA5, GFS-FNL, bias correction, rainfall prediction

1. INTRODUCTION

Rainfall prediction in tropical regions is one of the most persistent challenges in atmospheric science, due to the highly dynamic, multiscale processes governing convective activity. Southern Sumatra, particularly Lampung Province, exemplifies such complexity, where rainfall variability is modulated by a combination of large-scale atmospheric circulation, regional topography, and local convective processes. These interactions produce rainfall distributions that are not only spatially heterogeneous but also temporally volatile, contributing to recurrent hydrometeorological hazards such as floods and landslides. The ability to improve rainfall prediction in this region is thus critical for advancing disaster preparedness, agricultural planning, and sustainable water resource management.

The meteorological setting of southern Sumatra is influenced by several large-scale climate modes. The El Niño Southern Oscillation (ENSO) alters seasonal rainfall variability by modulating convective activity across the Maritime Continent, while the Madden-Julian Oscillation (MJO) introduces subseasonal fluctuations that often trigger heavy rainfall episodes [1]. In addition, the Indian Ocean Dipole (IOD) has been shown to strongly influence rainfall extremes over Sumatra through anomalous sea surface temperature gradients that modify moisture transport [2]. These coupled ocean-atmosphere processes interact with mesoscale circulations, such as land/sea breezes and mountain/valley winds, that are prominent in Lampung's complex terrain. Together, they form a challenging environment for numerical weather prediction models to accurately capture rainfall.

Global numerical weather prediction (NWP) systems, including the Global Forecast System (GFS) and reanalysis products such as ERA5, provide essential boundary and initial conditions for regional simulations. While these datasets have progressively improved in resolution and accuracy, their ability to resolve mesoscale convection in tropical regions remains limited [3]. ERA5, produced by the European Center for Medium-

Range Weather Forecasts (ECMWF), offers a spatial resolution of 0.25° and improved representation of atmospheric dynamics compared with its predecessors. It has demonstrated superior performance in representing rainfall and temperature over Asia and the tropics relative to other reanalysis [4], [5]. Nevertheless, biases persist, particularly in convective rainfall regimes, necessitating the use of regional downscaling and post-processing techniques.

The Weather Research and Forecasting (WRF) model provides a versatile platform for dynamical downscaling and high-resolution weather prediction. Its flexibility lies in the variety of physical parameterization schemes available, particularly for cumulus convection and cloud microphysics, which strongly influence rainfall simulations. Numerous sensitivity studies have highlighted that small differences in parameterization can yield large variations in rainfall magnitude and spatial distribution in convective environments [6], [7]. For instance, the WSM6 microphysics and Kain-Fritsch cumulus schemes have frequently been identified as producing more realistic convective rainfall over tropical regions, including parts of Indonesia [8]. However, parameterization performance is often case dependent, reinforcing the need for local evaluations in regions such as Lampung.

In addition to model physics and input datasets, systematic biases in WRF rainfall simulations are a recurring challenge. These biases arise from model structural uncertainties and sub grid-scale process representations. Statistical bias correction has thus become an indispensable component of regional forecasting studies. Linear Scaling (LS) and Quantile Mapping (QM) represent two commonly employed approaches, each with distinct strengths. LS is effective at adjusting mean rainfall values while retaining natural variability, whereas QM better corrects mismatches in the distribution tails [9]. Recent studies in Indonesia found that LS often outperforms QM under tropical conditions, offering more consistent improvements in rainfall prediction [10]. This suggests that post-processing methods tailored to the local climatic regime can significantly enhance the usability of model outputs. Another avenue for improving forecast reliability is the use of ensemble approaches. By integrating multiple simulations that differ in parameterization, boundary datasets, or initial conditions, ensemble methods provide a probabilistic representation of rainfall outcomes. Weighted ensembles, which assign differential importance to members based on performance metrics, have been shown to reduce forecast errors more effectively than simple averaging [11]. In tropical regions, ensemble strategies are particularly valuable for capturing the uncertainty inherent in convective processes, yet their application in southern Sumatra remains underexplored. Local studies in Lampung have mostly focused on single-scheme sensitivity tests [12], with limited investigation of how weighted ensembles might improve prediction reliability.

Lampung's climatological and geographical setting further underscores the need for such investigations. The province features coastal plains along the Java Sea, the mountainous Bukit Barisan range, and lowland basins, all of which exert strong orographic influences on rainfall distribution. Flash floods and landslides have been recurrent in upland districts, while agricultural areas in the lowlands depend heavily on reliable rainfall for crop scheduling. Historical analyses indicate decreasing rainfall trends since the 1990s in some areas, often linked to ENSO events that prolong the dry season and intensify water stress [13]. At the same time, extreme rainfall events remain frequent, highlighting the dual challenges of managing both drought and flood risk [14]. Improving rainfall prediction in this context is therefore both a scientific and socio-economic imperative.

This study addresses these challenges by systematically evaluating the sensitivity of WRF rainfall simulations to input datasets and parameterization schemes over southern Sumatra. Specifically, it compares GFS and ERA5 as initial and boundary conditions, assesses the performance of multiple cumulus and microphysics combinations, and applies bias correction to reduce systematic errors. Furthermore, it develops and evaluates weighted ensemble approaches based on correlation-RMSE and POD-FAR metrics, aiming to determine whether such methods provide tangible improvements over simple ensemble means. By focusing on a representative rainfall event in Lampung, the study provides insights into the relative suitability of ERA5 versus GFS for local rainfall prediction and demonstrates the potential of ensemble weighting to enhance forecast skill.

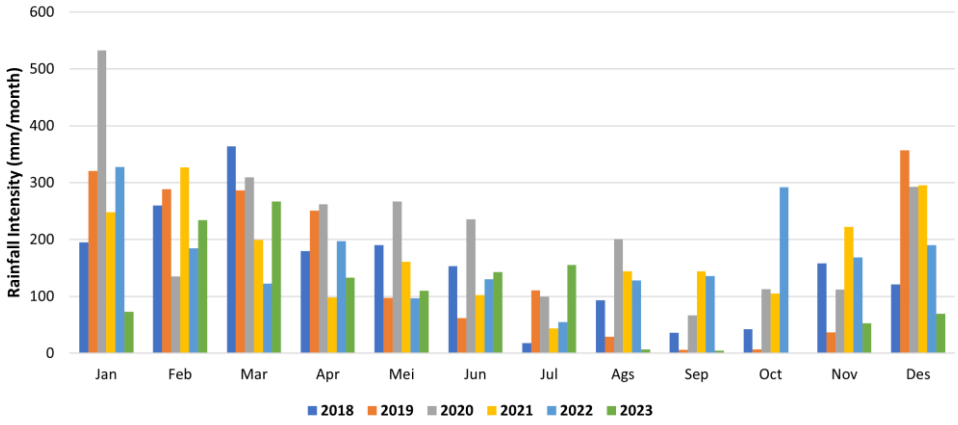
2. METHODS

2.1 Data and Methods

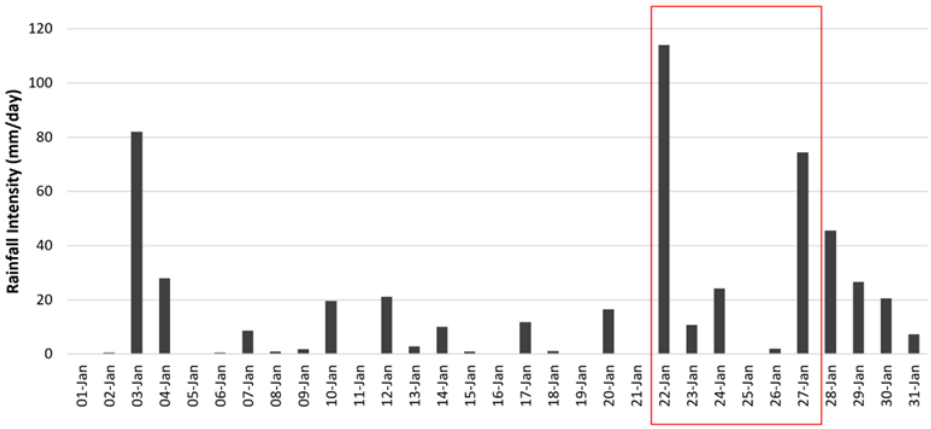
The study was conducted in Lampung Province, located at the southern tip of Sumatra, Indonesia, with observational reference data obtained from the Automatic Weather Station (AWS) operated by the Meteorology and Geophysics Unit of Institut Teknologi Sumatera (ITERA). The province is characterized by high rainfall variability influenced by both monsoonal circulation and local topographic effects. To determine the simulation period, a climatological screening was first performed using monthly rainfall accumulation data

Sumatra Journal of Disaster, Geography and Geography Education: June 2026. Vol.10. No.1. pp. 187-200
Disaster, Geography, Geography Education http://sjdgge.ppj.unp.ac.id/index.php/Sjdgge
ISSN: 2580 - 4030 (Print) 2580 - 1775 (Online), Indonesia

from 2018 to 2023. As shown in Figure 1 (a), January 2020 recorded the highest total monthly rainfall (>500 mm), making it an appropriate month for analyzing different rainfall intensities. Subsequently, daily rainfall patterns within January 2020 were examined to identify intervals that represent diverse rainfall classifications. Based on this analysis, the period 22–27 January 2020 was selected. The daily accumulation, illustrated in Figure 1 (b), covers a comprehensive range of rainfall categories: very heavy (22 January), heavy (27 January), moderate (24 January), light (23 and 26 January), and very light (25 January). This period, therefore, provides a robust dataset to test model sensitivity under multiple rainfall regimes. The selection ensures that the evaluation is not limited to a single extreme event, but instead captures the spectrum of tropical rainfall conditions typically observed in southern Sumatra.



(a)



(b)

Fig 1. (a) Monthly rainfall accumulation at ITERA (2018–2023), with January 2020 identified as the wettest month (>500 mm); (b) Daily rainfall accumulation at ITERA in January 2020, showing the selected period (22–27 January) that spans very light to very heavy rainfall events

Simulations were run 10 times for each input dataset (GFS and ERA5), yielding 10 members per dataset. Variations among members were introduced by altering the microphysics and cumulus parameterization schemes, while the radiation (Dudhia and RRTM) and planetary boundary layer (YSU) schemes were kept constant. A similar study was conducted by Ikranegara (2023) in the Soekarno–Hatta Airport area, Jakarta. Although the location differs from the present domain centered in Lampung, their findings showed that the combination of WSM6 microphysics, YSU planetary boundary layer, and Grell–Devenyi 3D cumulus schemes provided the best performance in representing weather variables [14]. This suggests that parameterization tests carried out in Jakarta can provide relevant insights and additional justification for the domain configuration adopted in this study.

This study employed multiple datasets to evaluate rainfall simulations over southern Sumatra during the selected period (22–27 January 2020). Four primary data sources were utilized.

- Observational rainfall data were obtained from the Automatic Weather Station (AWS) at Institut Teknologi Sumatera (ITERA). The AWS provides continuous measurements at 10-minute intervals, serving as the reference for validation of model outputs.
- Global Forecast System (GFS) Final Analysis (FNL) data from the National Centers for Environmental Prediction (NCEP) were used as one set of initial and boundary conditions for the Weather Research and Forecasting (WRF) model. The GFS dataset has a spatial resolution of 1.0° × 1.0° and a temporal resolution of 6 hours, and is available in GRIB2 format

Sumatra Journal of Disaster, Geography and Geography Education: June 2026. Vol.10. No.1. pp. 187-200
Disaster, Geography, Geography Education http://sjdgge.ppj.unp.ac.id/index.php/Sjdgge
ISSN: 2580 - 4030 (Print) 2580 - 1775 (Online), Indonesia

from the NCAR Research Data Archive. The dataset can be accessed through the NCAR Research Data Archive: https://rda.ucar.edu/datasets/d083002/

- Third, ERA5 reanalysis data from the European Center for Medium-Range Weather Forecasts (ECMWF) were employed as an alternative boundary condition. ERA5 provides improved spatiotemporal resolution (0.25° × 0.25°, 6-hourly) and includes both surface and pressure-level fields. Data were accessed via the Copernicus Climate Data Store using the official CDS API: https://cds.climate.copernicus.eu/

These combined datasets enable a comprehensive evaluation of WRF performance under different initial conditions and enhance the robustness of model verification against local observations.

2.2 WRF Model Configuration

The Weather Research and Forecasting (WRF-ARW) model was configured to investigate rainfall variability over southern Sumatra, with particular emphasis on the Lampung region. The model domain was designed using a Mercator projection and two nested grids, as illustrated in Figure 2. The outer domain was set at 9 km horizontal resolution, while the inner nest achieved 3 km resolution, with a parent-to-nest ratio of 3:1. This nesting strategy allows the model to capture both large-scale atmospheric circulation and local convective processes influenced by coastal and orographic effects. The simulation period covered 22–27 January 2020, with input and output intervals of 6 hours.

Fig 2. WRF model domain configuration with a two-way nested grid: the parent domain at 9 km resolution and the inner nest at 3 km resolution, centered over southern Sumatra

To examine the sensitivity of physical parameterizations, ten model members were generated using different combinations of cumulus and microphysics schemes. These configurations are summarized in Table 1. Among the tested schemes, well-established options such as Kain-Fritsch (KF), Betts-Miller-Janjic (BMJ), Grell-3D, and Tiedtke were used for cumulus parameterization, while the microphysics schemes included WSM6, Thompson, Morrison, and Goddard. A diverse set of schemes was selected to evaluate the model's ability to represent convective rainfall processes typical of tropical environments.

Table 1. List of cumulus and microphysics parameterization schemes used in the WRF model members for sensitivity testing

Member	Cumulus	Microphysics
1	-	WSM 6
2	-	Goddard
3	-	Thompson
4	Tiedtke	WSM 6
5	Kain Fritsch	WSM 3
6	Betts-Miller Janjic	Lin Purdue
7	Kain Fritsch	Eta Ferrier
8	Modified-Tiedtke	Morrison
9	Betts-Miller Janjic	WSM 6
10	Grell-3D	Goddard

Other physical options were held constant across all members to isolate the effects of cumulus and microphysics schemes. Specifically, the Dudhia scheme was applied for shortwave radiation, the Rapid Radiative Transfer Model (RRTM) for longwave radiation, and the Yonsei University (YSU) scheme for planetary boundary layer processes. This configuration ensured a consistent framework while enabling a rigorous assessment of cumulus–microphysics interactions. Through this design, the study aimed to identify the most suitable parameterization combinations for rainfall prediction in Lampung and to provide a robust foundation for subsequent ensemble modeling and bias correction.

2.3 Bias Correction and Ensemble Technique

Bias correction and ensemble approaches were applied to improve the reliability of WRF simulations. Numerical models often exhibit systematic errors due to uncertainties in initialization, physics representation, and spatial resolution. Two statistical methods were employed for bias adjustment. The first, Linear Scaling (LS), corrects model outputs by adjusting the mean relative to observations, as expressed in Equation (1). The second, Quantile Mapping (QM), aligns the cumulative distribution of model data with that of the observations, following Equation (2). Both methods were evaluated using correlation and Root Mean Square Error (RMSE), with the superior technique subsequently adopted for ensemble construction.

$$X_{corr} = X_{mod} \cdot \frac{\mu_{obs}}{\mu_{mod}} \quad \text{(Eq 1)}$$

$$X_{corr} = F_{obs}^{-1}\left(F_{mod}(X_{mod})\right) \quad \text{(Eq 2)}$$

Where: X_{corr} = bias-corrected value, X_{mod} = original model value, F_{mod} = cumulative distribution function (CDF) of model data, F_{obs}^{-1} = inverse CDF of observed data, μ_{obs} = observed mean, μ_{mod} = model mean

Ensemble modeling was then performed to reduce forecast uncertainty by integrating multiple WRF members. The simplest approach, Ensemble Scheme 1, computes the arithmetic mean of all members (Equation 3). A more advanced method, Ensemble Scheme 2, assigns weights based on correlation and RMSE, rewarding models with higher accuracy (Equations 4). Finally, a categorical performance–based ensemble (Scheme 3,) utilized detection metrics such as Probability of Detection (POD) and False Alarm Ratio (FAR) (Equation 5).

$$Y1_{ens} = \frac{1}{N}\sum_{n=1}^{N} y_n \quad \text{(Eq 3)}$$

$$Y2_{ens} = k_1y_1 + k_2y_2 + k_3y_3 + \cdots + k_ny_n \quad \text{(Eq 4)}$$
$$where: k_n = \frac{C_n + E_n}{2};$$
$$C_n = \frac{R_n}{\sum R}; E_n = \frac{S - RMSE_n}{(n-1)\,S}$$

$$Y3_{ens} = l_1y_1 + l_2y_2 + l_3y_3 + \cdots + l_ny_n \quad \text{(Eq 5)}$$
$$where: l_n = \frac{P_n + F_n}{2}; P_n = \frac{POD_n}{\sum POD};$$
$$F_n = \frac{T - FAR_n}{(n-1)\,T}$$

where y_n is the rainfall of the n −the member and N is the total number of members; R_n is the correlation of the n-th member and $RMSE_n$ is its root mean square error (Members with higher correlation and lower RMSE receive larger weights); POD_n and FAR_n are the detection and false alarm metrics of the n-th member (Higher POD and lower FAR lead to larger weights). By combining bias correction with weighted ensemble techniques, this study aimed to minimize systematic errors and enhance predictive skill, providing a more robust basis for rainfall forecasts over southern Sumatra.

2.4 Model Verification

Model verification was conducted using both continuous and categorical approaches to comprehensively assess WRF performance. Continuous verification focused on the ability of the model to reproduce rainfall intensity relative to observations. Two primary metrics were employed: the Pearson correlation coefficient (Equation 7), which measures the strength and direction of the linear relationship between simulated and observed rainfall, and the Root Mean Square Error (RMSE) (Equation 6), which quantifies the magnitude of prediction errors regardless of their sign. These statistics were further visualized using a Taylor diagram, which allows simultaneous comparison of correlations, standard deviations, and RMSEs with reference observations.

$$\mathbf{RMSE} = \left(\frac{\sum(y_i - \hat{y}_i)^2}{n}\right)^{\frac{1}{2}} \quad \text{(Eq 6)}$$

$$r = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad \text{(Eq 7)}$$

Categorical verification assessed the model's ability to correctly detect rainfall events. A contingency table (Table 2) was constructed to classify model outcomes into hits, misses, false alarms, and correct negatives. From this, three categorical scores were derived: Probability of Detection (POD) (Equation 8), which measures the fraction of observed rainfall events correctly predicted; False Alarm Ratio (FAR) (Equation 9), which evaluates the frequency of overpredicted rainfall; and the Critical Success Index (CSI)

Sumatra Journal of Disaster, Geography and Geography Education: June 2026. Vol.10. No.1. pp. 187-200
Disaster, Geography, Geography Education http://sjdgge.ppj.unp.ac.id/index.php/Sjdgge
ISSN: 2580 - 4030 (Print) 2580 - 1775 (Online), Indonesia

(Equation 10), which integrates both POD and FAR into a single performance measure.

Table 2. Contingency table for rainfall event verification, classifying model predictions against observations

	Observed Rain	Observed No rain
Model Rain	Hit	False Alarm
Model No Rain	Miss	Correct Negative

- Hit: Model correctly predicts rainfall when it occurs in observations.
- Miss: Model fails to predict rainfall that occurs in observations.
- False Alarm: Model predicts rainfall that does not occur in observations.
- Correct Negative: Model correctly predicts no rainfall when observations also show no rainfall.

$$POD = \frac{Hit}{Hit + Miss} \quad \text{(Eq 8)}$$

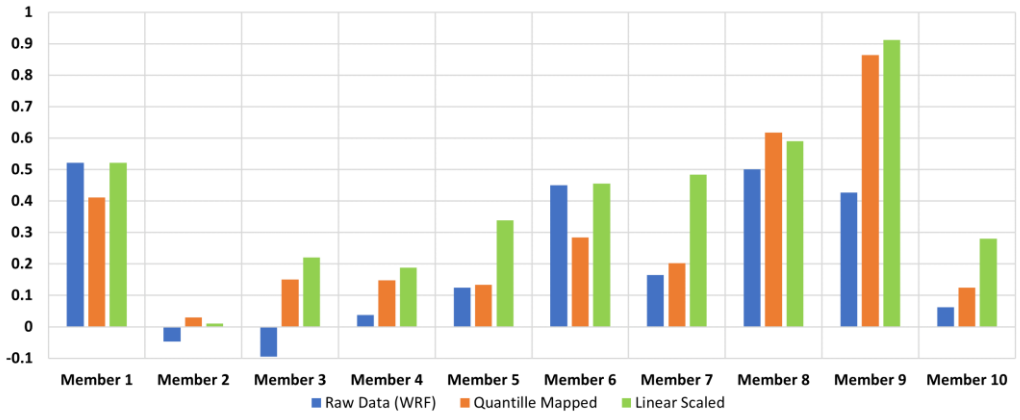
$$FAR = \frac{False\ Alarm}{False\ Alarm + Hit} \quad \text{(Eq 9)}$$

$$CSI = \frac{Hit}{Hit + False\ Alarm + Miss} \quad \text{(Eq 10)}$$

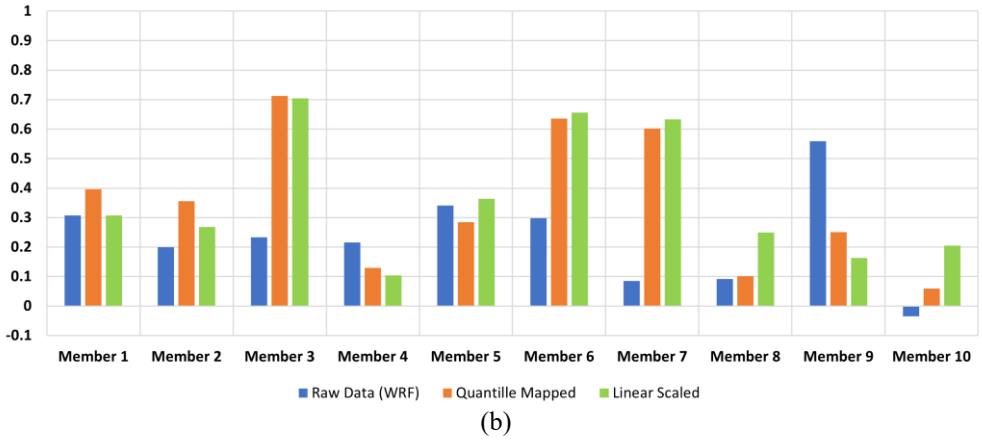
3. RESULTS AND DISCUSSION

3.1 Bias Correction Analysis

Numerical weather prediction models are often subject to systematic biases in rainfall simulations, particularly in regions dominated by convective processes, such as southern Sumatra. These biases arise from limitations in the initial and boundary conditions, as well as uncertainties in the model's physical parameterizations. To address this issue, two post-processing techniques, Linear Scaling (LS) and Quantile Mapping (QM), were applied to the raw outputs of WRF simulations driven by ERA5 and GFS. Their effectiveness was evaluated using correlation coefficients and Root Mean Square Error (RMSE), as illustrated in Figures 3 and 4.



(a)

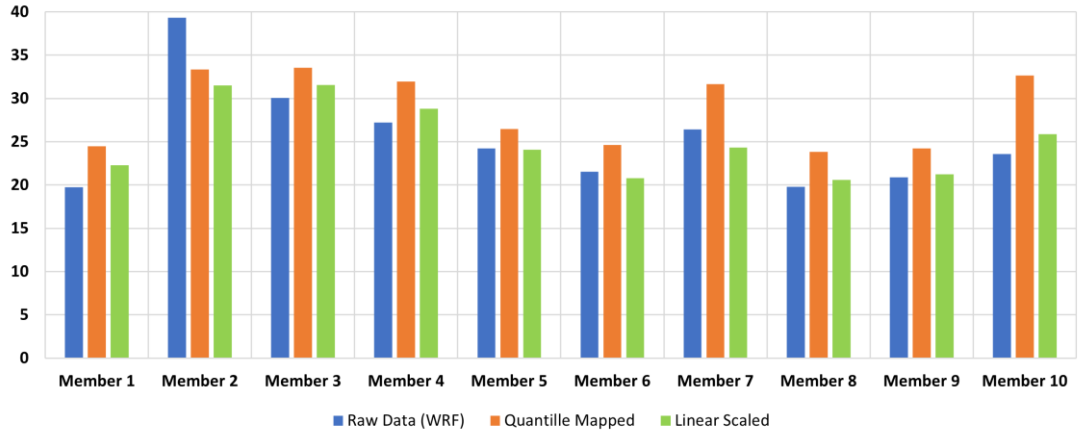


(b)

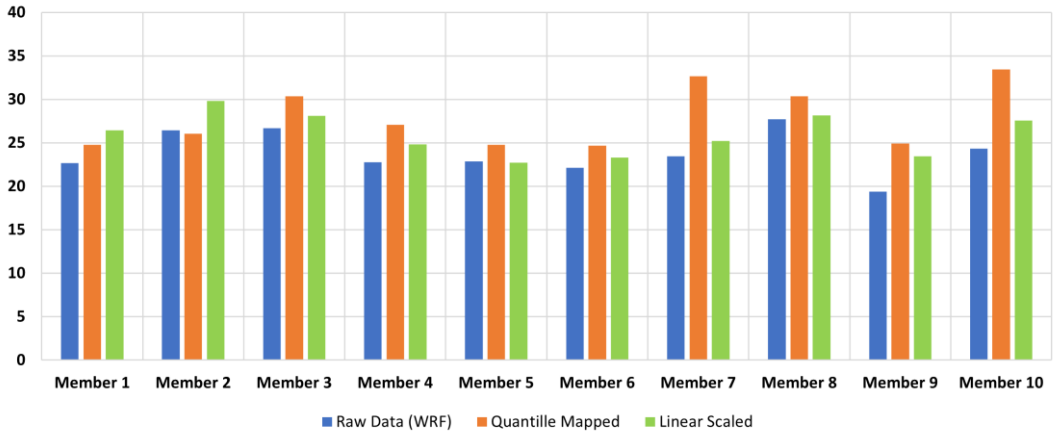
Fig 3. Correlation coefficients of WRF rainfall simulations before and after bias correction using Linear Scaling (LS) and Quantile Mapping (QM) for (a) ERA5 and (b) GFS-driven members.

Figure 3 shows that LS consistently improved correlation across most members, while QM frequently degraded performance. For ERA5-driven simulations (Fig. 3(a)), LS raised the correlation to ~0.40 on average, compared with only raising 0.30 after QM. GFS members displayed similar behavior (Fig 3(b)), with LS achieving ~0.35 correlation against 0.28 in the raw

outputs, while QM offered no significant benefit. These results highlight the ability of LS to preserve temporal coherence between simulated and observed rainfall, whereas QM occasionally distorted relationships due to overadjustment. Error magnitudes, shown in Figure 4, reinforce these findings. LS systematically reduced RMSE for both datasets, while QM sometimes increased it beyond the raw model values. For ERA5, RMSE decreased from 30.65 mm in the raw simulations to 25.10 mm after LS, whereas QM produced a higher error of 28.69 mm. Similarly, for GFS, RMSE declined from 29.14 mm to 25.95 mm with LS, but only to 27.92 mm with QM. This confirms that LS provides not only a stronger correlation but also a lower absolute error.



(a)



(b)

Fig 4. Root Mean Square Error (RMSE) of WRF rainfall simulations before and after bias correction using Linear Scaling (LS) and Quantile Mapping (QM) for (a) ERA5 and (b) GFS-driven members

The contrasting performance of LS and QM can be explained by data characteristics. LS adjusts only the mean, thereby avoiding overfitting and maintaining rainfall variability. QM, in contrast, attempts to match full distributions, but in tropical settings with short calibration records and high variability, this process can exaggerate biases, particularly for moderate and extreme rainfall. As a result, QM did not achieve its theoretical advantage in this case. Overall, the analysis demonstrates that bias correction is indispensable for rainfall prediction over southern Sumatra. Among the two tested methods, LS proved to be more reliable, stable, and effective than QM. Consequently, LS was adopted as the standard correction for subsequent ensemble construction and verification. This outcome aligns with recent evidence from Southeast Asia, where simple scaling approaches often outperform complex mapping techniques under conditions of data scarcity and strong convective variability [10]. In comparative terms, ERA5 combined with LS correction delivered the best overall performance, yielding higher correlations and lower RMSE than GFS with LS, underscoring ERA5's advantage as boundary forcing for tropical rainfall simulations.

3.2 Verification of Rainfall Intensity

Rainfall intensity verification was conducted to evaluate how well WRF simulations captured the magnitude and temporal variability of observed

rainfall over southern Sumatra. The analysis considered two boundary datasets (ERA5 and GFS), ten parameterization members for each, and a weighted ensemble. To construct weighted ensembles, individual WRF members were assigned weights (k) based on their statistical performance. The weighting approach prioritizes members with higher correlation and lower RMSE, ensuring that better-performing simulations contribute more to the ensemble mean. Table 3 originally presented the weights separately for ERA5 and GFS members. Here, the results are consolidated into a single comparison table to highlight differences between the two boundary datasets.

Table 3. Ensemble weights (k) for ERA5- and GFS-driven WRF members to construct Ensemble 2

Member	WRF-ERA5 Weight (*k*)	Member	WRF-GFS Weight (*k*)
1	0,159	1	0,120
5	0,130	5	0,190
6	0,151	6	0,266
7	0,152	7	0,257
8	0,171	9	0,167
9	0,237		

For ERA5-driven simulations, Member 9 (WSM6 + BMJ) received the largest weight (k = 0.237), reflecting its outstanding correlation (0.91) and low RMSE (~20 mm). Other high-performing members, such as Members 7 and 8, also received relatively large weights (0.150–0.180), while other performing members, including Member 5, were assigned minimal weight (0.13). This distribution demonstrates the dominance of a few configurations in the ERA5 ensemble, highlighting the stability of WSM6 microphysics and the BMJ/Kain–Fritsch cumulus schemes. In contrast, the GFS ensemble displayed a flatter distribution, with several members contributing nearly equally. The largest weight was assigned to Member 6 (k = 0.266), followed closely by Member 7 (k = 0.257) and Member 5 (k = 0.190). Other members, such as 1 (k = 0.120) and 9 (k = 0.167), had lower but still non-negligible contributions. Unlike ERA5, where a single member dominated, the GFS ensemble relied on multiple moderate contributors, reflecting greater variability in individual-member performance.

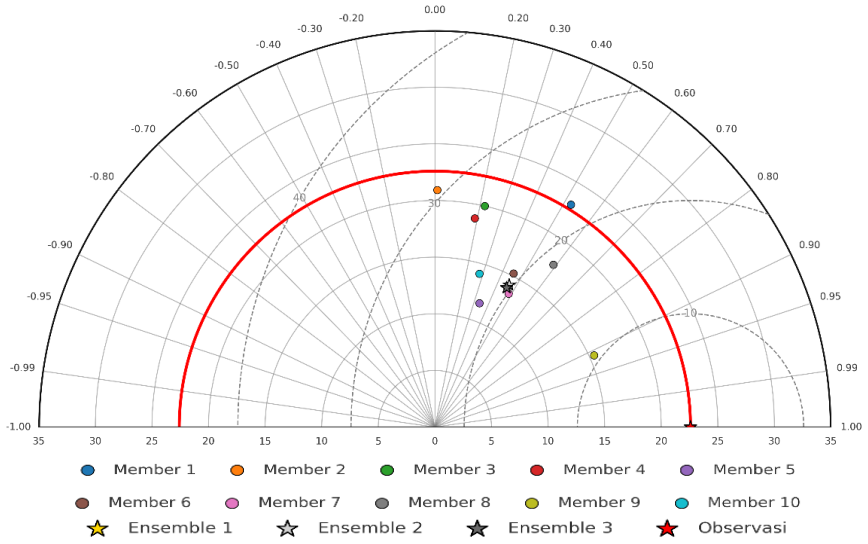


Fig 5. Taylor diagram and statistical verification of ERA5-driven WRF members for rainfall intensity during 22–27 January 2020.

ERA5-based members consistently achieved higher skill than those initialized with GFS. As shown in Figure 5, correlations for most ERA5 members ranged between 0.40 and 0.50, with several exceeding 0.70. The standout performer was Member 9 (WSM6 + BMJ), which produced a correlation of 0.91 and an RMSE of ~20 mm. Members using the WSM6 microphysics scheme, coupled with either BMJ or Kain-Fritsch cumulus parameterizations, performed particularly well, indicating their suitability for representing convective processes in tropical environments.

These findings are consistent with earlier studies in Southeast Asia. Fatmasari et al. (2017) also found that BMJ performed favorably in Lampung for convective rainfall [12], while Park et al. (2020) and Huang et al. (2020) highlighted the robustness of WSM6 microphysics in reproducing intense rainfall events in tropical and subtropical regions [6], [7]. The superior ERA5 performance is likely due to its higher resolution (0.25°) and assimilation of diverse observational data, which provide more accurate boundary conditions [3].

Simulations initialized with GFS displayed weaker performance overall (Figure 6). Correlations were generally clustered around 0.35–0.45, and RMSE values were slightly higher than for ERA5. The best configuration was Member 3 (BMJ + Lin Purdue), which achieved a correlation of ~0.70 and RMSE of 21 mm. Nevertheless, most GFS members did not match the accuracy of ERA5-driven members. This result aligns with recent intercomparisons that reported GFS products underestimate tropical convective activity compared with ERA5 [4], [5]. The coarser resolution of GFS (1.0°) may limit its representation of mesoscale features that strongly influence rainfall in southern Sumatra, such as sea-breeze convergence and orographic uplift along the Bukit Barisan range.

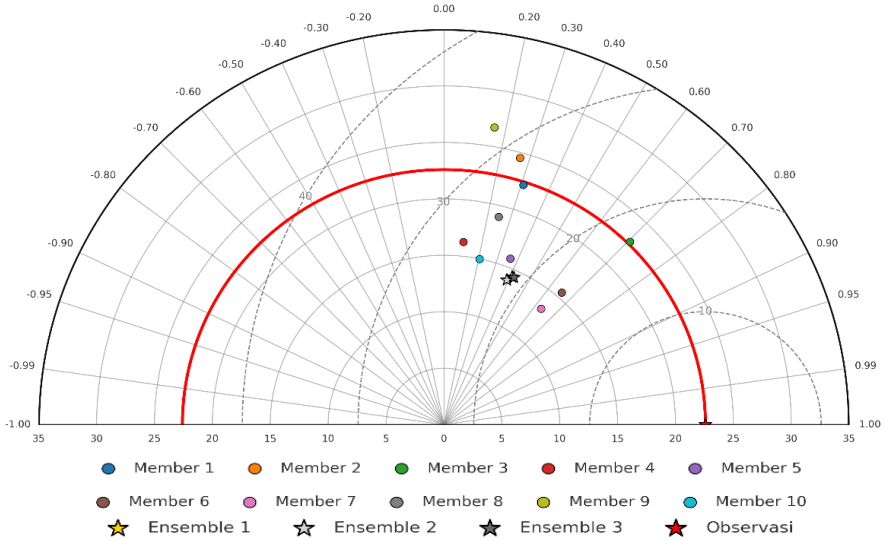


Fig 6. Taylor diagram and statistical verification of GFS-driven WRF members for rainfall intensity during 22–27 January 2020.

Comparing the two boundary datasets reveals a clear advantage for ERA5. Not only did ERA5-driven simulations achieve higher correlations, but they also showed lower RMSE values and clustered closer to the observational reference point in Taylor diagrams. For example, ERA5 members 7–9 exhibited both correlation values greater than 0.70 and normalized variability near unity, whereas GFS members rarely achieved these thresholds. These differences highlight the importance of boundary data quality in tropical rainfall simulations, echoing the findings of Stergiou et al (2022), who emphasized the sensitivity of WRF skill to initialization datasets [11].

Another outcome is the consistency of certain parameterizations across datasets. The WSM6 microphysics scheme repeatedly delivered strong performance, reflecting its detailed representation of mixed-phase cloud processes, which are critical for tropical convection [6]. Similarly, BMJ and Kain Fritsch cumulus schemes showed robust skill, likely due to their ability to represent convective triggering and moisture redistribution efficiently (see table 4). This aligns with Sofiati et al. (2025), who found these schemes to be among the most reliable for seasonal prediction across the Maritime Continent [8].

Table 4. Best-performing WRF members from ERA5- and GFS-driven simulations, showing correlation with observed rainfall

Boundary Dataset	Best Member Configuration	Correlation	Notes on Performance
ERA5	Member 9: WSM6 microphysics + BMJ cumulus	0,91	Highest overall skill; lowest RMSE (~20 mm)
GFS	Member 3: Lin Purdue microphysics + BMJ cumulus	~0,71	Best among GFS; RMSE ~21 mm

The variability among members under the same boundary conditions highlights parameterization uncertainty as a major factor in tropical rainfall prediction. While ERA5 Member 9 substantially outperformed the ensemble average, other members lagged, underscoring the need for ensemble approaches to mitigate uncertainty. Nonetheless, the identification of consistently superior schemes provides practical guidance: WSM6 with BMJ or Kain Fritsch should be prioritized in operational configurations for southern Sumatra. From a boundary dataset perspective, ERA5 provides stronger foundations for downscaling in this region than GFS. Its assimilation of global in-situ and satellite data more accurately captures large-scale moisture transport, which is critical for simulating localized convective rainfall [3].

3.3 Verification of Rainfall Occurrence

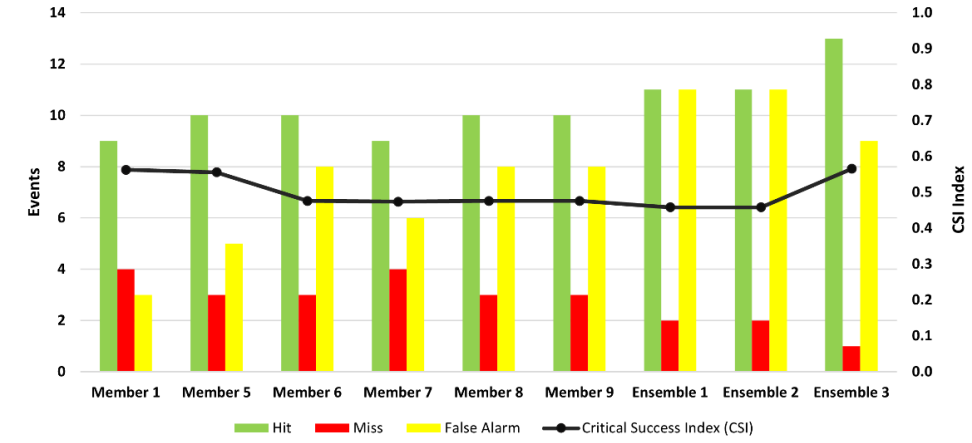
Beyond accurately representing rainfall intensity, numerical weather prediction models must also demonstrate their ability to correctly identify the occurrence or non-occurrence of rainfall events. This categorical verification is particularly important for operational applications in flood forecasting, early warning systems, and agricultural planning, where the detection of rainfall events can be as crucial as predicting their magnitude [15].

The weighting factors (l) for Ensemble 3, presented in Table 5, highlight contrasting characteristics between ERA5 and GFS-driven simulations. For ERA5, the weights were evenly distributed among selected members, ranging narrowly between 0.160 and 0.172, indicating that multiple members demonstrated comparable skill in detecting rainfall events. This balanced distribution suggests that ERA5 ensembles benefit from the collective strength of several reliable members rather than relying on a single configuration. By contrast, the GFS ensemble showed greater variability, with weights spanning from 0.150 to 0.210. Notably, Member 5 obtained the highest weight (l = 0.210), emphasizing its dominant role in driving ensemble skill. The remaining GFS members contributed less evenly, with values ranging from 0.140 to 0.180, underscoring a heavier dependence on one or two strong members. Taken together, these results confirm that ERA5 ensembles are stabilized by uniformly skillful members, whereas GFS ensembles are more sensitive to parameterization choices and disproportionately influenced by their best-performing member.

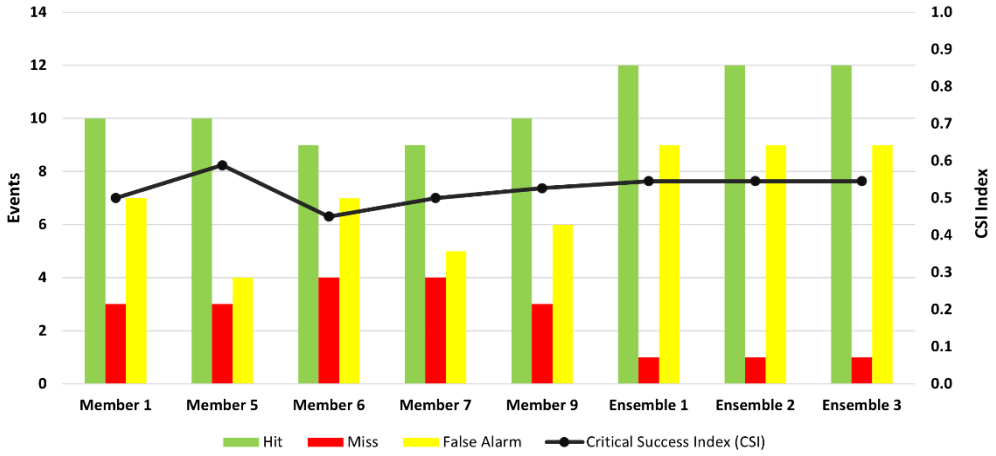
Table 5. Ensemble weights (l) for ERA5- and GFS-driven WRF members to construct Ensemble 3

Member	WRF-ERA5 Weight (*l*)	Member	WRF-GFS Weight (*l*)
1	0,167	1	0,202
5	0,172	5	0,210
6	0,167	6	0,189
7	0,160	7	0,195
8	0,167	9	0,204
9	0,167		

The results for ERA5-based WRF simulations indicate a strong ability to detect rainfall events while maintaining relatively low false alarms. As in Figure 7 (a), ERA5 ensembles displayed higher POD values, particularly for Ensemble 3, which combined weighting based on POD and FAR. This scheme achieved the best balance between successful detection and error minimization. Ensemble 3 recorded the highest CSI (~0.60), suggesting superior performance in capturing rainfall events while limiting false detections. Among individual members, Members 7, 8, and 9 showed consistently strong POD values, reflecting their skill in simulating convective initiation and rainfall triggering. However, these members occasionally experienced elevated FAR, particularly during periods of weak convection, which highlights the persistent challenge of overpredicting light rainfall in tropical regions. These findings align with previous studies in, where WRF simulations initialized with ERA5 have demonstrated improved event detection skills compared to other global datasets [16].



(a)



(b)

Fig 7. Categorical verification of rainfall occurrence (POD, FAR, CSI) for (a) ERA5- and (b) GFS-driven WRF simulations

For GFS-driven simulations, categorical performance was somewhat weaker overall. Figure 7 (b) shows that while several GFS members produced POD values comparable to ERA5, they often exhibited higher FAR, reducing their CSI values. Interestingly, Member 5 as an outlier: despite being a single member rather than an ensemble, it achieved one of the highest CSI scores (>0.60), surpassing even the ensemble means (ensemble 1). This suggests that for GFS, individual parameterization choices occasionally outperform weighted combinations, a finding also reported by Stergiou et al. (2022) in their WRF ensemble analysis over Europe [11]. The dominance of GFS Member 5 could be attributed to the combination of its cumulus and microphysics schemes, which proved particularly effective for moderate rainfall detection. Nonetheless, the general pattern indicated that GFS ensembles lacked the same level of stability as ERA5 ensembles. High FAR values were more frequent, pointing to overprediction of rainfall, a common limitation of global forecast datasets in the tropics [17].

A direct comparison between ERA5- and GFS-driven simulations highlights several key insights. First, ERA5 ensembles consistently delivered more stable and reliable detection skill. Ensemble 3 in ERA5 achieved the best CSI, reflecting an optimal trade-off between POD and FAR. This indicates that weighting schemes based on categorical scores can meaningfully enhance ensemble skill, a result consistent with findings from Ochoa-Rodriguez et al. (2019), who demonstrated that categorical weighting improved urban flood forecasting in convective storm environments [18]. The unexpected superiority of GFS Member 5 over GFS ensembles underscores the variability in parameterization sensitivity. While ensembles are generally designed to reduce uncertainty, they may occasionally dilute the performance of highly skilled members. This phenomenon has been noted in studies of tropical rainfall ensembles, where the spread of members can mask the skill of optimal configurations [19]. It suggests that ensemble design must carefully consider member selection rather than indiscriminately averaging all available members. The differences between ERA5 and GFS highlight the importance of boundary data in

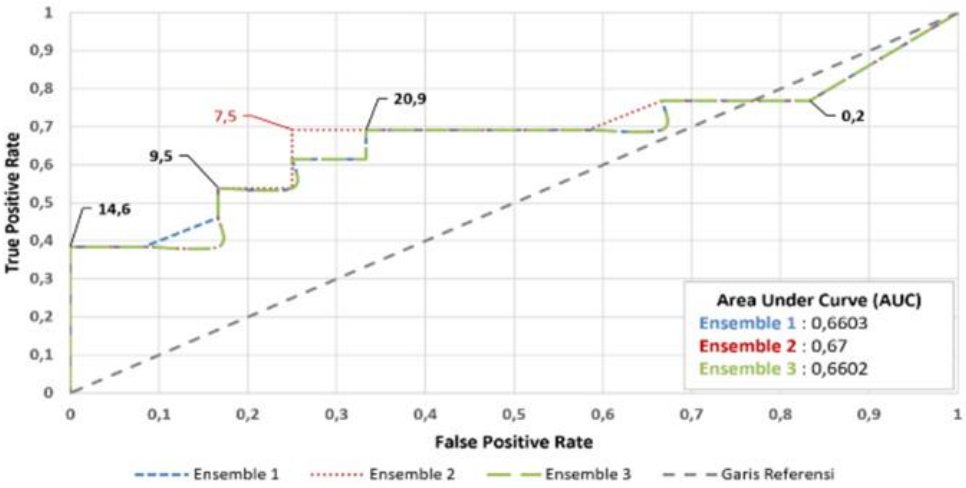
Sumatra Journal of Disaster, Geography and Geography Education: June 2026. Vol.10. No.1. pp. 187-200
Disaster, Geography, Geography Education http://sjdgge.ppj.unp.ac.id/index.php/Sjdgge
ISSN: 2580 - 4030 (Print) 2580 - 1775 (Online), Indonesia

categorical verification. ERA5's higher resolution and assimilated observations provided a better representation of the large-scale circulation and moisture convergence, leading to more accurate simulations of rainfall occurrence. This advantage was particularly evident in CSI scores, where ERA5 ensembles consistently outperformed GFS. These results reinforce earlier evaluations of ERA5 in the tropics [3], [5]. Finally, the results emphasize the importance of combining categorical and continuous verification approaches. While ERA5 clearly dominated in terms of intensity and occurrence detection.

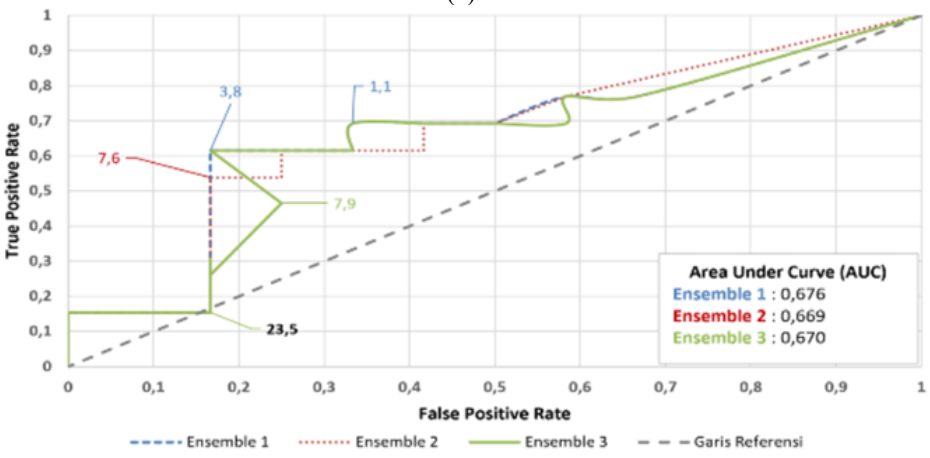
3.4 ROC (Receiver Operating Characteristics) Analysis

To complement the continuous and categorical evaluations, the performance of WRF simulations was further assessed using Receiver Operating Characteristic (ROC) curves. ROC analysis provides an integrated view of a model's ability to discriminate between rainfall and no rainfall events across varying thresholds. The curves in Figure 8 plot the True Positive Rate (equivalent to POD) against the False Positive Rate, with the Area Under the Curve (AUC) serving as a summary statistic. An AUC value closer to 1.0 indicates excellent discrimination, while a value near 0.5 reflects no skill [20].

ERA5-driven simulations yielded AUC values between 0.660 and 0.670, placing them in the "moderate skill" category (see Figure 8(a)). Importantly, the optimal probability thresholds for ERA5 ranged from 7.5 mm to 14.6 mm, suggesting that ERA5 is more effective at detecting moderate-to-heavy rainfall events. This aligns with findings from Southeast Asia where ERA5 has been shown to outperform forecast products in representing high-intensity convective rainfall [21], [22]. The results demonstrate ERA5's capability to capture large-scale moisture convergence and convective triggers associated with heavy precipitation.



(a)



(b)

Fig 8. ROC cureves and AUC values for (a) ERA5 and (b) GFS-driven WRF simulations, showing ERA5 skill at higher rainfall thresholds and GFS sensitivity to light rainfall events

By contrast, GFS-driven simulations achieved slightly higher AUC values, between 0.669 and 0.676 (see Figure 8(b)). However, their optimal thresholds were substantially lower, ranging from

0.2 mm to 7.6 mm, indicating stronger sensitivity to light and moderate rainfall. While this makes GFS valuable for early detection of light precipitation events, it also highlights its relative weakness in capturing intense convective episodes. Such findings are consistent with previous studies showing that GFS tends to produce more frequent low-intensity rainfall while underestimating heavy rainfall extremes [17].

The comparative analysis in Figure 8 thus reveals a complementary relationship between ERA5 and GFS. ERA5 is more reliable for forecasting heavy rainfall events, which often drive hydrometeorological hazards such as flooding and landslides, while GFS shows higher skill at identifying light to moderate rainfall, which can be relevant for agricultural applications. This complementarity suggests potential value in multi-source ensemble strategies, combining ERA5's strength in detecting extreme rainfall with GFS's capacity to identify light rainfall. Furthermore, the ROC analysis indicates that both ERA5 and GFS simulations achieved moderate but useful skill in distinguishing rainfall events from non-events. ERA5 is more skillful in higher-intensity thresholds, while GFS provides better sensitivity at lower thresholds. These results are consistent with international evaluations of reanalysis and forecast datasets, which emphasize that no single dataset is universally superior across all rainfall regimes [23], [24].

4. CONCLUSION

This study evaluated the performance of the WRF model in simulating rainfall over southern Sumatra using ERA5 and GFS as boundary datasets, with a focus on bias correction, rainfall intensity, rainfall occurrence, and ensemble construction. The results demonstrate that systematic errors in raw simulations can be effectively reduced using Linear Scaling (LS), which consistently improves correlation and lowers RMSE, while Quantile Mapping (QM) often introduces additional errors. In terms of rainfall intensity, ERA5-driven simulations outperformed GFS, with the best configuration (ERA5 Member 9, WSM6 + BMJ) achieving the highest performance. high correlation (0.91) and low RMSE (~20 mm). For rainfall occurrence, ERA5 ensembles, particularly Ensemble 3 based on POD–FAR weighting, delivered the most balanced skill, with CSI values near 0.60. In contrast, GFS performance was more variable, though individual members, such as GFS Member 5, occasionally exceeded the ensemble mean. ROC analysis further revealed that ERA5 was more effective at detecting moderate to heavy rainfall, while GFS was more sensitive to light rainfall events, suggesting complementary strengths between the two datasets.

Overall, these findings emphasize the advantage of ERA5 for high-resolution regional downscaling in the tropics, particularly when combined with LS bias correction and weighted ensemble methods. The robustness of WSM6 microphysics with BMJ or Kain–Fritsch cumulus schemes was consistently confirmed, providing practical guidance for operational forecasting. At the same time, the complementary characteristics of ERA5 and GFS highlight opportunities for multi-source ensemble approaches to improve rainfall prediction across the full spectrum of intensities. This study, therefore, contributes to advancing rainfall forecasting strategies in data-sparse tropical regions, with direct relevance for disaster risk management, water resource planning, and agricultural applications in southern Sumatra.

5. ACKNOWLEDGEMENTS

The author would like to thank Institut Teknologi Sumatera for providing the research grant (1998av/IT9.2.1/PT.01.03/2025) through "Hibah Penelitian Itera 2025".

6. REFERENCES

[1] P. Xavier, J. Gottschalck, and M. Wheeler, "Madden–Julian Oscillation and its predictability," *Nat. Rev. Earth \& Environ.*, vol. 3, no. 1, pp. 46–60, 2022, doi: 10.1038/s43017-021-00223-2.

[2] W. Cai *et al.*, "Climate impacts of the Indian Ocean Dipole," *Nat. Rev. Earth \& Environ.*, vol. 1, no. 1, pp. 51–64, 2019, doi: 10.1038/s43017-019-0010-6.

[3] B. Mueller, M. A. M. Maqueda, and S. Woolnough, "Evaluating ERA5 in the tropics," *Q. J. R. Meteorol. Soc.*, vol. 147, no. 739, pp. 2455–2472, 2021, doi: 10.1002/qj.4055.

[4] R. Liu and others, "Global-scale ERA5 product precipitation and temperature evaluation," *Ecol. Indic.*, vol. 166, p. 112481, 2024, doi: 10.1016/j.ecolind.2024.112481.

[5] Y. Xie, J. Wang, X. Wan, and Y. Lyu, "Assessing the monthly performance of daily precipitation products over Southeast Asia using the gauge-based analysis," *PLoS One*, vol. 20, no. 3, p. e0319477, 2025, doi: 10.1371/journal.pone.0319477.

[6] Y. Huang, Y. Wang, L. Xue, X. Wei, L. Zhang, and H. Li, "Comparison of microphysics parameterization schemes in the WRF model for an extreme rainfall event in the coastal metropolitan City of Guangzhou,

China," *Atmos. Res.*, vol. 240, p. 104939, 2020, doi: 10.1016/j.atmosres.2020.104939.

[7] J. Park, "Impact of cloud microphysics schemes on tropical cyclone forecast over the western North Pacific," *J. Geophys. Res. Atmos.*, vol. 125, no. 18, p. e2019JD032288, 2020, doi: 10.1029/2019JD032288.

[8] I. Sofiati and others, "Performance of weather research forecasting model for seasonal prediction of precipitation over Indonesian maritime continent," *Kuwait J. Sci.*, vol. 52, no. 1, p. 100293, 2025.

[9] P. Dhawan and others, "A comprehensive comparison of bias correction methods in climate model simulations: Application on ERA5-Land across different temporal resolutions," *Heliyon*, vol. 10, no. 23, p. e25123, 2024, doi: 10.1016/j.heliyon.2024.e25123.

[10] R. C. H. Hutauruk and others, "Evaluation of Bias Correction Method for Monthly Rainfall Prediction of ECMWF SEAS5 in Indonesia," *J. Meteorol. Geophys.*, vol. 25, no. 2, pp. 103–112, 2024, doi: 10.31172/jmg.2024.25.2.103.

[11] I. Stergiou, E. Tagaris, and R.-E. P. Sotiropoulou, "WRF physics ensemble performance evaluation over continental and coastal regions in Germany," *Atmosphere (Basel).*, vol. 14, no. 1, p. 17, 2022, doi: 10.3390/atmos14010017.

[12] E. Fatmasari, E. M. Saragih, E. Putra, and D. E. Nugraheni, "Sensitivity test of cumulus parameterization schemes in WRF for rainfall prediction in Lampung," in *IOP Conference Series: Earth and Environmental Science*, 2017, vol. 54, p. 12067. doi: 10.1088/1755-1315/54/1/012067.

[13] T. Manik, D. Rosadi, and S. Nurhayati, "Climate variability and rainfall trend analysis in Sumatra," *Procedia Environ. Sci.*, vol. 24, pp. 59–67, 2015, doi: 10.1016/j.proenv.2015.03.008.

[14] F. Welly, "The impact of topography on rainfall extremes in southern Sumatra," *Indones. J. Geogr.*, vol. 52, no. 3, pp. 215–229, 2020, doi: 10.22146/ijg.57355.

[15] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*, 4th ed. Cambridge, MA: Academic Press, 2019.

[16] T. Nguyen-Duy, T. Ngo-Duc, M. Choisy, I. P. Fernandez, and S. Sparrow, "Performance and added value of a high-resolution (2 km) rainfall product based on WRF-downscaled ERA5 for Ho Chi Minh City, Vietnam," *Theor. Appl. Climatol.*, vol. 156, no. 12, p. 656, 2025, doi: 10.1007/s00704-025-05894-1.

[17] O. P. Prat and B. R. Nelson, "Evaluation of precipitation estimates over the contiguous United States using multiple sensors and models," *J. Hydrometeorol.*, vol. 21, no. 3, pp. 569–587, 2020, doi: 10.1175/JHM-D-19-0128.1.

[18] S. Ochoa-Rodriguez, L.-P. Wang, A. Gires, and others, "Impact of high-resolution ensemble rainfall on urban flood modelling," *J. Hydrol.*, vol. 573, pp. 478–493, 2019, doi: 10.1016/j.jhydrol.2019.03.075.

[19] S. C. Chan, E. J. Kendon, N. M. Roberts, H. J. Fowler, and S. Blenkinsop, "The value of high-resolution rainfall ensembles for convective event prediction," *Weather Forecast.*, vol. 36, no. 2, pp. 431–450, 2021, doi: 10.1175/WAF-D-20-0131.1.

[20] Ş. K. Çorbacıoğlu and G. Aksel, "Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value.," *Turkish J. Emerg. Med.*, vol. 23, no. 4, pp. 195–198, 2023, doi: 10.4103/tjem.tjem_182_23.

[21] X. Yang, Q. Jiang, H. Chen, and X. Li, "Evaluation of ERA5 for rainfall simulation in Southeast Asia," *Atmos. Res.*, vol. 249, p. 105334, 2021, doi: 10.1016/j.atmosres.2020.105334.

[22] E. J. Becker, E. H. Berbery, and R. W. Higgins, "Understanding biases in reanalysis precipitation over the tropics," *Clim. Dyn.*, vol. 59, no. 5–6, pp. 1479–1496, 2022, doi: 10.1007/s00382-022-06057-3.

[23] R. Schiemann and others, "Evaluation of precipitation in global reanalyses," *J. Clim.*, vol. 33, no. 1, pp. 138–158, 2020, doi: 10.1175/JCLI-D-19-0552.1.

[24] X. Feng, A. Porporato, and I. Rodriguez-Iturbe, "Stochastic modeling of tropical rainfall: Implications for forecast evaluation," *J. Hydrometeorol.*, vol. 22, no. 2, pp. 345–359, 2021, doi: 10.1175/JHM-D-20-0178.1.