RESEARCH ARTICLE • OPEN ACCESS

Enhancing the Reliability of the Flood Early Warning System in Samarinda City Through the Hybrid SARIMA-RBFNN Model

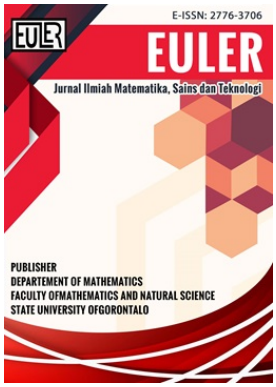
Syifa Mutia Rahmah, Andrea Tri Rian Dani, and Sri Wahyuningsih

Volume 13, Issue 3, pp. 328–337, Dec. 2025

Received 3 September 2025, Revised 5 November 2025, Accepted 9 November 2025, Published 1 December 2025

To Cite this Article : S. M. Rahmah, A. T. R. Dani, and S. Wahyuningsih,"Enhancing the Reliability of the Flood Early Warning System in Samarinda City Through the Hybrid SARIMA-RBFNN Model", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 13, no. 3, pp. 328–337, 2025, *https://doi.org/10.37905/euler.v13i3.34268*

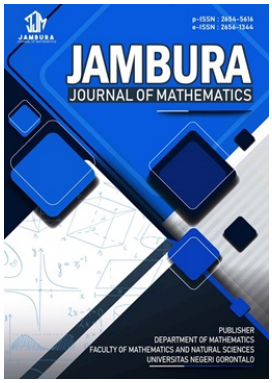
JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI



Homepage	:	http://ejurnal.ung.ac.id/index.php/euler/index
Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
Frequency	:	Three times a year
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/euler
Online ISSN	:	2776-3706
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/euler/oai
Google Scholar ID	:	QF_r-_gAAAAJ
Email	:	euler@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

Jambura Journal of Biomathematics



Jambura Journal of Mathematics

Jambura Journal of Mathematics Education

Jambura Journal of Probability and Statistics

Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi, Volume 13, Issue 3, Pages 328–337, Dec. 2025 *https://doi.org/10.37905/euler.v13i3.34268*

Research Article

Check for updates

Enhancing the Reliability of the Flood Early Warning System in Samarinda City Through the Hybrid SARIMA-RBFNN Model

Syifa Mutia Rahmah[1], **Andrea Tri Rian Dani**[1,*], **and Sri Wahyuningsih**[1]

[1]*Statistics Study Program, Universitas Mulawarman, Samarinda 75119, Indonesia*

ARTICLE HISTORY
Received 3 September 2025
Revised 5 November 2025
Accepted 9 November 2025
Published 1 December 2025

KEYWORDS
Forecast
Hybrid
Rainfall
RBFNN
SARIMA

ABSTRACT. *Rainfall data in Samarinda City exhibit seasonal patterns that play a crucial role in increasing flood risk during certain periods. To enhance the effectiveness of the early warning system, this study developed a hybrid SARIMA-RBFNN model. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model was applied to capture linear seasonal patterns, while the Radial Basis Function Neural Network (RBFNN) was employed to model nonlinear residuals from SARIMA. Model performance was assessed using the Symmetric Mean Absolute Percentage Error (SMAPE) and Root Mean Squared Error Prediction (RMSEP). Compared to the single SARIMA model (SMAPE = 34.699%, RMSEP = 82.255), the hybrid SARIMA–RBFNN achieved lower in-sample errors (SMAPE = 34.175%, RMSEP = 78.577) and demonstrated more stable performance for out-of-sample data. This indicates that the hybrid model provides a more balanced and reliable prediction by capturing nonlinear rainfall fluctuations that SARIMA alone could not model effectively. Forecasts for 2024 revealed a consistent seasonal trend, peaking mid-year. These findings indicate that the hybrid model can improve the reliability of the flood early warning system in Samarinda by providing more accurate rainfall predictions.*

1. Introduction

Forecasting is a crucial technique for predicting future values of a variable based on historical data analysis [1]. One of the most widely used methods for time series forecasting is the Autoregressive Integrated Moving Average (ARIMA) and its seasonal extension SARIMA, which are effective for handling linear and seasonal patterns [2]. However, ARIMA-based models struggle with capturing nonlinear relationships, which significantly affects forecasting accuracy, particularly in rainfall prediction [3]. To overcome this limitation, hybrid approaches have been increasingly adopted in time series forecasting. One of the most effective hybrid models is ARIMA-Neural Network (NN), which combines the linear modeling capability of ARIMA with the nonlinear pattern recognition ability of NN [4]. In this approach, ARIMA captures linear trends and seasonal components, while NN models the residuals, enhancing overall prediction accuracy and reducing errors.

This hybrid approach has demonstrated superior performance in various applications. Silfiani et al. [5] compared the hybrid Time Series Regression (TSR)-Radial Basis Function Neural Network (RBFNN) and SARIMA-RBFNN models for forecasting household electricity consumption in Jember, Indonesia, and found that the SARIMA–RBFNN model achieved higher forecasting accuracy. Similarly, Nguyen & Pham [6] empirically evaluated a hybrid ARIMA–RBFNN model in an additive framework and demonstrated that the hybrid model consistently outperformed standalone models in time series forecasting. These findings indicate that hybrid models have strong potential for improving forecasting accuracy across different domains, including meteorology.

Among various neural network architectures, the RBFNN stands out as particularly effective for time series forecasting due to its ability to model nonlinear relationships using radial basis activation functions. RBFNN operates through two learning stages: an unsupervised phase, where the hidden layer clusters input data, and a supervised phase, where the output layer adjusts weights to minimize prediction errors. Compared with traditional feedforward neural networks, RBFNN offers faster training and better adaptability to nonlinear data structures, making it highly suitable for forecasting applications [7].

Although hybrid ARIMA–NN models have shown promising results, previous studies on rainfall forecasting remain limited in several ways. Most applications have focused on other regions outside Kalimantan, and few have been tested using localized datasets. Furthermore, many existing studies employ standard feedforward or backpropagation neural networks rather than RBFNN, which offers stronger nonlinear modeling capabilities. Recent studies in Indonesia have also adopted advanced deep learning methods such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) for rainfall prediction, showing high accuracy in modeling temporal dependencies in meteorological data. For instance, Suranata [8] successfully applied both LSTM and GRU using BMKG datasets in Denpasar and found that LSTM achieved lower RMSE and MAE values compared to GRU, demonstrating its effectiveness for rainfall forecasting in tropical

*Corresponding Author.

Email : *andreatririandani@fmipa.unmul.ac.id* (A. T. R. Dani)
Homepage : http://ejurnal.ung.ac.id/index.php/euler/index / E-ISSN : 2776-3706

regions. These results reinforce the growing relevance of deep learning–based approaches in improving local-scale rainfall prediction, complementing hybrid statistical–neural models such as SARIMA–RBFNN. These results reinforce the growing relevance of deep learning–based approaches in improving local-scale rainfall prediction, complementing hybrid statistical–neural models such as SARIMA–RBFNN.

Samarinda is highly vulnerable to flooding due to significant variability in rainfall, influenced by topography, oceanic conditions, and climate change. Accurate rainfall forecasting is essential for disaster mitigation and effective decision-making [9]. However, modeling rainfall behavior in this region remains challenging due to its complex climatic patterns. Previous studies on rainfall forecasting in Indonesia, particularly in Kalimantan, have mostly applied conventional statistical models such as ARIMA or linear regression which often fail to capture the nonlinear rainfall characteristics typical of tropical regions. Ihsan et al. [4] found that ARIMA-based models in Makassar had limited ability to identify sudden rainfall fluctuations caused by local climatic variability. In East Kalimantan, research on rainfall prediction remains limited, with most studies focusing on descriptive climatology rather than predictive modeling.

This study addresses these gaps by applying and optimizing a hybrid SARIMA–RBFNN model for rainfall forecasting in Samarinda City, a flood-prone region in East Kalimantan. Unlike previous studies that primarily used conventional statistical or generic neural models, this research integrates SARIMA with RBFNN to simultaneously capture linear seasonal structures and nonlinear rainfall variations. The novelty of this study lies in both its methodological contribution enhancing rainfall prediction through hybridization and its regional focus on Samarinda, where such approaches have not been comprehensively explored. Ultimately, the proposed model aims to improve the reliability of Samarinda's flood early warning system by providing more accurate and locally adaptive rainfall forecasts to support disaster mitigation and preparedness efforts.

2. Methods

2.1. Periodogram Analysis

The periodogram is a function that represents the distribution of data variation based on frequency. Identifying repeating patterns or periodicity involves analyzing frequencies indicated by peak points in the periodogram graph [10]. According to Wei [11], the periodogram values can be expressed as eq. (1).

$$I(\omega_i) = \begin{cases} na_0^2, & i = 0, \\ \dfrac{n(a_i^2 + b_i^2)}{2}, & i = 1, 2, \ldots, \left[\dfrac{n-1}{2}\right], \text{ if } n \text{ is odd}, \\ \dfrac{n(a_i^2 + b_i^2)}{2}, & i = 1, 2, \ldots, \left[\dfrac{n}{2} - 1\right], \text{ if } n \text{ is even}, \\ na_{n/2}^2, & i = \dfrac{n}{2}, \end{cases} \tag{1}$$

where a_i and b_i are the Fourier coefficient obtained through eq. (2) and eq. (3).

$$a_i = \begin{cases} \dfrac{1}{n}\left[\displaystyle\sum_{t=1}^{n} Z_t \cos(\omega_i t)\right], & i = 0 \text{ and } i = \dfrac{n}{2} \text{ if } n \text{ is even}, \\ \dfrac{2}{n}\left[\displaystyle\sum_{t=1}^{n} Z_t \cos(\omega_i t)\right], & i = 1, 2, \ldots, \left[\dfrac{n-1}{2}\right] \text{ if } n \text{ is odd}, \\ \dfrac{2}{n}\left[\displaystyle\sum_{t=1}^{n} Z_t \cos(\omega_i t)\right], & i = 1, 2, \ldots, \left[\dfrac{n}{2} - 1\right] \text{ if } n \text{ is even}, \end{cases} \tag{2}$$

$$b_i = \frac{2}{n}\left[\sum_{t=1}^{n} Z_t \sin(\omega_i t)\right], \quad \begin{cases} i = 1, 2, \ldots, \left[\dfrac{n-1}{2}\right] & \text{if } n \text{ is odd}, \\ i = 1, 2, \ldots, \left[\dfrac{n}{2} - 1\right] & \text{if } n \text{ is even}, \end{cases} \tag{3}$$

$$\omega_i = \frac{2\pi i}{n}, \quad \begin{cases} i = 1, 2, \ldots, \left[\dfrac{n-1}{2}\right] & \text{if } n \text{ is odd}, \\ i = 1, 2, \ldots, \left[\dfrac{n}{2}\right] & \text{if } n \text{ is even}, \end{cases} \tag{4}$$

where,

$I(\omega_i)$	:	the periodogram value at the k-th Fourier frecuency
ω_i	:	the k-th Fourier frequency
a_i	:	parameter of the Fourier coefficient a
b_i	:	parameter of the Fourier coefficient b
π	:	3.142 or $\frac{22}{7}$.

To determine the seasonal period, eq. (5) can be used,

$$T^* = \frac{2\pi}{\omega_{(i)}}, \tag{5}$$

where $\omega_{(i)}$ represents the i-th Fourier frequency with the highest periodogram value.

2.2. Seasonal Autoregressive Integrated Moving Average

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is an extension of ARIMA designed to model time series data that exhibiting seasonal patterns. In the notation SARIMA(p, d, q) (P, D, Q)^S, lowercase letters indicate the non-seasonal components, while uppercase letters correspond to the seasonal components [12]. SARIMA is a linear model primarily applied to forecast stationary time series [13]. Variance non-stationarity can be addressed through power transformations, such as the Box–Cox transformation introduced by Box and Cox in 1964, as shown in eq. (6) [11].

$$Z_t^* = \begin{cases} \dfrac{Z_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln(Z_t), & \lambda = 0, \end{cases} \tag{6}$$

where, Z_t is time series data at period t, $t = 1, 2, \ldots, n$ and λ is transformation parameter.

Mean non-stationarity can be resolved through differencing. According to [14], this process involves subtracting the observation at time t from the observation at the previous time

$(t-1)$. The general differencing equation is given in eq. (7).

$$Z_t^{*1} = Z_t - Z_{t-1}, \tag{7}$$

where, Z_t^{*1} represents the time series data at period t after differencing of order 1. The SARIMA(p, d, q) (P, D, Q)$^{\text{S}}$ model is expressed in eq. (8) [11]:

$$\Phi_P(B^S)\,\phi_p(B)\,(1-B)^d(1-B^S)^D Z_t = \Theta_Q(B^S)\,\theta_q(B)\,\varepsilon_t, \tag{8}$$

where

S	:	seasonal period
D	:	seasonal differencing order
B	:	backshift operator
Φ_P	:	seasonal AR parameter of order P
ϕ_p	:	non-seasonal AR parameter of order p
Θ_Q	:	seasonal MA parameter of order$Q$
θ_q	:	non-seasonal MA parameter of order$q$.

The AR and MA orders are identified through the autocorrelation (ACF) and partial autocorrelation (PACF) plots [15]. After specifying the initial SARIMA model, parameters are estimated through maximum likelihood estimation (MLE). Diagnostic checks are then performed, including t-tests for parameter significance, the Kolmogorov–Smirnov test for residual normality, and the Ljung–Box test for residual independence [11].

2.3. Radial Basis Function Neural Network

The Radial Basis Function Neural Network (RBFNN) is structured as a feedforward model with three main layers: input, hidden, and output. Learning in RBFNN occurs in two stages: an unsupervised phase that maps inputs to the hidden layer, followed by a supervised phase that maps hidden layer outputs to the final output layer. Unlike conventional multilayer networks that use sigmoid activation functions, RBFNN employs radial basis functions (RBF) in the hidden layer to perform nonlinear transformations, while the output layer applies a linear combination to generate predictions [7]. RBFNN is more efficient in solving complex cases compared to standard multilayer networks [16].

The Gaussian function is the most widely used basis function in RBFNN due to its strong generalization capability and ease of implementation, as expressed in eq. (9) and eq. (10) [7].

$$\varphi_{k^*}(\mathbf{x}) = \exp\left[-\frac{r^2}{2\sigma_{k^*}^2}\right], \tag{9}$$

$$r = \|\mathbf{x} - \mathbf{c}_{k^*}\|, \tag{10}$$

where,

σ_{k^*}	:	maximum distance at the k^*-th hidden layer neuron
r	:	euclidean norm between the input vector and the centroid at the hidden layer
c_{k^*}	:	centroid at the k^*-th hidden layer neuron
$\mathbf{x}$	:	input variable vector

The Gaussian function is the most commonly used due to its compatibility with the radial basis concept, strong generalization ability, and ease of implementation. The steps in modeling an RBFNN are as follows:

1. Identify the number of neurons in the input layer.
2. Determining the number of neurons in the hidden layer using K-means clustering, where the number of neurons is based on the number of clusters formed. To determine the optimal number of clusters, the Silhouette Coefficient (SC) value is used as a measure of clustering effectiveness [17].
3. Calculate the centroids and the distances between input vectors and each neuron in the hidden layer.
4. Calculate the activation function values.
5. Determine the weights and biases using the least squares method with eq. (11):

$$\hat{\mathbf{w}} = \left(\boldsymbol{\varphi}^{\mathsf{T}}\boldsymbol{\varphi}\right)^{-1}\boldsymbol{\varphi}^{\mathsf{T}}\mathbf{x}_t. \tag{11}$$

6. Compute the output function values using Equation eq. (12):

$$Y_t = \sum_{k^*=1}^{K^*} w_{k^*}\,\varphi_{k^*}(\mathbf{x}) + w_b. \tag{12}$$

2.4. Hybrid SARIMA-RBFNN

The rapid advancement of computational technology has driven innovation in forecasting methods. One notable trend is the emergence of hybrid forecasting models, which combine two or more different forecasting techniques [18]. This hybrid approach is expected to significantly enhance forecasting accuracy compared to using a single method [19]. Generally, the combination of linear and nonlinear time series models can be expressed as follows [20]:

$$\hat{N}_t = \hat{Z}_t + \hat{Y}_t, \tag{13}$$

where, $\hat{Z}_t$ represents the forecasted values from the linear model (SARIMA), and $\hat{Y}_t$ represents the forecasted values from the nonlinear model (RBFNN).

2.5. Forecasting Accuracy

Model evaluation is conducted to assess the prediction accuracy of the model against actual data based on the residuals generated. Root Mean Squared Error Prediction (RMSEP) and Symmetric Mean Absolute Percentage Error (SMAPE) are key criteria used to identify the best model by quantifying prediction errors. The model with the lowest RMSEP and SMAPE values is considered the most suitable, as it provides more accurate and reliable predictions. RMSEP and SMAPE values can be calculated using eq. (14) and eq. (15) [21, 22].

$$\text{RMSEP} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Z_t + \hat{N}_t\right)^2}, \tag{14}$$

$$\text{SMAPE} = \left(\frac{2}{n}\sum_{t=1}^{n}\frac{\left|Z_t - \hat{N}_t\right|}{|Z_t| + \left|\hat{N}_t\right|}\right) \times 100\%. \tag{15}$$

3. Results and Discussion

3.1. Data Description

This study analyzes monthly rainfall data in Samarinda from January 2016 to December 2023, totaling 96 observations obtained from Badan Pusat Statistik (BPS). The analysis in this study was assisted using software R Studio for data processing, visualization, and model implementation. To gain deeper insights into the monthly rainfall patterns in Samarinda City, time series charts were created, as shown in Figure 1. These charts provide statistical visualization of rainfall changes month-to-month, helping to identify patterns within the data.

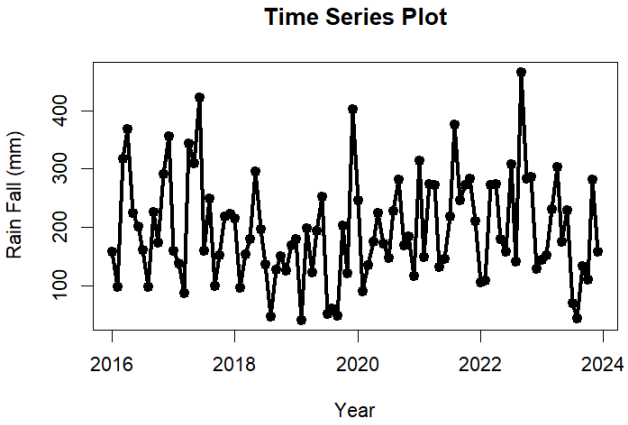


Figure 1. Time Series of Monthly Rainfall in Samarinda (January 2016–December 2023)

The time series visualization reveals pronounced seasonal fluctuations, with rainfall peaks generally occurring in March–May and October–December, and lower rainfall around mid-year. However, the magnitude and timing of these peaks vary considerably across years, showing highly unstable and irregular patterns likely influenced by large-scale climate anomalies such as El Niño and La Niña. This high level of fluctuation indicates that rainfall behavior in Samarinda is not purely seasonal but also highly dynamic, thereby justifying the use of a hybrid forecasting model such as SARIMA–RBFNN, which can capture both the seasonal structure and complex variability of the data. To ensure robust model training and evaluation, the dataset was divided into 90% in-sample data (January 2016–February 2023, 86 observations) for model construction and 10% out-of-sample data (February 2023–December 2023, 10 observations) for forecasting validation.

3.2. Periodogram

Before modeling, a periodogram analysis was perfomed to identify the seasonal components in rainfall data. The Fourier frequency, coefficients, and periodogram values are presented in Table 1. Based on Table 1, a periodogram chart is created, as shown in Figure 2.

From Figure 2, peaks can be observed in the periodogram, with the highest peak at the 13^{th} frequency, having a value of 65,449.145. The Fourier frequency for this peak is 0.950, which is then used to calculate the seasonal period using eq. (5):

$$T^* = \frac{2(3,143)}{0,950} = 6,610 \approx 6.$$

The calculation shows that the monthly rainfall in Samarinda City has a seasonal period of 6 months. This indicates that rainfall tends to increase or decrease every 6 months, suggesting a seasonal pattern in the monthly rainfall data.

3.3. SARIMA

SARIMA was applied to model the monthly rainfall data in Samarinda City from January 2016 to December 2023. The modeling process was conducted step by step to capture seasonal patterns and improve forecasting accuracy. The modeling procedure

Table 1. Periodogram analysis results

i	ω_i	a_i	b_i	$I(\omega_i)$
0	0.000	199.215	-	3,413,052.980
1	0.073	36.194	-11.619	62,136.257
2	0.146	-17.973	9.369	17,664.138
3	0.219	-4.986	-10.719	6,009.475
4	0.292	2.214	-4.663	1,145.473
5	0.365	6.976	-9.176	5,712.919
6	0.438	-5.911	-3.483	2,024.110
7	0.511	-15.714	4.242	11,391.939
8	0.584	-32.883	-2.240	46,712.190
9	0.658	3.300	11.542	6,196.463
10	0.731	-13.470	10.793	12,810.478
11	0.804	-8.060	16.035	13,849.596
12	0.877	-33.470	-18.020	62,132.459
13	0.950	-38.536	6.086	65,449.145
⋮	⋮	⋮	⋮	⋮
43	3.142	-3.883	-	1,296.386

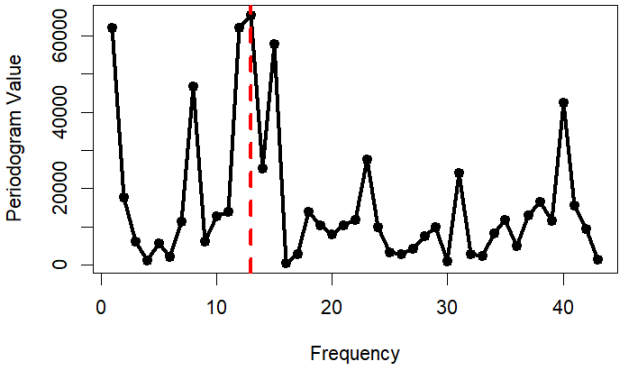
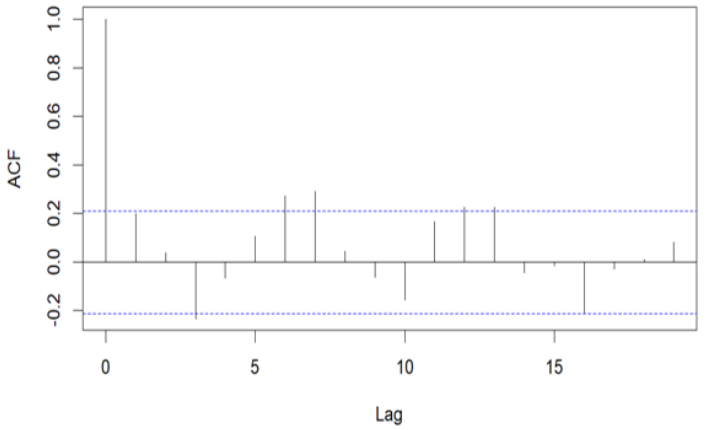


Figure 2. Periodogram chart of monthly rainfall data in samarinda city (January 2016 - February 2023)

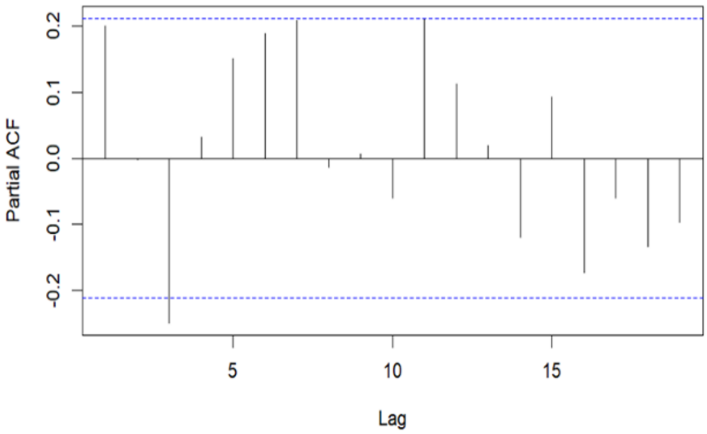
began with testing the stationarity of the data in both variance and mean. The time series plot showed clear fluctuations over time, suggesting non-stationarity. A Box–Cox transformation was then employed to stabilize variance, where the estimated λ value of 0.458 indicated that the data were not stationary in variance. After applying the power transformation ($\lambda = 0.458$), the variance stabilized, as λ approached 1.

Next, the stationarity data in mean after transformation is checked using the ADF test. Based on the ADF test, the results obtained are $p_{value} = 0.035$. This value is smaller than alpha, where $\alpha = 0.05$, which means that the data after transformation is stationary in mean. If the data is stationary in variance and average, an ACF and PACF plot is created, as in Figure 3.

Based on Figure 3a, the ACF exceeds the confidence interval limits at lags 3, 6, 7, 12, 13, and 16. According to the principle of parsimony, lag 3 is selected for the non-seasonal MA component, and lags 6 and 12 for the seasonal MA component. Therefore, the possible order for the non-seasonal MA(q) is 3, and the possible orders for seasonal MA(Q) are 1 and 2. In Figure 3b, the PACF exceeds the confidence interval at lag 3, indicating that the possible order for non-seasonal AR(p) is 3, and the possible order

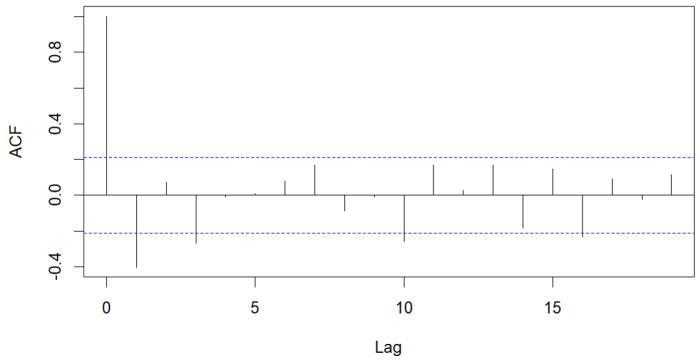


(a)

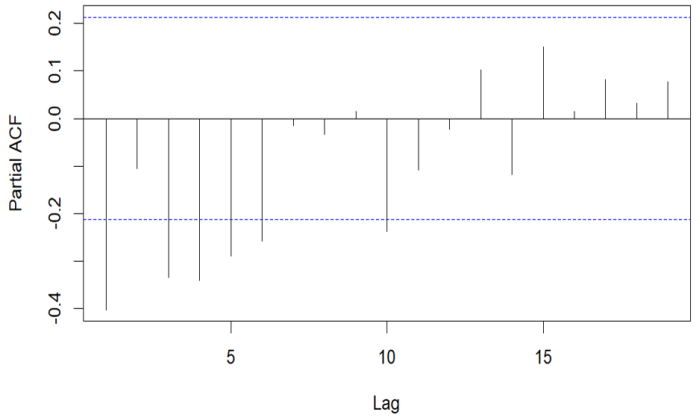


(b)

Figure 3. Plot of ACF(a) and PACF(b) after transformation



(a)



(b)

Figure 4. Plot of ACF(a) and PACF(b) after transformation and first-order differencing

for seasonal AR(P) is 0.

The next step is to perform parameter significance testing and residual diagnostics for the preliminary SARIMA model using R software. The results are presented in Table 2.

Table 2. Parameter significance testing and residual diagnostics for the preliminary SARIMA model

Model	Parameter Significance	Residual Normality	Residual Independence
SARIMA(0,0,3)(0,0,1)^6	×	✓	✓
SARIMA(3,0,3)(0,0,1)^6	×	✓	✓
SARIMA(3,0,0)(0,0,1)^6	×	✓	✓
SARIMA(0,0,3)(0,0,2)^6	×	✓	✓
SARIMA(3,0,3)(0,0,2)^6	×	✓	✓
SARIMA(3,0,0)(0,0,2)^6	×	✓	✓

Based on the results in Table 2, none of the SARIMA models exhibit significant parameters, although all models satisfy the assumptions of residual normality and independence. Therefore, a first-order differencing process was applied to the transformed data using eq. (4) to stabilize the series. The ACF and PACF plots of the rainfall data after transformation and first-order differencing are shown in Figure 4.

Based on Figure 4a, the ACF exceeds the confidence interval limits at lags 1, 3, 10, and 16. According to the principle of parsimony, only lag 1 is selected for the non-seasonal MA component, resulting in a possible order for non-seasonal MA(q) of 1, while the possible order for seasonal MA(Q) is 0. In Figure 4b, the PACF exceeds the confidence interval at lags 1, 3, 4, 5, 6, and 10. Applying the principle of parsimony, lags 1, 3, 4, and 5 are chosen for the non-seasonal AR component, while lag 6 is selected for the seasonal AR component. Thus, the possible orders for non-seasonal AR(p) are 1, 3, 4, and 5, while the possible order for seasonal AR(P) is 1. Consequently, the differencing order for the non-seasonal component (d) is 1, while for the seasonal component (D), it is 0.

The preliminary SARIMA models obtained after transformation and first-order differencing are SARIMA$(0,1,1)\,(1,0,0)^6$, SARIMA$(1,1,1)\,(1,0,0)^6$, SARIMA$(3,1,1)\,(1,0,0)^6$, SARIMA$(4,1,1)\,(1,0,0)^6$, SARIMA$(5,1,1)\,(1,0,0)^6$, SARIMA$(1,1,0)\,(1,0,0)^6$, SARIMA$(3,1,0)\,(1,0,0)^6$, SARIMA$(4,1,0)\,(1,0,0)^6$, and SARIMA$(5,1,0)\,(1,0,0)^6$. These models were subsequently subjected to parameter estimation and significance testing using R software. A summary of these estimation and significance test results is presented in Table 3.

Based on parameter significance testing in Table 3, it can be concluded that the models that have significant parameters are the SARIMA$(0,1,1)\,(1,0,0)^6$, SARIMA$(5,1,1)\,(1,0,0)^6$, and SARIMA$(5,1,0)\,(1,0,0)^6$ because each parameter in the model

Table 3. Parameter estimation and significance testing

Model	Parameter Estimation	p_{value}	Model	Parameter Estimation	p_{value}
SARIMA $(0,1,1)(1,0,0)^6$	$\theta_1 = -0.952$	0.000	SARIMA $(1,1,0)(1,0,0)^6$	$\varphi_1 = -0.452$	0.000
	$\Phi_1 = 0.297$	0.017		$\Phi_1 = 0.229$	0.047
SARIMA $(1,1,1)(1,0,0)^6$	$\varphi_1 = 0.118$	0.309	SARIMA $(3,1,0)(1,0,0)^6$	$\varphi_1 = -0.555$	0.000
	$\theta_1 = -1.000$	0.000		$\varphi_2 = -0.309$	0.013
	$\Phi_1 = 0.278$	0.021		$\varphi_3 = -0.317$	0.004
				$\Phi_1 = 0.216$	0.094
SARIMA $(3,1,1)(1,0,0)^6$	$\varphi_1 = 0.057$	0.605	SARIMA $(4,1,0)(1,0,0)^6$	$\varphi_1 = -0.616$	0.000
	$\varphi_2 = -0.044$	0.692		$\varphi_2 = -0.362$	0.002
	$\varphi_3 = -0.350$	0.007		$\varphi_3 = -0.515$	0.000
	$\theta_1 = -0.856$	0.000		$\varphi_4 = -0.345$	0.002
	$\Phi_1 = 0.088$	0.533		$\Phi_1 = 0.077$	0.567
SARIMA $(4,1,1)(1,0,0)^6$	$\varphi_1 = -0.087$	0.518	SARIMA $(5,1,0)(1,0,0)^6$	$\varphi_1 = -0.792$	0.000
	$\varphi_2 = -0.111$	0.307		$\varphi_2 = -0.621$	0.000
	$\varphi_3 = -0.434$	0.000		$\varphi_3 = -0.748$	0.000
	$\varphi_4 = -0.247$	0.074		$\varphi_4 = -0.746$	0.000
	$\theta_1 = -0.751$	0.000		$\varphi_5 = -0.560$	0.000
	$\Phi_1 = -0.041$	0.776		$\Phi_1 = -0.377$	0.002
SARIMA $(5,1,1)(1,0,0)^6$	$\varphi_1 = -0.531$	0.011			
	$\varphi_2 = -0.467$	0.004			
	$\varphi_3 = -0.659$	0.000			
	$\varphi_4 = -0.586$	0.000			
	$\varphi_5 = -0.427$	0.011			
	$\theta_1 = -0.357$	0.085			
	$\Phi_1 = -0.317$	0.026			

has a p_{value} smaller than the significance level $\alpha = 0.05$.

After testing the parameter significance, diagnostic testing was conducted to ensure that the models met the assumptions of residual independence and normality. The Ljung–Box test results, which compare p_{value} at various lags to a significance level of 0.05, indicate that SARIMA$(0,1,1)\,(1,0,0)^6$ and SARIMA$(5,1,0)\,(1,0,0)^6$ satisfy the independence assumption, as all p-values exceed 0.05, while SARIMA$(5,1,0)\,(1,0,0)^6$ does not. Meanwhile, the Kolmogorov–Smirnov test confirms that the residuals of all three models follow a normal distribution, with none of the p_{value} falling below 0.05.

Based on the diagnostic testing of the residual data, it was found that the SARIMA models, after transformation and first-order differencing, that satisfy both assumptions are SARIMA$(0,1,1)\,(1,0,0)^6$ and SARIMA$(5,1,0)\,(1,0,0)^6$. Therefore, these two models are used for prediction. Based on eq. (8) and Table 3, the SARIMA$(0,1,1)\,(1,0,0)^6$model can be expressed as follows:

$$\widehat{Z_t} = Z_{t-1} - 0.952Z_{t-7} + {}_t + 0.297_{t-1}. \tag{16}$$

The SARIMA$(5,1,0)\,(1,0,0)^6$ model can be expressed as:

$$\begin{aligned}\hat{Z}_t = {} & Z_{t-1} - 0.792Z_{t-1} + 0.792Z_{t-2} - 0.621Z_{t-2} + 0.621Z_{t-3} \\ & - 0.748Z_{t-3} + 0.748Z_{t-4} - 0.746Z_{t-4} + 0.746Z_{t-5} \\ & - 0.560Z_{t-5} + 0.560Z_{t-6} - 0.377Z_{t-6} + 0.377Z_{t-7} \\ & - 0.299Z_{t-7} + 0.299Z_{t-8} - 0.234Z_{t-8} + 0.234Z_{t-9} \\ & - 0.282Z_{t-9} + 0.282Z_{t-10} - 0.281Z_{t-10} + 0.281Z_{t-11} \\ & - 0.211Z_{t-11} + 0.211Z_{t-12} + \varepsilon_t \end{aligned} \tag{17}$$

The prediction values for the SARIMA$(0,1,1)\,(1,0,0)^6$ model are calculated using eq. (16), while for the SARIMA$(5,1,0)\,(1,0,0)^6$ model, predictions are made using eq. (17) and then transformed back to the original scale using the appropriate exponent. After obtaining the predictions from each SARIMA model, the next step is to evaluate model performance by calculating the Symmetric Mean Absolute Percentage Error (SMAPE) and Mean Squared Error Prediction (RMSEP) for in-sample data.

Table 4. SMAPE and RMSEP values of the SARIMA model for in-sample data

Evaluate Model	SARIMA$(0,1,1)\,(1,0,0)^6$	SARIMA$(5,1,0)\,(1,0,0)^6$
SMAPE	36.421%	34.699%
RMSEP	87.582	82.255

Based on the calculated SMAPE and RMSEP values shown in Table 4, SARIMA$(5,1,0)\,(1,0,0)^6$ was determined to be the best model for predicting rainfall in Samarinda due to its lower SMAPE and RMSEP values. Using SARIMA$(5,1,0)\,(1,0,0)^6$, rainfall was forecasted for 22 future periods (March 2023 to December 2024) and the results are shown in Table 5.

The forecast indicates that rainfall will continue to exhibit its usual seasonal pattern, peaking around May and gradually decreasing toward the end of the year. The out-of-sample forecast using the SARIMA$(5,1,0)\,(1,0,0)^6$ model produced a SMAPE of 45.921% and an RMSEP of 92.859. A comparison between the actual and predicted rainfall in Samarinda from January 2023 to December 2024 is presented in Figure 5.

The model effectively captured the overall trend and seasonal pattern of rainfall, though slight deviations were observed during extreme fluctuations. In general, the SARIMA$(5,1,0)\,(1,0,0)^6$ model provides a satisfactory forecast of future rainfall trends; however, further refinement may be re-

Table 5. Forecasting results of the SARIMA$(5,1,0)\,(1,0,0)^6$ model

Period	Forecast	Period	Forecast
87*	268.876	97	177.164
88*	266.132	98	185.004
89*	247.501	99	233.603
90*	185.000	100	250.632
91*	187.743	101	245.740
⋮	⋮	⋮	⋮
96*	191.168	108	213.375

*) forecast out-of-sample data

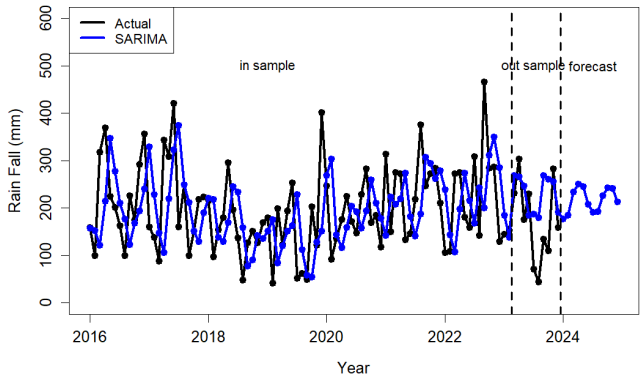


Figure 5. Comparison graph of actual data and predictions from the SARIMA$(5,1,0)\,(1,0,0)^6$ model

quired to enhance its accuracy, particularly in capturing extreme variations.

3.4. Hybrid SARIMA-RBFNN

Rainfall forecasting in Samarinda City using the SARIMA method produced an SMAPE of 34.699% and an RMSEP of 82.255. Although SARIMA effectively captures seasonal patterns, it may not fully account for nonlinear characteristics in the data. To address this limitation and enhance forecasting accuracy, a hybrid SARIMA–RBFNN model was employed. This approach combines SARIMA's strength in modeling seasonality with RBFNN's ability to capture nonlinear residual patterns, thereby improving overall predictive performance. The residual data were divided into in-sample (90%) and out-of-sample (10%) sets, covering January 2016–February 2023 and March–December 2023, respectively. RBFNN inputs were selected from significant SARIMA lags (1, 3, 4, 5, and 6) identified through the PACF plot. With lag 6 as the longest lag, the model utilized 80 data points starting from a later observation period.

The number of neurons in the hidden layer of the RBFNN model is determined using K-Means clustering with cluster numbers ranging from 2 to 8. The optimal number of clusters is identified based on the highest Silhouette Coefficient (SC) value, which measures clustering effectiveness.

From the analysis, the highest SC value (0.197) is obtained with 8 clusters, making it the most optimal configuration. However, since all SC values are below 0.25, the clusters are weakly defined due to the multicollinearity inherent in time series data,

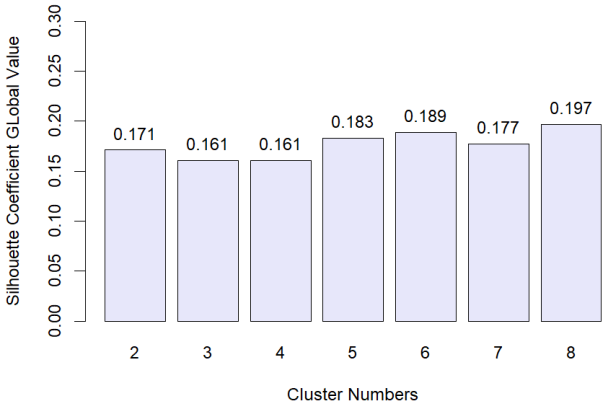


Figure 6. Evaluation results of clustering based on SC_{global} values

which hinders the formation of distinct groups. Consequently, while the RBFNN model uses eight clusters based on the highest SC value, the low coefficient indicates that the input feature space is not clearly separable, potentially reducing the network's ability to capture complex residual patterns. Despite this limitation, the 8 cluster configuration was adopted, corresponding to 8 neurons in the RBFNN hidden layer.

After determining the number of neurons in the hidden layer, the next step involves computing the activation function values to establish the corresponding weights. In this study, the Gaussian activation function was employed, as calculated using eq. (6). Based on these activation values, a Gaussian design matrix was constructed in a specific order, and the weight values were subsequently derived from this matrix, incorporating an additional bias column set to 1.

$$\varphi = \begin{bmatrix} 0.535 & 0.984 & 0.596 & \cdots & 0.674 & 1 \\ 0.535 & 0.645 & 0.376 & \cdots & 0.643 & 1 \\ 0.618 & 0.577 & 0.339 & \cdots & 0.624 & 1 \\ 0.522 & 0.832 & 0.627 & \cdots & 0.777 & 1 \\ 0.561 & 0.729 & 0.506 & \cdots & 0.947 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0.704 & 0.537 & 0.372 & \cdots & 0.617 & 1 \end{bmatrix}$$

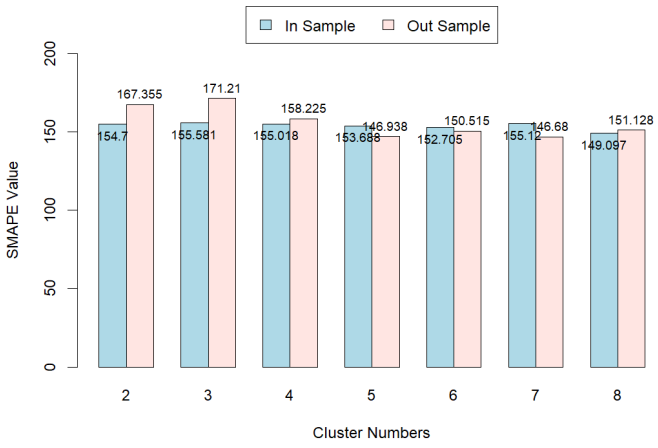
The optimal weights for 8 clusters were determined using the least squares method, as given in eq. (10).

$$\hat{\mathbf{w}} = \begin{bmatrix} -290.100 \\ -2.416 \\ 206.198 \\ 7.987 \\ 86.049 \\ 304.983 \\ -283.358 \\ 152.763 \\ -125.917 \end{bmatrix}$$

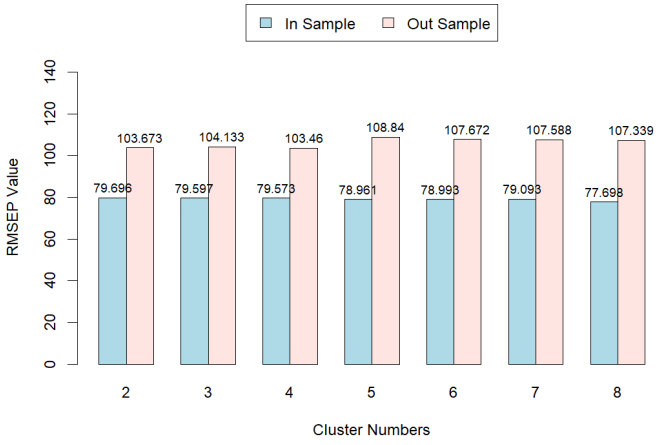
The SARIMA residual outputs for the in-sample data, representing predicted values, were obtained using eq. (9) for 80 data points spanning from period 7 to period 86. The RBFNN model's

predictive performance was then evaluated by calculating SMAPE and RMSEP using eq. (11) and eq. (12) in R, resulting in SMAPE = 149.097% and RMSEP = 77.698.

For the out-of-sample data, predictions were generated using the same approach. Gaussian activation values were derived from the in-sample centroids and maximum distances to construct the design matrix, while the weight parameters remained constant. Predictions were produced for 10 periods (87–96), and several cluster configurations were assessed to ensure the model's reliability and optimal performance.



(a)



(b)

Figure 7. SMAPE and RMSEP RBFNN model

Based on Figure 7, the configuration with 8 clusters produces the best SMAPE and RMSEP values for in-sample data, with SMAPE at 149.097% and RMSEP at 77.698, the lowest among the configurations. However, for out-of-sample data, the lowest SMAPE is seen with 5 clusters at 146.938%, and the lowest RMSEP with 4 clusters at 103,460. Despite this, the 8-cluster configuration is chosen as optimal due to its balance in accuracy for both in-sample and out-of-sample data, and it has the highest SC value of 0.185, indicating better cluster structure. Thus, 8 clusters are used, supporting the decision to utilize 8 neurons in the hidden layer of the RBFNN model to improve forecasting accuracy.

Forecasting was performed using the same method as out-of-sample residual model SARIMA output calculations. Gaussian activation functions were computed sequentially, with predictions for each period using the values from the prior period. The results are shown in Table 6.

Table 6. Residual forecasting results of the SARIMA model using the RBFNN model

Period	Forecast	Period	Forecast
97	-3.103	103	20.235
98	15.029	104	20.188
99	8.905	105	20.252
100	16.149	106	19.304
101	23.983	107	18.751
102	23.598	108	19.215

Based on the calculated output values, a comparison graph of the residual data from the SARIMA model and the predicted results using the RBFNN model with 8 neurons in the hidden layer is created, as shown in Figure 8.

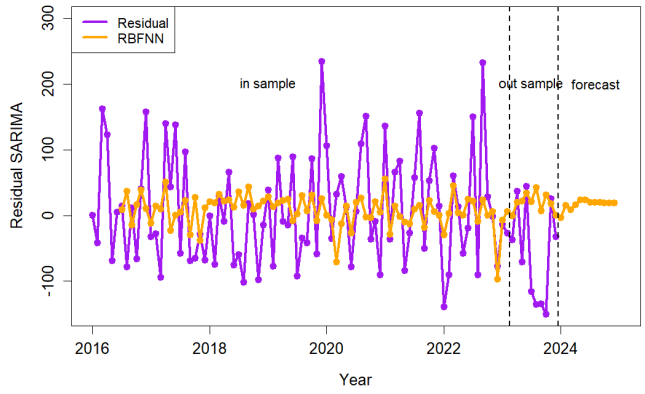


Figure 8. Comparison graph of residual SARIMA$(5,1,0)\,(1,0,0)^6$ and predictions from the RBFNN model

The comparison graph in Figure 8 indicates that SARIMA $(5,1,0)\,(1,0,0)^6$ residuals exhibit significant fluctuations, suggesting that individual SARIMA model does not fully capture data patterns. The RBFNN model reduced variability but struggled to capture all fluctuations, implying potential underfitting. The SMAPE values of 149.097% (in-sample) and 151.128% (out-of-sample) indicate a relatively high error, suggesting room for improvement in forecasting accuracy.

The hybrid SARIMA$(5,1,0)\,(1,0,0)^6$-RBFNN model was constructed using eq. (10) by combining predictions from both methods. The performance evaluation demonstrates improvements, with SMAPE values of 34.175% for in-sample predictions and 47.802% for out-of-sample predictions. Similarly, RMSEP values were measured at 78.577 for in-sample predictions and 107.339 for out-of-sample predictions, as presented in Table 7. The forecasting results for 12 periods from January 2024 to December 2024 using the hybrid model are presented in Table 8.

Based on Table 8, the rainfall pattern in Samarinda remains seasonal. The forecast indicates higher rainfall from March to

Table 7. SMAPE and RMSEP values of the hybrid SARIMA$(5,1,0)\,(1,0,0)^6$-RBFNN Model

Data	SMAPE	RMSEP
In-Sample	34.175%	78.577
Out-Sample	47.802%	107.339

Table 8. Forecasting results of the hybrid SARIMA$(5,1,0)\,(1,0,0)^6$-RBFNN model

Period	SARIMA $(5,1,0)\,(1,0,0)^6$	RBFNN	*Hybrid* SARIMA $(5,1,0)\,(1,0,0)^6$-RBFNN
97	177.164	-3.103	174.061
98	185.004	15.029	200.033
99	233.603	8.905	242.508
100	250.632	16.149	266.781
101	245.740	23.983	269.723
102	207.694	23.598	231.292
103	191.851	20.235	212.086
104	192.632	20.188	212.820
105	226.928	20.252	247.179
106	242.516	19.304	261.820
107	242.323	18.751	261.074
108	213.375	19.215	232.590

May 2024, reaching values between 242.508 mm and 269.723 mm, suggesting the potential for heavy rains. Conversely, lower rainfall is expected in January and December 2024, ranging from 174.061 mm to 232.590 mm, still within the high rainfall category. Overall, the forecast shows consistently high rainfall in Samarinda for 2024, without drastic reductions throughout the year. Despite fluctuations, Samarinda is likely to experience frequent rains. The hybrid SARIMA-RBFNN model captures the seasonal pattern well, with peak rainfall in early to mid-year and slight decreases towards year-end.

Based on the calculated predict and forecast values, a comparison graph of actual and forecasted rainfall in Samarinda from January 2016 to December 2024 is created, as shown in Figure 9.

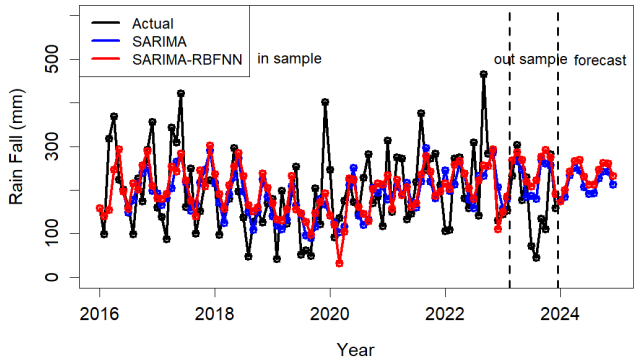


Figure 9. Comparison graph of actual and forecasted rainfall in Samarinda

Figure 9 illustrates the comparison between actual rainfall data in Samarinda and the forecasts generated by the SARIMA$(5,1,0)\,(1,0,0)^6$ and the hybrid SARIMA$(5,1,0)\,(1,0,0)^6$-RBFNN models. The hybrid model demonstrates forecasts that more closely align with the actual rainfall trend for both in-sample and out-of-sample periods, particularly during peak rainfall events. This improvement shows that the RBFNN component effectively captures nonlinear variations that the SARIMA model alone cannot identify, resulting in higher predictive accuracy and a smoother overall forecast.

Compared with the ARIMA–LSTM model by Mahajan et al. [3], the proposed SARIMA–RBFNN achieved competitive accuracy despite using far less data. While Mahajan's deep learning model required large, high-resolution datasets and complex training, SARIMA–RBFNN delivered reliable forecasts with only 96 monthly observations. Its simpler structure, interpretability, and low computational demand make it more practical for regional applications such as Samarinda's flood early warning system.

4. Conclusion

The results indicate that the hybrid SARIMA(5,1,0)(1,0,0)□–RBFNN model effectively enhances the reliability of the flood early warning system in Samarinda City by providing more accurate and robust rainfall forecasts. Compared to the single SARIMA model, the hybrid approach yields lower RMSEP and SMAPE values, indicating superior predictive performance and a better ability to capture both seasonal and nonlinear patterns in rainfall data. These improvements significantly contribute to reducing uncertainty in rainfall estimation, which is crucial for timely flood prediction and mitigation. In practical implementation, the rainfall forecasts generated by the hybrid SARIMA–RBFNN model can be operationalized by local agencies such as BMKG Samarinda and BPBD to issue early flood warnings, schedule drainage system maintenance, and allocate emergency response resources during high-risk months. The model's computational efficiency also allows seamless integration with existing IoT-based rainfall sensors and real-time monitoring platforms, enhancing the overall responsiveness and effectiveness of the city's flood early warning system. Consequently, this modeling approach supports early preparedness, minimizes potential flood impacts, and strengthens community resilience in flood-prone areas of Samarinda City.

Author Contributions. **Syifa Mutia Rahmah**: Problem identification, case study selection, methodology, data analysis, and manuscript writing. **Andrea Tri Rian Dani**: Conceptualization, manuscript review, and supervision.**Sri Wahyuningsih**: Conceptualization, manuscript review, and supervision. All authors discussed the results and contributed to the final manuscript.

Acknowledgement. The authors would like to express their gratitude to all parties who contributed to this research and the preparation of the manuscript. We highly appreciate the editor and reviewers for their valuable feedback and support in improving this work.

Funding. This research did not receive any external funding.

Conflict of interest. The authors declare no conflict of interest related to this article.

Data availability. Not applicable.

References

[1] M. Kolambe and S. Arora, "Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques," *J. Electr. Syst.*, vol. 20, no. 2s, pp. 575–586, 2024, doi: 10.52783/jes.1478.

[2] K. Szostek, D. Mazur, G. Dralus, and J. Kusznier, "Analysis of the Effectiveness of ARIMA, SARIMA, and SVR Models in Time Series Forecasting: A Case Study of Wind Farm Energy Production," *Energies*, vol. 17, no. 4803, 2024, doi: 10.3390/en17194803.

[3] D. Mahajan, S. Sharma, and K. Saini, "Real-time rainfall projection in hilly areas using ARLSTM model: a case of Dharamshala, India," *Eng. Res. Express*, vol. 6, no. 4, 2024, doi: 10.1088/2631-8695/ad9c17.

[4] H. Ihsan, Irwan, and A. I. E. Nensi, "Implementation of Backpropagation and Hybrid Arima-Nn Methods in Predicting Accuracy Levels of Rainfall in Makassar City," *Barekeng*, vol. 18, no. 4, pp. 10–14, 2024, doi: 10.30598/barekengvol18iss4pp2435-2448.

[5] M. Silfiani, F. N. Hayati, D. Nurlaily, and I. Fitria, "Household Electrical Load Forecasting: a Hybrid of Linear Models and Radial Basis Function Neural Network," in *2021 Int. Conf. Adv. Mechatronics, Intell. Manuf. Ind. Autom.*, pp. 253–257, 2021, doi: 10.1109/ICAMIMIA54022.2021.9807693.

[6] T. S. Nguyen and C. C. Pham, "Empirical Evaluation of the Time Series Forecasting Method by Combining ARIMA with RBFNN under the Additive Model," *J. Tech. Educ. Sci.*, vol. 19, no. 1, pp. 1–7, 2024, doi: 10.54644/jte.2024.1520.

[7] K. L. Du and M. N. S. Swamy, *Neural Networks and Statistical Learning*, 2nd ed., 2019, doi: 10.1007/978-1-4471-7452-3.

[8] I. W. A. Suranata, "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model," *J. Sist. dan Inform.*, vol. 18, no. 1, pp. 64–73, 2003, doi: 10.30864/jsi.v18i1.603.

[9] S. D. Latif et al., "Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches," *Alexandria Eng. J.*, vol. 82, pp. 16–25, 2023, doi: 10.1016/j.aej.2023.09.060.

[10] M. Wildi, *Lecture Notes in Economics and Mathematical Systems: Preface*, vol. 614, Heidelberg: Springer, 2005.

[11] W. W. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed., New York: Pearson, 2006, doi: 10.2307/2289741.

[12] J. Kaur, K. S. Parmar, and S. Singh, "Autoregressive models in environmental forecasting time series: a theoretical and application review," *Environ. Sci. Pollut. Res.*, vol. 30, no. 8, pp. 19617–19641, 2023, doi: 10.1007/s11356-023-25148-9.

[13] M. K. Sharma, M. Omer, and S. Kiani, "Time series analysis on precipitation with missing data using stochastic SARIMA," *Mausam*, vol. 71, no. 4, pp. 617–624, 2020.

[14] J. D. Cryer and K. S. Chan, *Time Series Analysis with Applications in R*, 2nd ed., USA: Springer, 2008.

[15] U. A. Yakubu and M. P. A. Saputra, "Time Series Model Analysis Using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for E-wallet Transactions during a Pandemic," *Int. J. Glob. Oper. Res.*, vol. 3, no. 3, pp. 80–85, 2022, doi: 10.47194/ijgor.v3i3.168.

[16] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesus, *Neural Network Design*, 2014, doi: 10.1007/1-84628-303-5.

[17] I. T. Utami, F. Suryaningrum, and D. Ispriyanti, "K-Means Cluster Count Optimization With Silhouette Index Validation and Davies Bouldin Index," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 0707–0716, 2023, doi: 10.30598/barekengvol17iss2pp0707-0716.

[18] L. J. Slater et al., "Hybrid forecasting: blending climate predictions with AI models," *Hydrol. Earth Syst. Sci.*, vol. 27, no. 9, pp. 1865–1889, 2023, doi: 10.5194/hess-27-1865-2023.

[19] D. Takeyasu, H. Yamashita, and K. Takeyasu, "A Hybrid Method to Improve Forecasting Accuracy in the Case of Sanitary Materials Data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 5, pp. 59–67, 2014, doi: 10.14569/ijacsa.2014.050509.

[20] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, 2003.

[21] N. M. Faber, "Estimating the uncertainty in estimates of root mean square error of prediction: Application to determining the size of an adequate test set in multivariate calibration," *Chemom. Intell. Lab. Syst.*, vol. 49, no. 1, pp. 79–89, 1999, doi: 10.1016/S0169-7439(99)00027-1.

[22] S. Makridakis and M. Hibon, "The M3-Competition: Results, Conclusions and Implications," *Int. J. Forecast.*, vol. 16, pp. 451–476, 2000, doi: 10.1016/S0169-2070(00)00057-1.