



Statistical Downscaling for Climate Change Studies: Comparing TCPG-GLM and Gradient-Boosted TCPG Models for Local-Scale Rainfall in Purwakarta

Fatkhurokhman Fauzi^{1, 2}, Anik Djuraidah^{1*}, Agus Mohamad Soleh¹, and Cici Suhaeni¹

¹ School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

² Department of Statistics, Universitas Muhammadiyah Semarang, Semarang, Indonesia

* Correspondence: anikdjuraidah@apps.ipb.ac.id

Citation:

Fauzi, F.; Djuraidah, A.; Soleh, A.M.; Suhaeni, C. Statistical downscaling for climate change studies: comparing TCPG-GLM and gradient-boosted TCPG models for local-scale rainfall in Purwakarta. *ASEAN J. Sci. Tech. Report.* **2026**, 29(9), e263108. <https://doi.org/10.55164/ajstr.v29i9.263108>

Article history:

Received: August 4, 2025

Revised: May 17, 2026

Accepted: June 1, 2026

Available online: July 26, 2026

Publisher's Note:

This article is published and distributed under the terms of the Thaksin University.

Abstract: Accurate rainfall projections are vital for climate adaptation and water resource planning in tropical regions, where rainfall patterns are highly variable and zero-inflated. This study evaluates the performance of two statistical downscaling approaches based on the Tweedie Compound Poisson-Gamma (TCPG) distribution for monthly rainfall modeling in Purwakarta, Indonesia: the generalized linear model with Tweedie distribution (GLM-Tweedie) and the nonlinear Tweedie distribution boosting model (TDboost). Using observational rainfall data from five stations and predictor variables from three Earth System Models (BNU-ESM, MIROC-ESM, and MIROC-ESM-CHEM), model performance was assessed using root mean squared error of prediction (RMSEP) and Pearson correlation coefficient. The results show that TDboost generally outperformed GLM-Tweedie in terms of predictive accuracy and temporal consistency. At Cibukamanah station, TDboost achieved the lowest RMSEP under the MIROC-ESM-CHEM scenario (90.63 mm), while its highest correlation was obtained under the BNU-ESM scenario ($r = 0.88$). Similarly, at Cisomang station, TDboost under the MIROC-ESM-CHEM scenario achieved an RMSEP of 109.75 mm and a correlation of $r = 0.78$, indicating improved predictive performance compared with GLM-Tweedie. These findings suggest that distribution-aware boosting techniques can better capture the nonlinear and zero-inflated characteristics of tropical rainfall. The results also indicate that the best-performing ESM may depend on the evaluation metric used, with MIROC-ESM-CHEM providing lower prediction errors at some stations and BNU-ESM showing strong temporal correlation when combined with TDboost. Overall, this study highlights the importance of combining appropriate statistical downscaling models with suitable ESM predictors for improving local-scale rainfall prediction in complex tropical environments.

Keywords: Climate change; generalized linear models; machine learning; statistical downscaling; tweedie boosting

1. Introduction

Rainfall prediction in tropical regions has become increasingly important because of the growing demand for accurate climate information to support food security, hydrometeorological disaster mitigation, and sustainable water resource management. Tropical areas, such as Indonesia, exhibit highly complex rainfall patterns influenced by multiscale atmospheric dynamics, including

regional phenomena such as the El Niño–Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD), which introduce significant uncertainty into seasonal rainfall [1, 2]. Moreover, local factors such as mountainous terrain, hilly landscapes, and microclimatic effects intensify spatial rainfall variability, as observed in Purwakarta Regency, West Java—a region characterized by complex geography and substantial intra-regional rainfall variation. These challenges are further exacerbated by the sparse observational networks and irregular rainfall distributions, which are poorly represented by coarse-resolution outputs from global climate models [3], [4].

This challenge is further compounded by the coarse spatial resolution of Earth System Models (ESMs), which typically operate at resolutions of tens to hundreds of kilometers [5]. Consequently, ESM outputs often fail to capture local rainfall heterogeneity, particularly in topographically complex areas such as Purwakarta Regency, West Java [6]. Such inaccuracies result in biased rainfall estimates that do not reflect actual field conditions. To bridge this gap, statistical downscaling methods are employed to link large-scale predictors from ESMs to local observational data through statistical relationships [7–9]. These approaches not only enhance spatial resolution but also correct systematic biases in ESM outputs based on historical data [5, 10].

Numerous studies have demonstrated that statistical downscaling produces more accurate rainfall projections in tropical regions when combined with statistical models that align with the unique distributional characteristics of rainfall, such as the Tweedie distribution [11]. Monthly rainfall in the tropics typically displays highly skewed, non-negative values with a substantial proportion of zeros [12]. These features are not well modeled using the normality assumptions underlying many classical regression models. In this context, the Tweedie Compound Poisson-Gamma (TCPG) distribution is an appropriate choice, as it captures both the discrete (rain occurrence) and continuous (rain amount) components within a single framework [13]. Furthermore, as part of the exponential family, the Tweedie distribution is compatible with the Generalized Linear Model (GLM) framework and offers flexibility in modeling zero-inflated and highly skewed data [14].

A study by Hayati et al. [15] revealed that applying GLMs with a Tweedie distribution significantly improves monthly rainfall prediction compared with conventional linear models. This flexibility is partly due to the index parameter (p), which lies within the range $1 < p < 2$, enabling the model to adapt its shape to match local data characteristics [16]. However, GLMs remain limited by the assumption of a linear relationship between predictors and the response variable. In tropical climate systems with complex, nonlinear interactions among sea surface temperature, atmospheric pressure, and global circulation indices, linear models often fall short of accurately capturing these dynamics [14].

To address this limitation, machine learning approaches such as tree-based boosting methods have been developed to capture nonlinear patterns and complex interactions among variables [17]. A notable example is Tweedie Boosting (TDboost), which combines decision tree structures with a loss function based on the Tweedie distribution [18, 19]. This method iteratively learns from complex data structures and significantly improves predictive performance, particularly for highly skewed or non-normally distributed data [11, 19].

Against this background, this study aims to evaluate and compare two statistical downscaling approaches based on the Tweedie distribution—namely, the Generalized Linear Model (GLM-Tweedie) and TDboost—for modeling monthly rainfall in Purwakarta Regency, Indonesia, a tropical region with high spatial complexity. Beyond comparing the performance of linear and nonlinear models, this study also evaluates the capability of three CMIP5-based Earth system models (BNU-ESM, MIROC-ESM, and MIROC-ESM-CHEM) to represent monthly rainfall at the local scale in Indonesia.

The main contribution of this study is not the development of a new boosting algorithm, but the evaluation of a distribution-aware nonlinear downscaling framework for local monthly rainfall in a tropical Southeast Asian setting. While TDboost has been introduced in previous statistical literature, its application for local-scale rainfall downscaling using ESM predictors in Indonesia remains limited. This study examines whether a Tweedie-based boosting model can improve the representation of zero-inflated, right-skewed, and nonlinear rainfall behavior compared with a linear Tweedie GLM under the same ESM input conditions. In addition, by comparing three CMIP5-based ESM scenarios across five rainfall stations, this study provides evidence on how the choice of ESM input may influence local downscaling performance in a complex tropical

region. Therefore, the novelty of this work lies in its regional application, distributional consistency, and comparative evaluation of linear and nonlinear Tweedie-based downscaling approaches for Southeast Asian rainfall modeling.

2. Materials and Methods

This study focuses on Purwakarta Regency, located in West Java, Indonesia, a region characterized by high spatial and temporal rainfall variability driven by orographic influences and tropical atmospheric dynamics. To capture this spatial heterogeneity, rainfall data were collected from five observation stations: Purwakarta, Pondok Salam, Cibukamanah, Cisomang, and Wanayasa. Monthly rainfall observations were obtained from the Meteorological, Climatological, and Geophysical Agency of Indonesia (BMKG) and spanned 16 years from January 2005 to December 2020. For modeling purposes, the dataset was divided into 144 months for training and the most recent 12 months (January–December 2020) for testing. Details of the data used are shown in Table 1.

Table 1. Summary of Data.

Data	Institution	Scale	Spatial Resolution
Station	BMKG	Monthly	-
Model for Interdisciplinary Research on Climate (MIROC-ESM) ^[20]	JAMSTEC	Monthly	2.8° × 2.8°
MIROC-ESM-CHEM (with Chemistry) ^[20]	JAMSTEC	Monthly	2.8° × 2.8°
Beijing Normal University (BNU-ESM) ^[21]	Beijing Normal University	Monthly	2.8° × 2.8°

As predictors in the statistical modeling process, this study used data from three global climate models (Earth System Models (ESMs)) under the historical scenario framework of the Coupled Model Intercomparison Project Phase 5 (CMIP5), namely BNU-ESM, MIROC-ESM, and MIROC-ESM-CHEM. These models were selected for their extensive application in regional climate studies and their demonstrated ability to capture tropical rainfall variability, particularly in the Southeast Asian region. MIROC-ESM and its variant, MIROC-ESM-CHEM, were developed by Japanese institutions with a strong focus on the Asian climate system, making their parameterizations relatively more sensitive to regional climatic conditions. In contrast, BNU-ESM is recognized for its robust performance in simulating monsoon dynamics and maritime tropical climates. Monthly ESM data were extracted from the nearest grid point to each station location without applying additional bias correction.

In this study, the predictor set used in the statistical downscaling models consisted of nine ESM-derived climate predictor variables, each extracted from the nearest grid point to each rainfall observation station. To account for short-term temporal dependence, one-month lagged predictors were also generated from the original ESM predictors. In addition, seasonal information was incorporated using three seasonal dummy variables representing DJF, MAM, and JJA, with SON treated as the reference season. Thus, the final predictor structure included the original ESM-derived predictors, their one-month lagged counterparts, and seasonal dummy variables. The same predictor structure was used for both GLM-Tweedie and TDboost to ensure a fair comparison between the two statistical downscaling approaches. No separate variable selection or dimensionality reduction procedure was applied before modeling.

To model monthly rainfall data characterized by nonlinearity, zero-inflation, and a tendency toward overdispersion, a statistical approach capable of flexibly capturing these distributional characteristics is required. Therefore, this study implemented two modeling approaches based on the Tweedie distribution: the Generalized Linear Model (GLM-Tweedie), representing a classical linear model, and Tweedie Distribution

Boosting (TDboost), a nonlinear model based on a boosting algorithm. These two models were selected for comparative evaluation in terms of their predictive performance on rainfall data derived from a combination of ground-based observations and global climate model projections, to assess the effectiveness of each approach in statistically representing tropical rainfall dynamics.

2.1 GLM-Tweedie

The Generalized Linear Model (GLM) with a Tweedie distribution is a widely used statistical approach for analyzing data that are non-negative, highly right-skewed, and contain a substantial number of zeros—such as monthly rainfall data in tropical regions. The main advantage of the Tweedie distribution lies in its ability to represent a compound distribution as a compound Poisson–Gamma distribution, particularly when the power index parameter $1 < p < 2$. In this range, the distribution is referred to as the Tweedie Compound Poisson-Gamma (TCPG), which is especially well-suited for rainfall data that include both zero events (no rainfall) and positive continuous rainfall amounts. In the GLM-Tweedie framework, the expected value of the response variable Y_i is linked to the linear predictor via a logarithmic link function:

$$\log(\mu_i) = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} \quad (1)$$

where μ_i is the expected value of the response variable Y_i , $\mathbf{x}_i$ is the vector of predictor variables (climate features), and $\boldsymbol{\beta}$ is the vector of regression coefficients. The relationship between the variance and the mean is assumed to follow the power-law form:

$$\text{Var}(Y_i) = \phi \mu_i^p \quad (2)$$

where ϕ is the dispersion parameter and p is the Tweedie power index that governs the distributional shape, this formulation allows the model to accommodate overdispersion and zero inflation, which are commonly observed in rainfall data.

The parameters $\boldsymbol{\beta}$, ϕ , and p are estimated using the Maximum Likelihood Estimation (MLE) approach. The log-likelihood function of the Tweedie distribution for a single observation can be expressed as follows:

$$\log f(y_i, \mu_i, \phi, p) = \frac{1}{\phi} \left(\frac{y_i \mu_i^{1-p}}{1-p} - \frac{\mu_i^{2-p}}{2-p} \right) + \log a(y_i, \phi) \quad (3)$$

Since the normalization function $a(y_i, \phi)$ does not have a closed-form expression, the likelihood must be computed numerically. Estimation of the regression coefficients $\boldsymbol{\beta}$ is typically performed using the Iteratively Reweighted Least Squares (IRLS) algorithm. Meanwhile, the Tweedie power index p is commonly selected via a profile likelihood approach, in which a range of candidate values of p is evaluated, and the value that maximizes the overall log-likelihood is chosen.

2.2 TDboost

To model rainfall data characterized by nonlinearity, overdispersion, and a high frequency of zero values, the Tweedie Distribution Boosting (TDboost) approach was developed by Yang et al. [19] as an extension of the gradient boosting algorithm, specifically designed to accommodate a loss function based on the Tweedie Compound Poisson-Gamma (TCPG) distribution. This model is tailored to predict the mean value μ of the response variable under the assumption of non-negative and zero-inflated outcomes, without relying on the linear relationship assumption inherent in GLM-Tweedie. The strength of TDboost lies in its ability to iteratively build predictive models using decision trees, where each iteration aims to minimize the negative log-likelihood of the Tweedie distribution. By defining a loss function that matches the data distribution, TDboost adjusts the predicted estimates to align with the statistical characteristics of the response while simultaneously capturing nonlinearities and complex interactions among predictors. The following is the TDboost algorithm as developed by Yang et al. [19]:

TDboost Algorithm

Let $\{(x_i, y_i)\}_{i=1}^n$ be the training dataset, where x_i is a vector of predictors and y_i is the response variable (monthly rainfall). The algorithm proceeds as follows:

Inputs:

- Training dataset $\{(x_i, y_i)\}_{i=1}^n$
- Tweedie power index p and dispersion parameter ϕ
- Number of boosting iterations M
- Learning rate $\nu \in (0, 1]$
- Maximum tree depth J

Steps:

1. **Initialize** the prediction function:

$$\hat{f}^{(0)}(x) = \log(\bar{y})$$

where $\bar{y}$ is the average of the response variable.

2. **For** each boosting iteration $m = 1, 2, \dots, M$, perform the following:
 - Compute the current predicted mean:

$$\hat{\mu}_i^{(m-1)} = \exp(\hat{f}^{(m-1)}(x_i))$$

- Calculate the pseudo-residuals using the derivative of the Tweedie deviance:

$$r_i^{(m)} = \hat{\mu}_i^{(m-1)} \left(\frac{y_i}{(\hat{\mu}_i^{(m-1)})^p} - (\hat{\mu}_i^{(m-1)})^{1-p} \right)$$

- Fit a regression tree $\hat{h}^{(m)}(x)$ to the pseudo-residuals $r_i^{(m)}$
 - Update the prediction function:

$$\hat{f}^{(m)}(x) = \hat{f}^{(m-1)}(x) + \nu \hat{h}^{(m)}(x)$$

Final prediction for the mean response:

$$\hat{\mu}(x) = \exp(\hat{f}^{(M)}(x))$$

2.3 Model Evaluation

To evaluate and compare the predictive accuracy of the two statistical downscaling approaches—GLM-Tweedie and TDboost—this study employed two performance metrics: Root Mean Square Error of Prediction (RMSEP) and the Pearson correlation coefficient (r). These metrics were calculated using the testing set, consisting of the most recent 12 months of the rainfall dataset, which was not used during model training. The use of a 12-month testing period was intended to evaluate the models over one complete annual rainfall cycle, which is relevant for Indonesia's seasonal climate regime characterized broadly by wet and dry seasons. Therefore, the testing period includes the main seasonal transition within a single year. However, this design should be interpreted as an initial independent evaluation rather than a comprehensive assessment of interannual generalization. This data separation was intended to ensure that the model evaluation reflects predictive performance on unseen data, a critical aspect for operational climate projections. The RMSEP was calculated using the following equation:

$$\text{RMSEP} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

where $\hat{y}_i$ denotes the predicted value from the model, y_i represents the actual observed value, and n is the number of observations in the testing dataset. A lower RMSEP value indicates smaller average prediction

errors relative to the actual data, signifying better quantitative accuracy of the model. Meanwhile, the Pearson correlation coefficient (r) is calculated using the following equation:

$$r = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

where $\bar{\hat{y}}$ and $\bar{y}$ represent the mean of the predicted and observed values, respectively. A correlation coefficient closer to 1 indicates a strong, positive linear relationship between the predicted and observed values, suggesting consistent patterns between the two series.

Model evaluation was conducted separately for each rainfall station under the three ESM scenarios: BNU-ESM, MIROC-ESM, and MIROC-ESM-CHEM. The results of RMSEP and correlation metrics offer a comprehensive view of the spatial and temporal performance of both modeling approaches. Overall, TDboost demonstrated superior predictive performance, yielding more accurate and highly correlated predictions, particularly in regions with high rainfall variability.

3. Results and Discussion

3.1 Exploration of Rainfall Data

Understanding the seasonal distribution of rainfall is critical for effective water resource management, agricultural planning, and climate adaptation strategies in tropical regions such as Indonesia. This study investigates the intra-annual variability of monthly rainfall across five meteorological stations in West Java — Wanayasa, Cisomang, Pondok Salam, Cibukamanah, and Purwakarta—using boxplots (Figure 1). These locations represent diverse topographical and climatic zones, enabling a comparative assessment of spatial rainfall patterns and seasonal behavior.

The results reveal a predominantly monsoonal rainfall pattern across the five stations, characterized by a single main wet season and a single main dry season within the annual cycle. The wet season generally runs from January to March, while the dry season runs from July to September. Based on Figure 1, peak rainfall occurs at the beginning of the year, with median values exceeding 600 mm/month, particularly at Wanayasa and Pondok Salam stations. In contrast, July and August show very low rainfall, with median values approaching zero, especially at Cisomang and Cibukamanah. These seasonal fluctuations reflect the influence of the monsoonal system over western Indonesia and are consistent with a unimodal annual rainfall regime in Purwakarta Regency. Although rainfall regimes across the Indonesian maritime continent may vary from unimodal to bimodal depending on location, longitude, and the influence of the ITCZ, the observed rainfall distribution in Purwakarta is better described as monsoonal, with a single dominant wet season.

Among the five stations, Pondok Salam and Wanayasa exhibit the highest monthly rainfall variability, likely due to orographic enhancement resulting from their proximity to mountainous terrain. Meanwhile, Purwakarta station consistently records lower rainfall, with monthly median values generally below 400 mm. These spatial differences underscore the importance of local topographic and atmospheric dynamics. These findings are supported by previous studies indicating that rainfall patterns in this region are influenced by large-scale phenomena such as the Australasian monsoon and ENSO events [22], as well as mesoscale convective processes in areas with complex topography [23].

Extending the previous analysis, the rainfall density distributions from the five stations provide a more nuanced view of the monthly rainfall frequency and distributional patterns (Figure 2). While seasonal patterns were clearly observed in the boxplots, the density plots illustrate variations in distributional shape across locations. For example, Purwakarta shows a high-density peak at lower rainfall levels (below 200 mm), suggesting that light rainfall events are much more common. Conversely, stations like Pondok Salam and Cibukamanah exhibit broader distributions with peaks in the mid-range (200–500 mm) and right-skewed tails, indicating that extreme rainfall events may occur, although infrequently.

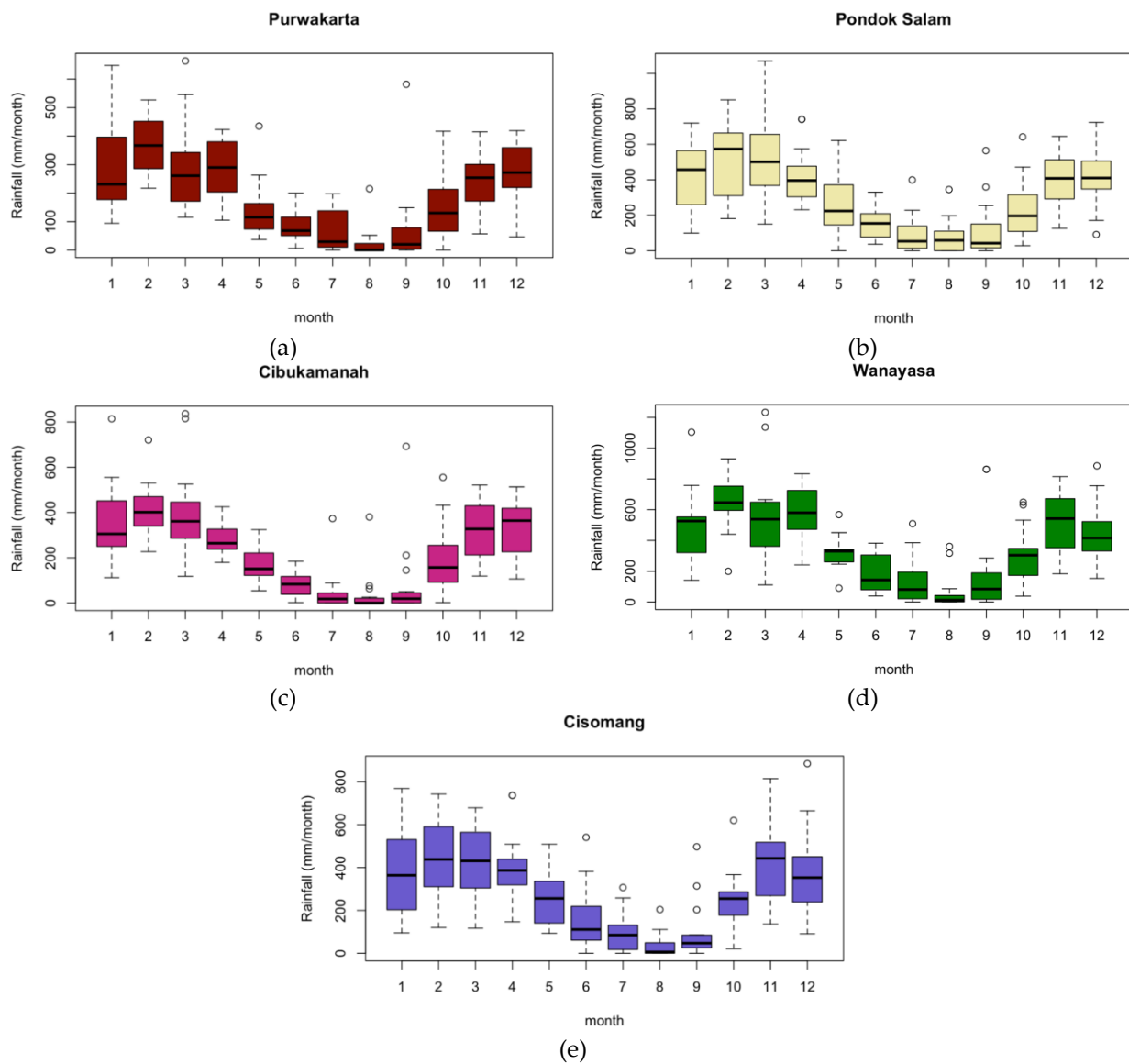


Figure 1. Monthly rainfall distribution patterns at five meteorological stations: (a) Purwakarta, (b) Pondok Salam, (c) Cibukamanah, (d) Wanayasa, and (e) Cisomang. Boxplots show intra-annual rainfall variability, with marked differences in intensity and seasonal rainfall distribution among the stations.

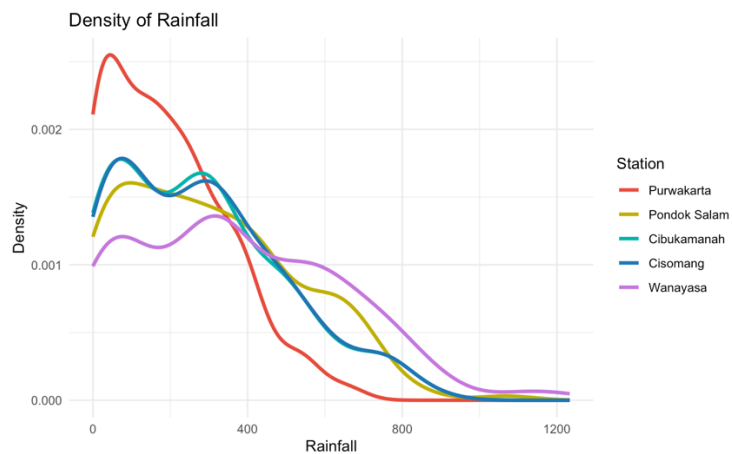


Figure 2. Rainfall density distribution at five stations in Purwakarta Regency.

The observed distribution patterns closely resemble the statistical properties of the Tweedie Compound Poisson-Gamma (TCPG) distribution, a mixed distribution that effectively accommodates data characterized by both a high frequency of zero events and extreme positive values. This distribution is particularly well-suited for modeling rainfall data, as it captures two underlying processes: a Poisson process representing the frequency of rainfall events, and a Gamma distribution modeling the intensity of rainfall when it occurs. Such a dual mechanism reflects the inherently sporadic nature of tropical rainfall, which is not only irregular but can also exhibit extreme intensities. Previous studies have confirmed the suitability of the TCPG distribution in hydrological and meteorological contexts, especially for modeling daily and monthly precipitation in tropical regions [24]. The positively skewed shapes and long right tails in the rainfall density plots for nearly all stations suggest that the rainfall data in this region are likely to follow a TCPG distribution. Statistically, this provides a more accurate modeling framework, and practically, it has important implications for the development of early warning systems, water resource management, and risk assessment for extreme weather events affecting agriculture and infrastructure.

3.2. Parameter & Hyperparameter Tuning

Based on the estimated power parameter p obtained through the profile likelihood approach in the GLM-Tweedie model (as shown in Table 2), the optimal p value for most stations—namely Purwakarta, Cibukamanah, Cisomang, and Wanayasa—is 1.3. The only exception is Pondok Salam, which yielded a higher optimal value of $p = 1.7$. This parameter plays a critical role in defining the shape of the Tweedie distribution family: values of p between 1 and 2 indicate that the data follow a compound Poisson-Gamma distribution, representing a mixture of discrete rainfall occurrences (i.e., rainfall > 0) and continuous rainfall amounts. For stations with $p = 1.3$, the distribution implies a predominance of light to moderate rainfall events, with a moderate frequency of zero values. In contrast, the higher p value at Pondok Salam suggests a tendency toward heavier rainfall intensities and a distribution with a fatter right tail. This indicates that extreme rainfall events are more frequent at Pondok Salam, possibly due to localized orographic effects or microclimatic conditions specific to that station.

Table 2. Estimated Tweedie Power Parameter p for Each Rainfall Station using Profile Likelihood Approach

Station	p
Purwakarta	1.3
Pondok Salam	1.7
Cibukamanah	1.3
Cisomang	1.3
Wanayasa	1.3

The observed variation indicates that, statistically, Pondok Salam exhibits a more variable and more extreme rainfall pattern than other locations. This aligns with previous findings from the boxplots and density graphs, which showed that Pondok Salam had a wider spread and a higher likelihood of intense rainfall events. A higher p value suggests that the Tweedie model captures an increased contribution from the Gamma component, reflecting greater rainfall intensity when precipitation occurs. In contrast, other stations with $p = 1.3$ more accurately represent light to moderate rainfall patterns that occur more frequently, with less extreme intensities. These findings reinforce the relevance of using the TCPG distribution for rainfall modeling in tropical regions and demonstrate that the GLM-Tweedie model can be effectively adapted to local rainfall characteristics by accurately estimating the power parameter p [25].

Table 3 presents the results of the hyperparameter tuning process for the TDboost model, which was used to model monthly rainfall across various tested hyperparameter combinations. The tuned parameters include: α_{trv} , representing the Tweedie power index p , with values tested between 1.3 and 1.7; depth_{trv} , the maximum depth of decision trees (either 2 or 3); eta_{trv} , the learning rate, set at 0.005 or 0.003; and ntrees , the maximum number of trees, fixed at 6500 for all model runs. These combinations were optimized to

minimize the prediction error on the validation set and capture the nonlinear and zero-inflated nature of tropical rainfall patterns.

Table 3. Optimized Hyperparameters of the TDboost Model for Each Rainfall Station

Station	p	Depth	Best Iteration
Purwakarta	1.7	3	5732
Pondok Salam	1.6	2	5904
Cibukamanah	1.6	2	6500
Cisomang	1.7	2	6495
Wanayasa	1.7	2	6054

Based on the hyperparameter tuning results, the optimal values of the Tweedie index parameter p varied across stations, reflecting the distinct distributional characteristics of rainfall at each location. Stations such as Purwakarta, Cisomang, and Wanayasa exhibited an optimal p value of 1.7, indicating greater rainfall variability and a greater frequency of high-intensity precipitation events. In contrast, Pondok Salam and Cibukamanah showed optimal p values of 1.6, which still fall within the compound Poisson-Gamma range but with slightly more moderate rainfall intensity contributions. Additionally, the optimal tree depth for the TDboost model was predominantly 2, except for Purwakarta, where a depth of 3 was required. This suggests that the relationship between the predictors and rainfall in Purwakarta is more complex and requires a richer model structure for accurate learning.

The difference in the selected Tweedie power index between GLM-Tweedie and TDboost should be interpreted in relation to the different model structures and optimization procedures. In GLM-Tweedie, the power index is selected within a parametric log-link regression framework, where the mean–variance relationship is evaluated under a linear predictor structure. In contrast, in TDboost, the Tweedie power index is treated as part of the loss-function tuning process and is selected together with other hyperparameters, such as tree depth, learning rate, and the number of boosting iterations, based on validation performance. Therefore, the optimal power index in TDboost does not necessarily have the same interpretation as the power index selected in GLM-Tweedie. A higher selected value in TDboost, such as $p = 1.7$ at Purwakarta, may indicate that the nonlinear boosting model gives greater weight to positive rainfall intensity and heavier-tailed variability after accounting for nonlinear predictor effects. Meanwhile, the lower value selected by GLM-Tweedie may reflect the best mean–variance compromise under a more restrictive linear structure. Thus, the discrepancy suggests that the two models capture different aspects of the rainfall-generating process, driven by their distinct optimization objectives and functional flexibility.

Regarding the optimal number of boosting iterations, models at stations such as Cibukamanah and Cisomang nearly reached the maximum tree limit (6500), indicating that a full boosting process was necessary to achieve optimal accuracy. This may imply high data variability or noise, requiring extensive learning to capture stable patterns. On the other hand, stations like Purwakarta and Pondok Salam reached convergence earlier (around 5700–5900 trees), suggesting that the model learned the predictive patterns at these locations more efficiently. Overall, these findings underscore the importance of fine-tuning the Tweedie distribution parameters to capture the unique rainfall dynamics at each location. Furthermore, they demonstrate the flexibility of the TDboost approach in accommodating spatial heterogeneity through precise parameter adjustment.

The relatively high number of boosting iterations at Cibukamanah and Cisomang indicates that the TDboost model required a longer learning process to capture the rainfall variability at these stations. In this study, the optimal number of trees was selected using an internal validation set within the training data, rather than solely on training error, which helps reduce the risk of overfitting. Nevertheless, the fact that the optimal iteration nearly reached the maximum tree limit suggests that potential overfitting cannot be fully ruled out, especially given the relatively short 12-month testing period. Therefore, the superior performance of TDboost at these stations should be interpreted with caution. Future studies should apply more rigorous validation schemes, such as leave-one-year-out cross-validation, k -fold cross-validation, or blocked temporal cross-validation, to further assess the model's robustness and generalization.

3.3. Performance Comparison

Predictive performance evaluation was conducted to compare the effectiveness of the GLM-Tweedie and TDboost models in modeling monthly rainfall, using two key metrics: Root Mean Square Error of Prediction (RMSEP) and Pearson correlation coefficient, calculated for the testing dataset. RMSEP quantifies the magnitude of prediction error, while the correlation coefficient indicates the model's ability to reproduce the actual rainfall pattern. Together, these metrics provide a comprehensive assessment of model accuracy and consistency relative to observed data.

To provide a more transparent numerical comparison, the complete performance results for all station–ESM–model combinations are summarized in Table 4. The table reports RMSEP and Pearson correlation coefficient values for GLM-Tweedie and TDboost across the five rainfall stations and three ESM scenarios. This table complements the graphical summaries in Figures 3 and 4 by providing exact numerical values for verification and comparison.

Table 4. Performance comparison of GLM-Tweedie and TDboost across rainfall stations and ESM scenarios

Station	ESM Scenario	GLM-Tweedie RMSEP	TDboost RMSEP	GLM-Tweedie r	TDboost r
Purwakarta	BNU-ESM	174.53	150.91	0.62	0.78
Pondok Salam	BNU-ESM	191.44	183.27	0.59	0.72
Cibukamanah	BNU-ESM	119.57	102.82	0.75	0.88
Cisomang	BNU-ESM	134.45	128.37	0.66	0.68
Wanayasa	BNU-ESM	192.31	200.98	0.68	0.72
Purwakarta	MIROC-ESM	200.84	187.26	0.56	0.65
Pondok Salam	MIROC-ESM	257.49	194.22	0.29	0.65
Cibukamanah	MIROC-ESM	180.04	140.68	0.51	0.73
Cisomang	MIROC-ESM	228.69	185.53	0.18	0.50
Wanayasa	MIROC-ESM	229.83	157.65	0.52	0.84
Purwakarta	MIROC-ESM-CHEM	149.84	161.62	0.81	0.71
Pondok Salam	MIROC-ESM-CHEM	169.80	166.35	0.72	0.73
Cibukamanah	MIROC-ESM-CHEM	91.74	90.64	0.87	0.87
Cisomang	MIROC-ESM-CHEM	126.01	109.76	0.69	0.78
Wanayasa	MIROC-ESM-CHEM	177.10	188.13	0.77	0.71

Table 4 shows that TDboost generally outperformed GLM-Tweedie across most station–ESM combinations. Based on RMSEP, TDboost produced lower prediction errors in 12 out of 15 combinations, indicating its stronger predictive accuracy for monthly rainfall downscaling. Under the MIROC-ESM scenario, TDboost consistently achieved lower RMSEP values and higher Pearson correlation coefficients across all five stations, suggesting that the nonlinear boosting structure was more effective at extracting predictive information from this ESM. However, under the MIROC-ESM-CHEM and BNU-ESM scenarios, several exceptions were observed. GLM-Tweedie produced lower RMSEP at Wanayasa under BNU-ESM and at Purwakarta and Wanayasa under MIROC-ESM-CHEM, indicating that the simpler linear Tweedie model may still perform competitively in some local settings. The lowest RMSEP was obtained at Cibukamanah using TDboost under the MIROC-ESM-CHEM scenario, while the highest correlation was achieved at the same station using TDboost under the BNU-ESM scenario. These results suggest that the best-performing ESM may vary across evaluation metrics, but overall, TDboost provides more consistent improvements in both prediction error and temporal agreement with observed rainfall.

As presented in Figure 3, the TDboost model demonstrated superior predictive performance compared to GLM-Tweedie in most cases, as indicated by consistently lower RMSEP values. For instance, at the Cibukamanah station, the TDboost model under the MIROC-ESM-CHEM scenario achieved the lowest

RMSEP of 90.63, outperforming both the GLM-Tweedie model under the same scenario (91.73) and all other scenario combinations. A similar trend was observed at Cisomang station, where TDboost (RMSEP = 109.75) significantly outperformed GLM-Tweedie (RMSEP = 126.01) under the MIROC-ESM-CHEM scenario. At Purwakarta station, TDboost also outperformed GLM-Tweedie across nearly all scenarios, with the most notable improvement seen under the MIROC-ESM scenario (187.25 vs. 200.84), indicating a consistent enhancement in predictive accuracy.

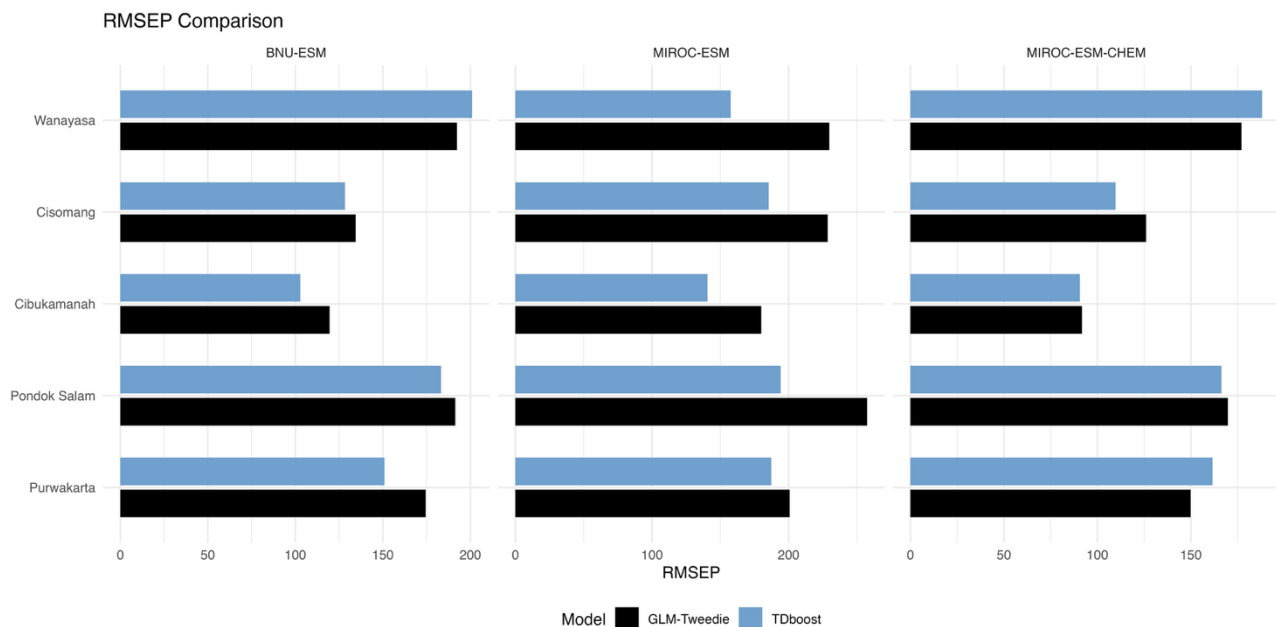


Figure 3. Comparison of RMSEP Values Between GLM-Tweedie and TDboost Models Across Stations and ESM Scenarios

Despite the TDboost model's overall superiority, several exceptions suggest that its performance may not universally surpass that of the GLM-Tweedie approach across all locations and scenarios. For instance, at the Wanayasa station under the MIROC-ESM-CHEM scenario, GLM-Tweedie achieved a lower RMSEP (177.10) than TDboost (188.13), suggesting slightly better predictive accuracy. This observation highlights the importance of considering local rainfall characteristics and the variability introduced by different ESM predictors. Nevertheless, TDboost generally outperforms GLM-Tweedie due to its ability to model nonlinear relationships and complex interactions among atmospheric variables. The boosting approach, especially when optimized via hyperparameter tuning, demonstrates significantly improved predictive performance compared to traditional GLMs for rainfall forecasting based on the Tweedie distribution.

Further evaluation of Pearson correlation coefficients between predicted and observed rainfall values during the testing period reinforces the advantage of TDboost. In almost all stations and ESM scenarios, TDboost consistently yields higher correlation values, indicating its superior ability to capture rainfall temporal patterns. At the Cibukamanah station, TDboost achieved the highest correlation of 0.88 under the BNU-ESM scenario, surpassing GLM-Tweedie's 0.75. A similar improvement was observed at the Cisomang station, where the correlation increased markedly from 0.19–0.69 (GLM-Tweedie) to 0.50–0.78 with TDboost. These findings underscore TDboost's effectiveness in detecting complex and nonlinear rainfall dynamics that are often underrepresented in classical linear models. The results align with previous studies demonstrating that gradient boosting significantly outperforms linear models in capturing intricate atmospheric interactions and nonlinear dependencies [26, 27].

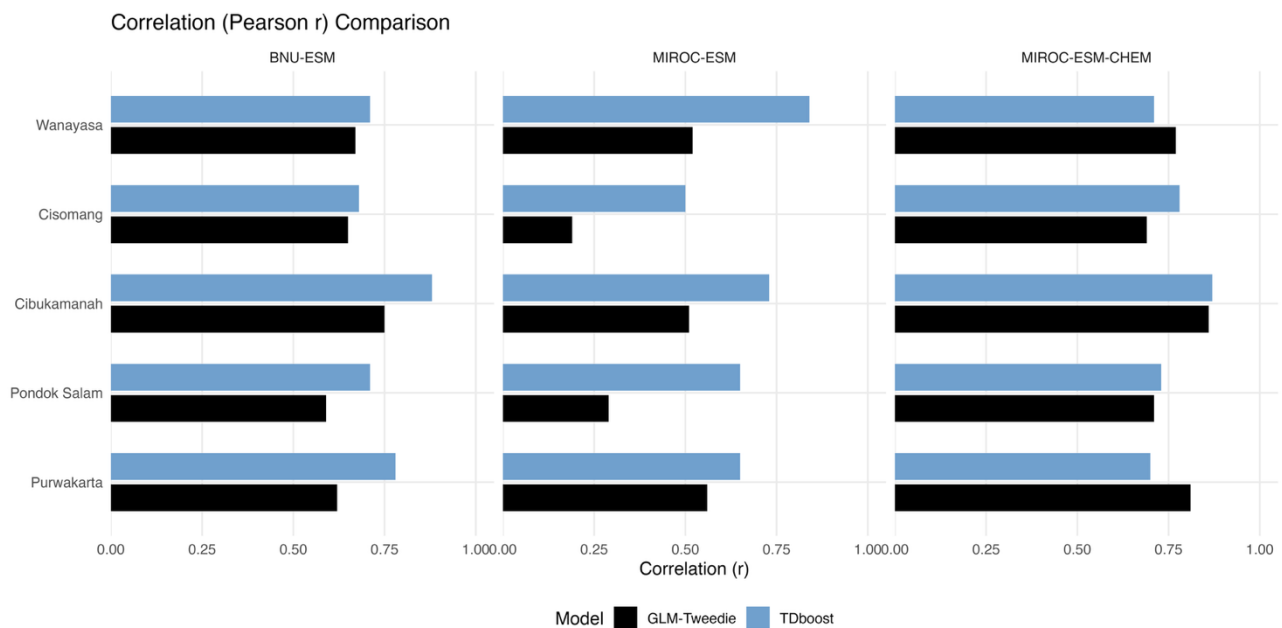


Figure 4. Correlation Coefficients Between Observed and Predicted Monthly Rainfall Across Stations and ESM Scenarios for GLM-Tweedie and TDboost Models

In the comparison of ESM inputs, BNU-ESM consistently delivered the highest correlation values for TDboost across nearly all stations. For example, at Cibukamanah and Purwakarta, BNU-ESM achieved the highest correlations of 0.88 and 0.78, respectively, when used with TDboost, outperforming the MIROC-ESM and MIROC-ESM-CHEM scenarios. This indicates that rainfall projections from BNU-ESM better align with local climatic characteristics in the study area, thereby providing more informative inputs for nonlinear models such as TDboost. Conversely, GLM-Tweedie appeared more sensitive to the choice of ESM, with its best performance often found under the MIROC-ESM-CHEM scenario, as observed in Purwakarta (correlation = 0.81) and Wanayasa (0.77). This can be attributed to the linear structure of GLM-Tweedie, which tends to rely on stable and informative predictors for accurate modeling ^[14].

In this study, ESM outputs were used without applying additional bias correction before statistical downscaling. This design choice was made to isolate the comparative performance of GLM-Tweedie and TDboost under the same raw ESM input conditions. However, this choice also represents an important limitation. ESM outputs may contain systematic biases, particularly for precipitation-related variables in tropical regions with complex terrain. Therefore, the absence of bias correction may have influenced the magnitude of prediction errors and the relative performance of the two models. The greater flexibility of TDboost may allow the model to capture nonlinear relationships between ESM predictors and local rainfall more effectively, but it may also make the model more sensitive to biased or noisy predictor patterns. In contrast, GLM-Tweedie has a more parsimonious structure but may be less able to adjust for nonlinear bias patterns in the ESM inputs. Future studies should compare raw and bias-corrected ESM predictors to determine whether bias correction further improves downscaling performance and model robustness.

Overall, these findings confirm that TDboost not only achieves lower prediction errors (RMSEP) but also shows greater consistency in representing rainfall patterns, as evidenced by stronger correlations with observed data. The choice of ESM plays a crucial role; BNU-ESM performed better when combined with TDboost, whereas MIROC-ESM-CHEM performed more compatibly with GLM-Tweedie. These results emphasize the importance of selecting both appropriate climate scenarios and statistical modeling approaches to achieve reliable rainfall predictions—an essential aspect for supporting climate adaptation planning and managing hydrometeorological risks in tropical regions. Following the strong performance of TDboost in terms of both predictive accuracy (RMSEP) and correlation strength, a time-series analysis was conducted to assess how well each model captures the temporal dynamics of observed

rainfall. The visual comparison between predicted and observed rainfall provides a more intuitive understanding of the models' consistency in tracking seasonal fluctuations and extreme events across stations and climate scenarios.

The stronger performance of TDboost can be explained from both statistical and climatological perspectives. Statistically, TDboost is more flexible than GLM-Tweedie because it combines the Tweedie loss function with tree-based boosting, allowing the model to capture nonlinear predictor–response relationships and interaction effects that are difficult to represent in a linear GLM framework. This flexibility is particularly relevant for tropical rainfall, where complex interactions among large-scale climate predictors, seasonal transitions, and local topographic effects influence rainfall occurrence and intensity. From a climatological perspective, TDboost's ability to better track rainfall peaks suggests that nonlinear models may be more suitable for representing highly variable rainfall regimes in complex tropical regions. This has practical implications for extreme rainfall monitoring and climate adaptation planning, because improved representation of high-rainfall periods can support more reliable hydrometeorological risk assessment, water resource planning, and local adaptation strategies. However, given the limited testing period, these implications should be interpreted cautiously and further validated using longer records and uncertainty-aware evaluation.

Based on the time-series visualization under the MIROC-ESM scenario (Figure 5), TDboost consistently demonstrated a superior ability to capture both seasonal and extreme rainfall patterns compared to GLM-Tweedie. At almost all stations, the orange line representing TDboost closely tracked observed rainfall trends, particularly during the testing period. In contrast, the blue line representing GLM-Tweedie often produced smoother, less responsive predictions to annual variability. This result confirms TDboost's greater ability to capture nonlinear patterns and better represent actual rainfall dynamics under the MIROC-ESM climate scenario.

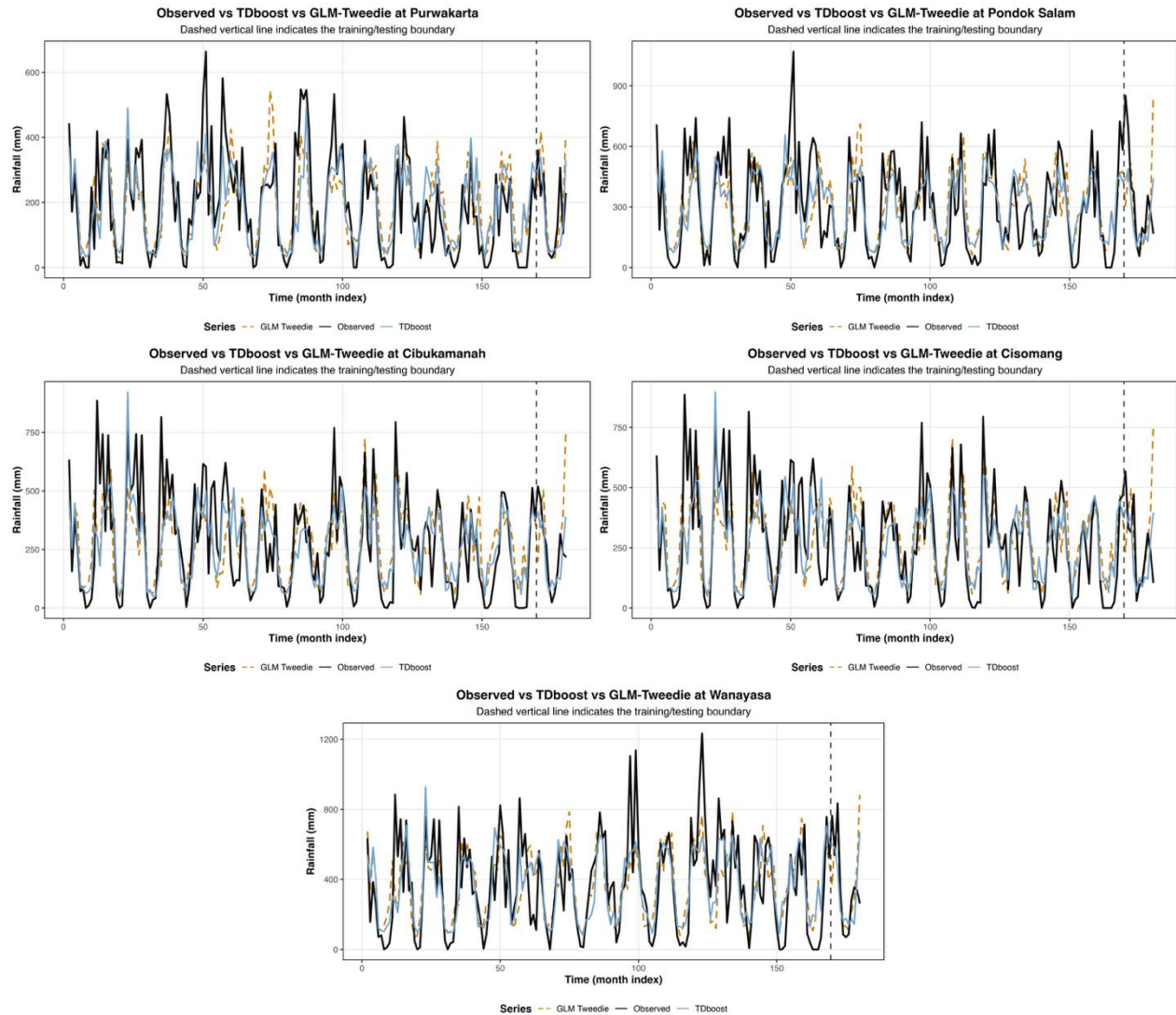


Figure 5. Comparison of Observed Rainfall and Model Predictions (TDboost vs GLM-Tweedie) under MIROC-ESM Scenario

Under the MIROC-ESM-CHEM scenario (Figure 6), the differences between the two models become even more pronounced. TDboost once again demonstrates a greater adaptability to rainfall patterns, with its prediction curves more closely following the observed data. Stations such as Wanayasa and Cisomang exhibit better alignment of TDboost predictions compared to GLM-Tweedie, particularly in capturing extreme rainfall events. This further supports previous findings that the boosting model, when optimally tuned, is well-suited for modeling complex, skewed rainfall distributions produced by the MIROC-ESM-CHEM scenario.

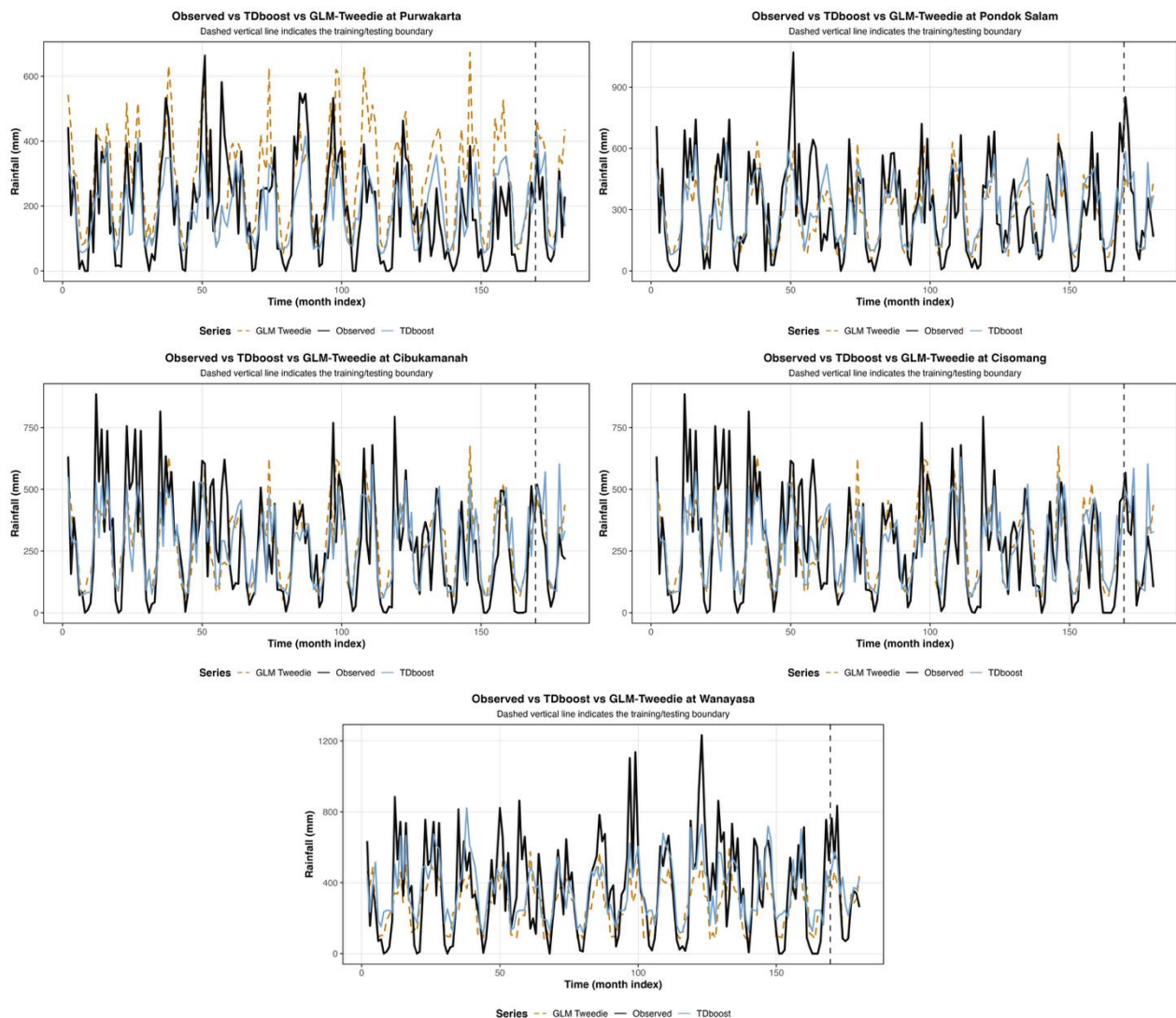


Figure 6. Comparison of Observed Rainfall and Model Predictions (TDboost vs GLM-Tweedie) under MIROC-ESM-CHEM Scenario

A similar pattern was also observed under the BNU-ESM scenario (Figure 7), where the TDboost model consistently maintained superior predictive performance compared with GLM-Tweedie. At several stations, such as Cibukamanah and Pondok Salam, TDboost more accurately replicated seasonal rainfall peaks, whereas GLM-Tweedie tended to smooth them and produce more conservative predictions. This advantage of TDboost highlights its flexibility in adapting to the spatial and temporal variability of rainfall data, even despite the uncertainty associated with the BNU-ESM input.

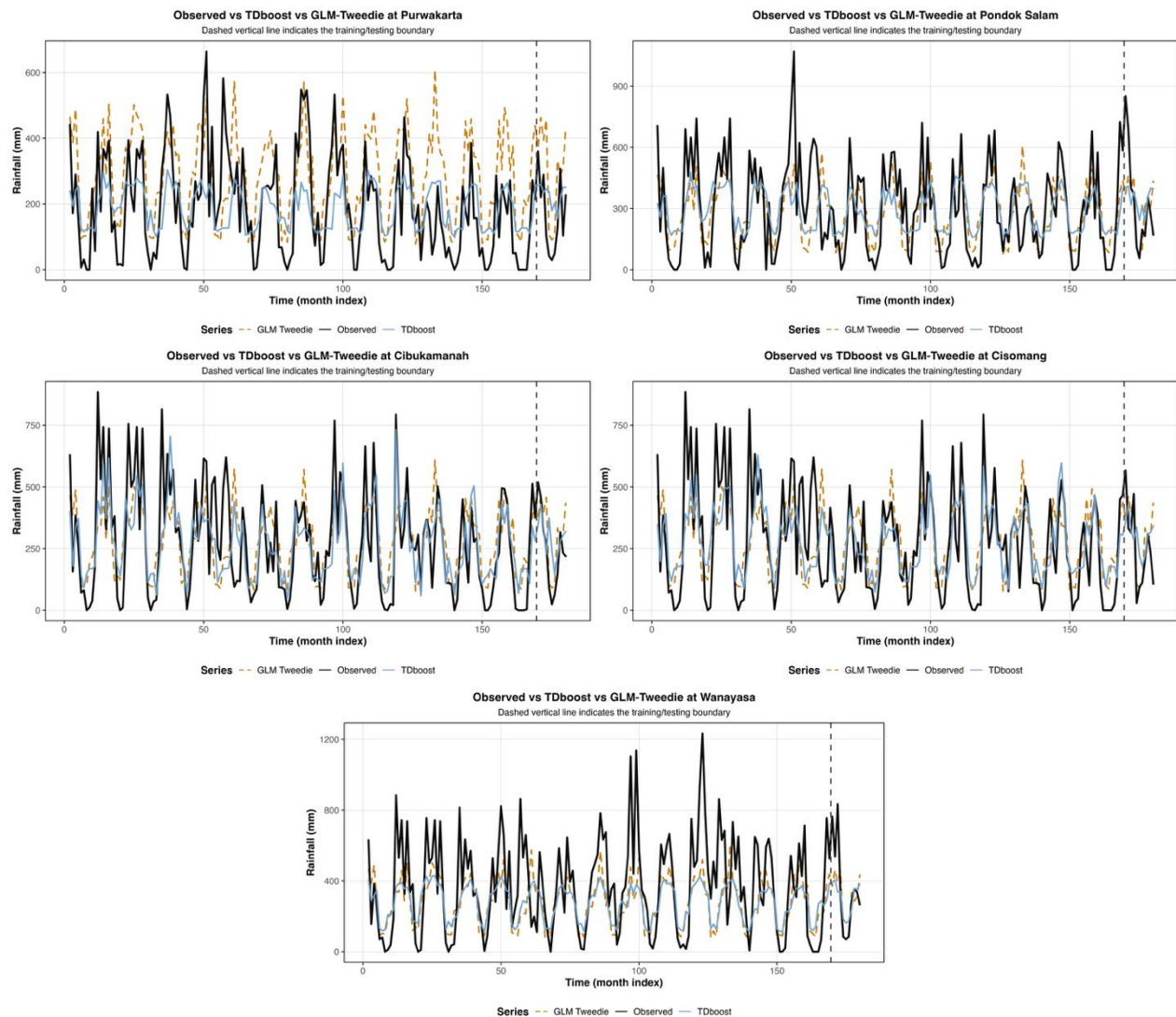


Figure 7. Comparison of Observed Rainfall and Model Predictions (TDboost vs GLM-Tweedie) under BNU-ESM Scenario

Overall, the visualization results across the three climate scenarios demonstrate that the TDboost model clearly outperforms the GLM-Tweedie model in approximating actual rainfall patterns. This superiority is evident not only in quantitative metrics such as RMSEP and correlation, but also qualitatively in the observed time-series patterns. Thus, the TDboost approach has proven more adaptive and accurate for modeling rainfall based on global climate projections, particularly for medium- to long-term forecasting in tropical regions.

This study has several limitations. First, the testing period covers only 12 months, from January to December 2020. Although this period represents one complete annual rainfall cycle and captures the broad wet-dry seasonal pattern in Indonesia, it may not fully reflect interannual rainfall variability. Therefore, the reported performance should be interpreted as an initial assessment of predictive ability rather than definitive evidence of long-term model generalization. Future studies should consider longer testing periods or more rigorous validation schemes, such as leave-one-year-out or blocked temporal cross-validation. Although no specific evidence of rainfall data disruption was identified for 2020, observational consistency during the COVID-19 period should also be examined carefully. Second, this study does not yet include formal uncertainty quantification. RMSEP and correlation summarize predictive accuracy and temporal agreement, but they do not provide uncertainty intervals related to model structure or ESM inputs. Future work should

incorporate bootstrap resampling, ensemble-based evaluation, or prediction intervals to better assess the reliability of downscaled rainfall estimates for climate risk applications.

4. Conclusions

This study demonstrates the effectiveness of applying Tweedie distribution-based models for statistical downscaling of monthly rainfall in tropical regions, using Purwakarta Regency, Indonesia, as a case study. By comparing the performance of a Generalized Linear Model (GLM) with a Tweedie distribution to a more flexible boosting-based approach (TDboost), the findings consistently show that TDboost outperforms GLM in terms of predictive accuracy (as indicated by lower RMSEP values) and the strength of association with observational data (as shown by higher Pearson correlation coefficients). These advantages are evident across most locations and climate scenarios, highlighting the superior capability of boosting methods in modeling highly variable and extreme rainfall patterns. Notably, the BNU-ESM climate scenario consistently provided the most stable and robust results, especially when paired with TDboost. In stations such as Cibukamanah and Purwakarta, this combination yielded the highest correlation with observed rainfall, suggesting that BNU-ESM projections align more closely with local rainfall characteristics compared with MIROC-ESM and MIROC-ESM-CHEM. These findings underscore the critical role of selecting an appropriate climate scenario to optimize the performance of statistical downscaling models, particularly in data-driven climate adaptation planning. For future research, incorporating spatial dependencies among locations should be considered, as explicitly modeling spatial relationships or integrating spatial features may enhance predictive performance, particularly in geographically interconnected areas. Moreover, extending the temporal resolution of the study (e.g., to daily or seasonal scales) and integrating additional atmospheric predictors, such as sea surface temperature or ENSO indices, could further improve model robustness and support the development of more reliable hydrometeorological early warning systems under climate change conditions.

5. Acknowledgements

The authors would like to thank the Meteorological, Climatological, and Geophysical Agency of Indonesia (BMKG) for providing rainfall observation data used in this study.

Author Contributions: Conceptualization, F.F. and A.D.; methodology, F.F., A.D., A.M.S., and C.S.; software, F.F.; validation, A.D., A.M.S., and C.S.; formal analysis, F.F.; investigation, F.F.; resources, F.F.; data curation, F.F.; writing—original draft preparation, F.F.; writing—review and editing, A.D., A.M.S., and C.S.; visualization, F.F.; supervision, A.D., A.M.S., and C.S.; project administration, F.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] Scaife, A. A.; Ferranti, L.; Alves, O.; Athanasiadis, P.; Baehr, J.; Dequé, M.; Dippe, T.; Dunstone, N.; Fereday, D.; Gudgel, R. G.; et al. Tropical Rainfall Predictions from Multiple Seasonal Forecast Systems. *Int. J. Climatol.* **2019**, *39*(2), 974–988. <https://doi.org/10.1002/joc.5855>
- [2] Hendrawan, I. G.; Asai, K.; Triwahyuni, A.; Lestari, D. V. The Interannual Rainfall Variability in Indonesia Corresponding to El Niño Southern Oscillation and Indian Ocean Dipole. *Acta Oceanol. Sin.* **2019**, *38*(7), 57–66. <https://doi.org/10.1007/s13131-019-1457-1>
- [3] Fauzi, F.; Kuswanto, H.; Atok, R. M.; Salamah, M. Statistical Downscaling and Bias Correction of the Earth System Models Outputs for Future Climate Projection under Solar Geoengineering: Case Study Indonesia. *J. Mech. Eng.* **2021**, *10*(1), 33–49.
- [4] Shabu, S. L. J.; Refonaa, J.; Devi, D.; Aishwarya, D.; Babu, K. K.; Reddy, K. P. Rainfall Prediction Using Machine Learning Techniques. *AIP Conf. Proc.* **2024**, *2850*(1), 030004. <https://doi.org/10.1063/5.0208435>

- [5] Fauzi, F.; Kuswanto, H.; Atok, R. M. Bias Correction and Statistical Downscaling of Earth System Models Using Quantile Delta Mapping (QDM) and Bias Correction Constructed Analogues with Quantile Mapping Reordering (BCCAQ). In *J. Phys.: Conf. Ser.*; 2020; 1538, p 012050. <https://doi.org/10.1088/1742-6596/1538/1/012050>
- [6] Soleh, A. M.; Wigena, A. H.; Djuraidah, A.; Saefuddin, A. Statistical Downscaling to Predict Monthly Rainfall Using Linear Regression with L1 Regularization (LASSO). *Appl. Math. Sci.* **2015**, 9(105–108), 5361–5369. <https://doi.org/10.12988/ams.2015.56434>
- [7] Rachmawati, R. N.; Sungkawa, I.; Rahayu, A. Extreme Rainfall Prediction Using Bayesian Quantile Regression in Statistical Downscaling Modeling. *Procedia Comput. Sci.* **2019**, 157, 406–413. <https://doi.org/10.1016/j.procs.2019.08.232>
- [8] Sadeghfam, S.; Khatibi, R.; Moradian, T. Statistical Downscaling of Precipitation Using Inclusive Multiple Modelling (IMM) at Two Levels. *J. Water Clim. Change* **2021**, 12(7), 3373–3387.
- [9] Zhang, X.; Shen, M.; Chen, J.; Homan, J. W.; Busted, P. R. Evaluation of Statistical Downscaling Methods for Simulating Daily Precipitation Distribution, Frequency, and Temporal Sequence. *Trans. ASABE* **2021**, 64(3), 771–784. <https://doi.org/10.13031/trans.14097>
- [10] Kuswanto, H.; Kravitz, B.; Miftahurrohman, B.; Fauzi, F.; Sopahaluwakan, A.; Moore, J. Impact of Solar Geoengineering on Temperatures over the Indonesian Maritime Continent. *Int. J. Climatol.* **2021**, 41, 1–20. <https://doi.org/10.1002/joc.7391>
- [11] Dewanti, D.; Djuraidah, A.; Sartono, B.; Sopaheluwakan, A. Bias Correction and Ensemble Techniques in Statistical Downscaling Model for Rainfall Prediction Using Tweedie-LASSO in West Java, Indonesia. *J. Agrometeorol.* **2024**, 26(3), 324–330. <https://doi.org/10.54386/jam.v26i3.2614>
- [12] Adam, F. F.; Kurnia, A.; Purnaba, I. G.; Mangku, I. W.; Soleh, A. M. Tweedie Compound Poisson Model with First Order Autoregressive Time Random Effect. *Commun. Math. Biol. Neurosci.* **2020**, 2020, 1–16. <https://doi.org/10.28919/cmbn/4960>
- [13] Dunn, P. K. Occurrence and Quantity of Precipitation Can Be Modelled Simultaneously. *Int. J. Climatol.* **2004**, 24(10), 1231–1239. <https://doi.org/10.1002/joc.1063>
- [14] Hayati, M.; Permatasari, R. Comparison of Generalized Linear Model between Gamma and Tweedie Compound Response for Rainfall Prediction in Lampung Province. *Asian J. Probab. Stat.* **2024**, 26(1), 41–49. <https://doi.org/10.9734/ajpas/2024/v26i1583>
- [15] Hayati, M.; Wigena, A. H.; Djuraidah, A.; Kurnia, A. A New Approach to Statistical Downscaling Using Tweedie Compound Poisson Gamma Response and Lasso Regularization. *Commun. Math. Biol. Neurosci.* **2021**, 2021. <https://doi.org/10.28919/cmbn/5936>
- [16] Halder, A.; Mohammed, S.; Dey, D. K. Bayesian Variable Selection in Double Generalized Linear Tweedie Spatial Process Models. *N. Engl. J. Stat. Data Sci.* **2023**, 187–199. <https://doi.org/10.51387/23-nejsds37>
- [17] Iranzad, R.; Liu, X.; Chaovalitwongse, W. A.; Hippe, D.; Wang, S.; Han, J.; Thammasorn, P.; Duan, C.; Zeng, J.; Bowen, S. Gradient Boosted Trees for Spatial Data and Its Application to Medical Imaging Data. *IJSE Trans. Healthc. Syst. Eng.* **2022**, 12(3), 165–179. <https://doi.org/10.1080/24725579.2021.1995536>
- [18] Zhou, H.; Qian, W.; Yang, Y. Tweedie Gradient Boosting for Extremely Unbalanced Zero-Inflated Data. *Commun. Stat. Simul. Comput.* **2020**, 1–23. <https://doi.org/10.1080/03610918.2020.1772302>
- [19] Yang, Y.; Qian, W.; Zou, H. Insurance Premium Prediction via Gradient Tree-Boosted Tweedie Compound Poisson Models. *J. Bus. Econ. Stat.* **2018**, 36(3), 456–470. <https://doi.org/10.1080/07350015.2016.1200981>
- [20] Watanabe, S.; Hajima, T.; Sudo, K.; Nagashima, T.; Takemura, T.; Okajima, H.; Nozawa, T.; Kawase, H.; Abe, M.; Yokohata, T.; et al. MIROC-ESM 2010: Model Description and Basic Results of CMIP5-20c3m Experiments. *Geosci. Model Dev.* **2011**, 4(4), 845–872. <https://doi.org/10.5194/gmd-4-845-2011>
- [21] Ji, D.; Wang, L.; Feng, J.; Wu, Q.; Cheng, H.; Zhang, Q.; Yang, J.; Dong, W.; Dai, Y.; Gong, D.; et al. Description and Basic Evaluation of Beijing Normal University Earth System Model (BNU-ESM) Version 1. *Geosci. Model Dev.* **2014**, 7(5), 2039–2064. <https://doi.org/10.5194/gmd-7-2039-2014>
- [22] Nur'utami, M. N.; Hidayat, R. Influences of IOD and ENSO to Indonesian Rainfall Variability: Role of Atmosphere–Ocean Interaction in the Indo-Pacific Sector. *Procedia Environ. Sci.* **2016**, 33, 196–203. <https://doi.org/10.1016/j.proenv.2016.03.070>

- [23] Hidayat, R.; Ando, K.; Masumoto, Y.; Luo, J. J. Interannual Variability of Rainfall over Indonesia: Impacts of ENSO and IOD and Their Predictability. *IOP Conf. Ser.: Earth Environ. Sci.* **2016**, 31(1), 012043. <https://doi.org/10.1088/1755-1315/31/1/012043>
- [24] Dunn, P. K.; Smyth, G. K. *Generalized Linear Models with Examples in R*, 2nd ed.; Springer: New York, 2018. <https://doi.org/10.1007/978-1-4419-0118-7>
- [25] Dunn, P. K.; Smyth, G. K. Series Evaluation of Tweedie Exponential Dispersion Model Densities. *Stat. Comput.* **2005**, 15(4), 267–280. <https://doi.org/10.1007/s11222-005-4070-y>.
- [26] Monego, V. S.; Anochi, J. A.; de Campos Velho, H. F. South America Seasonal Precipitation Prediction by Gradient-Boosting Machine-Learning Approach. *Atmosphere* **2022**, 13(2), 243. <https://doi.org/10.3390/atmos13020243>
- [27] Anwar, M. T.; Winarno, E.; Hadikurniawati, W.; Novita, M. Rainfall Prediction Using Extreme Gradient Boosting. In *Journal of Physics: Conference Series*; IOP Publishing Ltd, 2021; 1869, p 012078. <https://doi.org/10.1088/1742-6596/1869/1/012078>
- [28] Dunn, P. K.; Smyth, G. K. Series Evaluation of Tweedie Exponential Dispersion Model Densities. *Stat. Comput.* **2005**, 15(4), 267–280. <https://doi.org/10.1007/s11222-005-4070-y>
- [29] Dunn, P. K.; Smyth, G. K. *Generalized Linear Models with Examples in R*, 2nd ed.; Springer: New York, 2018. <https://doi.org/10.1007/978-1-4419-0118-7>
- [30] Hidayat, R.; Ando, K.; Masumoto, Y.; Luo, J. J. Interannual Variability of Rainfall over Indonesia: Impacts of ENSO and IOD and Their Predictability. In *IOP Conference Series: Earth and Environmental Science*; 2016; 31, p 012043. <https://doi.org/10.1088/1755-1315/31/1/012043>
- [31] Nur'utami, M. N.; Hidayat, R. Influences of IOD and ENSO to Indonesian Rainfall Variability: Role of Atmosphere–Ocean Interaction in the Indo-Pacific Sector. *Procedia Environ. Sci.* **2016**, 33, 196–203. <https://doi.org/10.1016/j.proenv.2016.03.070>
- [32] Ji, D.; Wang, L.; Feng, J.; Wu, Q.; Cheng, H.; Zhang, Q.; Yang, J.; Dong, W.; Dai, Y.; Gong, D.; et al. Description and Basic Evaluation of Beijing Normal University Earth System Model (BNU-ESM) Version 1. *Geosci. Model Dev.* **2014**, 7(5), 2039–2064. <https://doi.org/10.5194/gmd-7-2039-2014>
- [33] Watanabe, M.; Kakuta, H. Retinoid X Receptor Antagonists. *Int. J. Mol. Sci.* **2018**, 19(8), 2354. <https://doi.org/10.3390/ijms19082354>
- [34] Yang, Y.; Qian, W.; Zou, H. Insurance Premium Prediction via Gradient Tree-Boosted Tweedie Compound Poisson Models. *J. Bus. Econ. Stat.* **2018**, 36(3), 456–470. <https://doi.org/10.1080/07350015.2016.1200981>
- [35] Zhou, H.; Qian, W.; Yang, Y. Tweedie Gradient Boosting for Extremely Unbalanced Zero-Inflated Data. *Commun. Stat. Simul. Comput.* **2020**, 1–23. <https://doi.org/10.1080/03610918.2020.1772302>
- [36] Iranzad, R.; Liu, X.; Chaovalitwongse, W. A.; Hippe, D.; Wang, S.; Han, J.; Thammasorn, P.; Duan, C.; Zeng, J.; Bowen, S. Gradient Boosted Trees for Spatial Data and Its Application to Medical Imaging Data. *IJSE Trans. Healthc. Syst. Eng.* **2022**, 12(3), 165–179. <https://doi.org/10.1080/24725579.2021.1995536>
- [37] Halder, A.; Mohammed, S.; Dey, D. K. Bayesian Variable Selection in Double Generalized Linear Tweedie Spatial Process Models. *N. Engl. J. Stat. Data Sci.* **2023**, 187–199. <https://doi.org/10.51387/23-nejsds37>
- [38] Hayati, M.; Wigena, A. H.; Djuraidah, A.; Kurnia, A. A New Approach to Statistical Downscaling Using Tweedie Compound Poisson Gamma Response and Lasso Regularization. *Commun. Math. Biol. Neurosci.* **2021**, 2021, Article ID 2021. <https://doi.org/10.28919/cmbn/5936>
- [39] Hayati, M.; Permatasari, R. Comparison of Generalized Linear Model between Gamma and Tweedie Compound Response for Rainfall Prediction in Lampung Province. *Asian J. Probab. Stat.* **2024**, 26(1), 41–49. <https://doi.org/10.9734/ajpas/2024/v26i1583>
- [40] Dunn, P. K. Occurrence and Quantity of Precipitation Can Be Modelled Simultaneously. *Int. J. Climatol.* **2004**, 24(10), 1231–1239. <https://doi.org/10.1002/joc.1063>
- [41] Adam, F. F.; Kurnia, A.; Purnaba, I. G.; Mangku, I. W.; Soleh, A. M. Tweedie Compound Poisson Model with First Order Autoregressive Time Random Effect. *Commun. Math. Biol. Neurosci.* **2020**, 2020, 1–16. <https://doi.org/10.28919/cmbn/4960>

- [42] Dewanti, D.; Djuraidah, A.; Sartono, B.; Sopaheluwakan, A. Bias Correction and Ensemble Techniques in Statistical Downscaling Model for Rainfall Prediction Using Tweedie-LASSO in West Java, Indonesia. *J. Agrometeorol.* **2024**, 26(3), 324–330. <https://doi.org/10.54386/jam.v26i3.2614>
- [43] Kuswanto, H.; Kravitz, B.; Miftahurrohman, B.; Fauzi, F.; Sopahaluwaken, A.; Moore, J. Impact of Solar Geoengineering on Temperatures over the Indonesian Maritime Continent. *Int. J. Climatol.* **2021**, 41, 1–20. <https://doi.org/10.1002/joc.7391>
- [44] Zhang, X.; Shen, M.; Chen, J.; Homan, J. W.; Busteed, P. R. Evaluation of Statistical Downscaling Methods for Simulating Daily Precipitation Distribution, Frequency, and Temporal Sequence. *Trans. ASABE* **2021**, 64(3), 771–784. <https://doi.org/10.13031/trans.14097>
- [45] Sadeghfam, S.; Khatibi, R.; Moradian, T. Statistical Downscaling of Precipitation Using Inclusive Multiple Modelling (IMM) at Two Levels. *J. Water Clim. Change* **2021**, 12(7), 3373–3387.
- [46] Rachmawati, R. N.; Sungkawa, I.; Rahayu, A. Extreme Rainfall Prediction Using Bayesian Quantile Regression in Statistical Downscaling Modeling. *Procedia Comput. Sci.* **2019**, 157, 406–413. <https://doi.org/10.1016/j.procs.2019.08.232>
- [47] Soleh, A. M.; Wigena, A. H.; Djuraidah, A.; Saefuddin, A. Statistical Downscaling to Predict Monthly Rainfall Using Linear Regression with L1 Regularization (LASSO). *Appl. Math. Sci.* **2015**, 9(105–108), 5361–5369. <https://doi.org/10.12988/ams.2015.56434>
- [48] Fauzi, F.; Kuswanto, H.; Atok, R. M. Bias Correction and Statistical Downscaling of Earth System Models Using Quantile Delta Mapping (QDM) and Bias Correction Constructed Analogues with Quantile Mapping Reordering (BCCAQ). In *Journal of Physics: Conference Series*; 2020; 1538, p 012050. <https://doi.org/10.1088/1742-6596/1538/1/012050>
- [49] Shabu, S. L. J.; Refonaa, J.; Devi, D.; Aishwarya, D.; Babu, K. K.; Reddy, K. P. Rainfall Prediction Using Machine Learning Techniques. *AIP Conf. Proc.* **2024**, 2850(1), 030004. <https://doi.org/10.1063/5.0208435>
- [50] Fauzi, F.; Kuswanto, H.; Atok, R. M.; Salamah, M. Statistical Downscaling and Bias Correction of the Earth System Models Outputs for Future Climate Projection under Solar Geoengineering: Case Study Indonesia. *J. Mech. Eng.* **2021**, 10(1), 33–49.