International Journal of Advances in Data and Information Systems
Vol. 7, No. 2, August 2026, pp. 671~686
ISSN: 2721-3056, DOI: 10.25008/ijadis.v7i2.1549

Modeling Earth Skin Temperature and Cloud Amount: A Hybrid Fourier Series – Neural Network for Kalimantan Environmental Patterns

Andrea Tri Rian Dani[1,2], Nur Chamidah[3,4], I Nyoman Budiantara[5], Dursun Aydin[6], Ahmad Hakiim Jamaluddin[7], Naufal Ramadhan Al Akhwal Siregar[8]

[1] Doctoral Study Program MIPA, Faculty of Science and Technology, Universitas Airlangga, Indonesia
[2] Statistics Study Program, Department of Mathematics, FMIPA, Mulawarman University, Indonesia
[3] Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Indonesia
[4] Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Universitas Airlangga, Indonesia
[5] Department of Statistics, Faculty of Science and Data Analytics, Sepuluh Nopember Institute of Technology, Indonesia
[6] Department of Statistics, Faculty of Science, Mugla Sitki Kocman University, Turkey
[7] Department of Mathematics & Statistics, Universiti Putra Malaysia, Malaysia
[8] Master of Mathematics Study Program, Faculty of Science and Technology, Airlangga University, Indonesia

Article Info

Article history:
Received Feb 24, 2026
Revised May 30, 2026
Accepted June 06, 2026

Keywords:
Climate Action
Cloud Amount
Earth Skin Temperature
Fourier Series
Hybrid Model

ABSTRACT

This study focuses on the development and application of a Hybrid Fs-DLFFNN model to analyze Earth Skin Temperature and Cloud Amount in Balikpapan and Samarinda. For comparison purposes, Fourier Series (Fs) and Double Layer Feed Forward Neural Network (DLFFNN) models are also considered. Model performance is evaluated using MSE, RMSE, and MAPE on both training and testing data. The results indicate that the Hybrid Fs-DLFFNN model consistently achieves lower error values on the training data, reflecting its ability to capture both periodic and nonlinear patterns within a unified framework. However, testing results show that model performance varies across variables and locations, highlighting the importance of generalization. Forecasting for the next 18 periods provides indicative patterns of temporal variation in both cities. These projections are presented as exploratory insights rather than definitive conclusions. Overall, the Hybrid Fs-DLFFNN model offers a flexible and integrative approach for modeling complex environmental data. This study contributes to the advancement of hybrid modeling frameworks for environmental time series and provides a foundation for further methodological development and climate-related analysis, in line with the objectives of SDG 13 (Climate Action).

This is an open access article under the CC BY-SA license.

Corresponding Author:
Nur Chamidah,
Department of Mathematics, Faculty of Science and Technology
Universitas Airlangga, Indonesia
Email: nur-c@fst.ac.id

1. INTRODUCTION

Sustainable Development Goal (SDG) 13, focused on climate change and global warming, is a critical issue that must be addressed seriously [1]. Climate change impacts many aspects of life, including weather patterns, temperature, and humidity. One measurable phenomenon related to climate change is earth skin temperature and cloud amount [2]. Earth skin temperature reflects the temperature directly affected by global warming [3]. Meanwhile, cloud amount refers to the

average proportion of clouds in the atmosphere over a given period, which can change as climate patterns shift. Cloud plays a vital role in the climate system because it can influence both earth skin temperature and rainfall patterns, which are closely tied to global climate change[4]. In this context, measuring and analyzing cloud amount provides essential insights into the dynamics of a changing climate [5].

As a space research and observation agency, the National Aeronautics and Space Administration (NASA) continuously records hydrological phenomena in real time. One invaluable data source is the MERRA-2 satellite, managed by NASA, which provides data on earth skin temperature and cloud amount [6]. The availability of data from the MERRA-2 satellite significantly supports research and modeling to address the challenges related to SDG 13, namely, taking urgent action to address climate change and its impacts [7], [8]. These two issues are fascinating to analyze through statistical modeling using time series analysis. Time series analysis is a statistical technique used to model sequential data over time, to identify patterns and trends, and make predictions for future periods [9].

In the context of climate change, time series analysis can provide essential insights into how earth skin temperature and rainfall may behave in the future and their impact on human life and ecosystems [10]. There are many methods for modeling time series data, such as exponential smoothing [11], moving averages [12], Autoregressive Integrated Moving Average (ARIMA)[13], Singular Spectrum Analysis [14], and Fourier Series [15]. In machine learning, there are Neural Network methods with various architectures that can also be applied to time series data. Each method has its advantages and characteristics when modeling time series data [16]. In this study, researchers are interested in examining two relevant methods, the Fourier Series and the Double Layer Feed Forward Neural Network (DLFFNN).

Fourier series is an estimator in nonparametric regression that can be generalized to time series data, using time as a predictor [17]. The main advantage of the Fourier series is its ability to model data with seasonal or periodic patterns [18]. This is achieved by utilizing the oscillations or periodic waves generated by the Fourier series. The optimal oscillation is selected based on the Generalized Cross-Validation (GCV) criterion, which allows the selection of the model that best fits the existing data, resulting in the most accurate model in predicting seasonal patterns. Additionally, a Double Layer Feed Forward Neural Network (DLFFNN) is a type of neural network that uses two hidden layers in its architecture [19]. This is intended to improve the model's ability to capture more complex patterns in data, both linear and non-linear. DLFFNN has the advantage of modeling more complicated relationships between input and output variables, which is particularly useful in modeling hydrological data that is heavily influenced by multiple factors.

The novelty of this research lies in developing the Hybrid Fourier Series – DLFFNN (Fs-DLFFNN) model. The advantage of the Hybrid Fs-DLFFNN model is its ability to combine the power of Fourier Series analysis that can capture periodic patterns in data, with the ability of DLFFNN to model more complex non-linear patterns. This combination is expected to produce a more accurate and flexible model in predicting earth skin temperature and cloud amount. Based on the literature review, no research has examined the development of this model comprehensively, so this research can fill the research gap and become a significant innovation in the field of hydrological modeling. This research aims to develop the Hybrid Fs-DLFFNN model and compare it with the individual Fourier Series and DLFFNN models, to see the accuracy of each in modeling earth skin temperature and cloud amount data. The results of this research are expected to provide an essential contribution to the development of predictive models for climate change, as well as contribute to the achievement of SDG 13, namely, reducing the impact of climate change through more accurate and innovative modeling.

2. RESEARCH METHOD

This section outlines the research methods, including explanations of the time series data, Fourier series, neural network, hybrid model, evaluation model, data sources and research variables, and the research stages.

2.1. Time Series Analysis

Time series data is a collection of observational data recorded sequentially following a time sequence with consistent intervals. Time series analysis is a statistical method used to predict future events based on historical data [9]. This analysis process is carried out by processing data from the current period (y_t), which is influenced by one or more observations from the previous period (y_{t-k}). Patterns in time series data can be categorized into four main groups in Figure 1 [20].

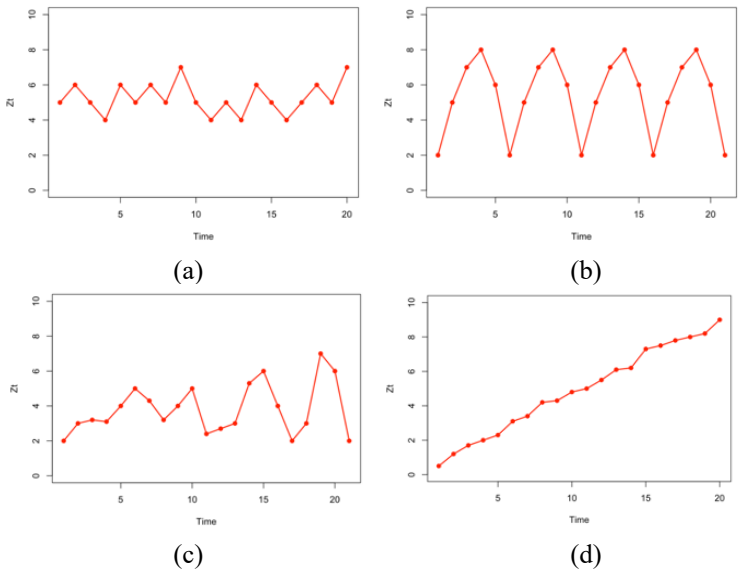


Figure 1. Time series data patterns (a) horizontal; (b) seasonal; (c) cyclical; (d) trend

2.2. Fourier Series

The Fourier series estimator is a highly flexible trigonometric polynomial, allowing it to optimally adapt to local data characteristics [21]. This method is generally used when the data pattern is unknown and there is a tendency for recurring values. This pattern repetition refers to the reappearance of the response variable value in each observation of the predictor variable [17]. For example, given paired observation data (v_t, y_t) that follow the nonparametric Fourier series regression model in Equation (1).

$$y_t = g(v_t) + \varepsilon_t \tag{1}$$

with $t = 1,2, \dots, n$.

The regression function $g(v_t)$ is unknown in form and will be estimated using this study's nonparametric Fourier series estimator regression approach. Suppose it is assumed that the regression function $g(v_t)$ is contained in the Hilbert space $L_2[a,b]$, namely $g(v_t) \in L_2[a,b]$,, so that $g(v_t)$ can be expressed as a linear combination of the basis elements of $L_2[a,b]$($Rahma\ et\ al.$, 2024). Unknown Fourier series coefficients are estimated by determining the optimal λ value. The optimal λ value states the number of Fourier coefficients (γ_w) that define the smoothness of the estimated regression curve. The complex exponential function can be expressed as Sine and Cosine functions, so the complete orthonormal system of $L_2[a,b]$ can be expressed in Equation (2).

$$x_w(v_t) = e^{2\pi i w v_t} \ ; w = 0, \pm 1, \pm 2, \dots ; t = 1,2, \dots, n \ ; i = \sqrt{-1} \tag{2}$$

In the context of the Fourier series, i refers to the unit imaginary number. Suppose n paired samples (v_t, y_t) are taken that satisfy Equation (2) to obtain Equation (3).

$$\boldsymbol{y} = \boldsymbol{V}(\lambda)\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{3}$$

The estimator for the parameter $\hat{\boldsymbol{\gamma}}$ is given in Equation (4).

$$\hat{\boldsymbol{\gamma}} = \left(\boldsymbol{V}(\lambda)^T \boldsymbol{V}(\lambda)\right)^{-1} \boldsymbol{V}(\lambda)^T \boldsymbol{y} \tag{4}$$

$\boldsymbol{V}(\lambda)^T$ is the complex transpose matrix of the matrix $\boldsymbol{V}(\lambda)$. Based on Equation (4), the estimator for γ_w is obtained in Equation (5).

$$\gamma_w = \frac{1}{n}\sum_{t=1}^{n} y_t e^{-2\pi i w v_t} \; ; w = -\lambda, (-\lambda+1) \dots, \lambda \; ; t = 1,2, \dots, n \; ; i = \sqrt{-1} \tag{5}$$

If $[a, b] = [0,1]$ and v_t is assumed to be equidistant in the range $[0,1]$ then the notation for v_t can be written as $v_t = \frac{(t-1)}{n}$.

$$\hat{g}(v_t) = \sum_{w=-\lambda}^{\lambda} \hat{\gamma}_j \, e^{2\pi i w \frac{(t-1)}{n}} \tag{6}$$

Based on Equation (6), the function $g(v_t)$ will have real and imaginary components, respectively, where both are real-valued functions that can be expressed as sine and cosine functions. If $\hat{\gamma}_{(-w)}$ is defined as a complex number conjugate of $\hat{\gamma}_w$, so that $\rho_w = \hat{\gamma}_w + \hat{\gamma}_{(-w)}$ and $\vartheta_w = -i\left(\hat{\gamma}_w - \hat{\gamma}_{(-w)}\right)$ then the values of ρ_w and ϑ_w can be represented as a combination of sine and cosine functions based on Euler's identity.

$$e^{ix} = \cos(x) + i \sin(x) \tag{7}$$

and

$$e^{-ix} = \cos(x) - i \sin(x) \tag{8}$$

The values of ρ_w and ϑ_w can be described as follows:

$$\rho_w = \frac{2}{n}\sum_{t=1}^{n} y_t \cos\left(2\pi w\left(\frac{(t-1)}{n}\right)\right) \tag{9}$$

and

$$\vartheta_w = \frac{2}{n}\sum_{t=1}^{n} y_t \sin\left(2\pi w\left(\frac{(t-1)}{n}\right)\right) \tag{10}$$

Based on the description in Equations (9) and (10), the form of the Fourier series estimator which is a function of time can be written in Equation (11).

$$\hat{g}(v_t) = \hat{\gamma}_0 + \sum_{w=1}^{\lambda}\left(\hat{\rho}_w \cos\left(2\pi w\left(\frac{t-1}{n}\right)\right) + \hat{\vartheta}_w \sin\left(2\pi w\left(\frac{t-1}{n}\right)\right)\right) \tag{11}$$

Determining the optimal oscillation (symbolized with λ) based on data using the minimum Generalized Cross-Validation (GCV) criterion in Equation (12).

$$GCV(\lambda_{opt}) = Min\frac{MSE(\lambda)}{\left(n^{-1} trace\,(\boldsymbol{I} - \boldsymbol{H}(\lambda)\right)^2} \tag{12}$$

where:

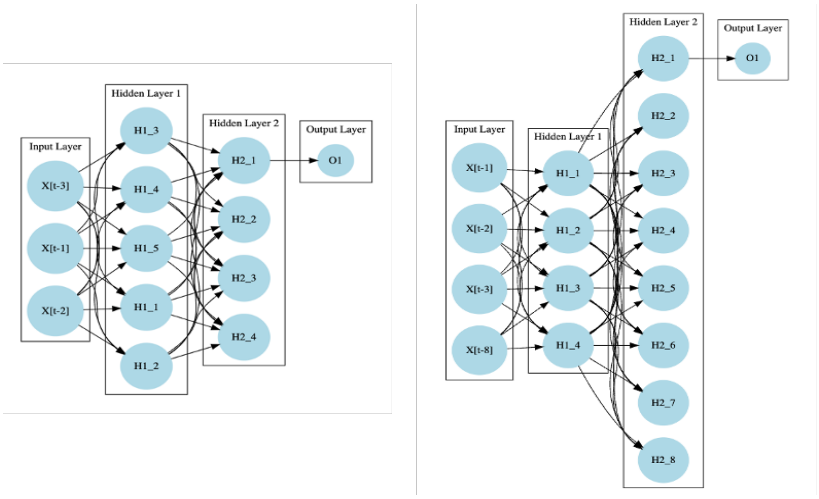
$$MSE(\lambda) = n^{-1}\sum_{t=1}^{n}\left(y_t - \hat{m}(v_t)\right)^2$$

$$\boldsymbol{H}(\lambda) = \begin{pmatrix} \frac{(2\lambda+1)}{n} & \frac{(2\lambda+1)}{n} & \cdots & \frac{(2\lambda+1)}{n} \\ \frac{(2\lambda+1)}{n} & \frac{(2\lambda+1)}{n} & \cdots & \frac{(2\lambda+1)}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(2\lambda+1)}{n} & \frac{(2\lambda+1)}{n} & \cdots & \frac{(2\lambda+1)}{n} \end{pmatrix}.$$

2.3. Neural Network

A Double Layer Feed Forward Neural Network (DLFFNN) is a type of Artificial Neural Network (ANN) that uses two or more hidden layers [23]. Each layer consists of several neurons

connected, forming a deep network. Each neuron in a hidden layer receives input from the previous neuron, processes the information, and sends the result to the next neuron. DLFFNN can model more complex non-linear relationships between inputs and outputs compared to Single Layer Perceptron (SLP), because multiple hidden layers allow the network to capture more complicated patterns in the data. The training process for DLFFNN typically involves a backpropagation algorithm.



(a) (b)

Figure 2. Architecture of DLFFNN

Figure 2 shows two diagrams of NN architectures with two hidden layers. Figure 2 (a) has three input nodes $(X_{(t-1)}, X_{(t-2)}, X_{(t-3)})$, with five neurons in the first hidden layer and four neurons in the second hidden layer, resulting in one output neuron. Figure 2 (b) has four input nodes $(X_{(t-1)}, X_{(t-2)}, X_{(t-3)}, X_{(t-8)})$, with four neurons in the first hidden layer and eight neurons in the second hidden layer, and one output neuron.

The DLFFNN architecture consists of three main layers that are very effective for modeling time series with non-linear and complex patterns:

- **Input Layer:** The first layer receives input data, such as historical values from a time series. Each neuron in this layer represents a single data feature, such as temperature or rainfall at a specific point in time.
- **Hidden Layers:** Hidden layers that process data non-linearly to extract more complex patterns. DLFFNN can capture seasonal trends and periodic fluctuations in time series data with multiple hidden layers.
- **Output Layer:** The final layer that generates predictions based on data processing from the previous layers. In time series, the output predicts future values, such as temperature or rainfall in the coming period.

The calculation of the output value $\hat{y}_t$ in the DLFFNN model can be generalized from SLP in Equation (13).

$$\hat{y}_t = f_0\left[\sum_{l=1}^{r}\left[w_l^o f_l^h\left(\sum_{j=1}^{q} w_{lh}^h f_j^h\left(\sum_{i=1}^{p} w_{ji}^h x_{i(k)} + b_j^h\right) + b_l^h\right)\right] + b^o\right] \quad (13)$$

where:

$x_{i(k)}$ is the input variable;

$\hat{y}_t$ is the output value of the time series data;

w_{ji}^h is the weight of input i leading to neuron j in the first hidden layer;

f_j^h is the activation function on neuron j in the first hidden layer;

b_j^h is the common factor on neuron j in the first hidden layer;
w_l^o is the weight of neuron l in the second hidden layer leading to the output layer;
b^o is the common factor on neurons in the output layer;
f_0 is the activation function in the output layer [24].

2.4. Hybrid Model

Zhang (2005) proposed a hybrid model combining ARIMA and ANN to forecast time series data [25]. The stages of the ARIMA-ANN hybrid modeling proposed by Zhang are as follows:

Step 1: ARIMA is used to model and forecast the linear portion of the data. After ARIMA is applied, the residuals from the ARIMA model are calculated.

Step 2: The residuals from the ARIMA model, suspected to represent nonlinear patterns, are then modeled with an ANN. The ANN is trained based on these residuals, identifying inputs based on the ACF and PACF graphs.

Step 3: The final prediction result is obtained by summing the predictions from the ARIMA model and the ANN output.

The Hybrid Model Formula proposed by Zhang can be written in Equation (14).

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{14}$$

$\hat{L}_t$ is the linear component of the ARIMA results, and $\hat{N}_t$ is the nonlinear component of the ANN results. The Hybrid Fourier Series-DLFFNN model is generalized using the same steps. This hybrid model combines the power of the Fourier Series (to capture the main linear and periodic patterns) with the DLFFNN (to capture nonlinear or residual patterns).

2.5. Evaluation Metrics

To assess the performance of the forecasting model used in this study, an evaluation measure is needed that can describe the level of error between the predicted value and the actual value. Some error measures used in the analysis are Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Mean Squared Error (MSE) measures the average of the squared differences between the actual values y_t and the predicted values $\hat{y}_t$ resulting from a model/estimation result. MSE is widely used in statistics and machine learning to evaluate the accuracy of predictive models. The MSE formula is written in Equation (15).

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \tag{15}$$

The Root Mean Squared Error (RMSE) is the square root of the MSE value. RMSE returns the error units to the actual data units, making the results more intuitive. The RMSE formula is written in Equation (16).

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{16}$$

The Mean Absolute Percentage Error (MAPE) measures the average absolute error between actual and predicted values as a percentage. A smaller MAPE value indicates a lower prediction error in percentage terms. In the literature, MAPE of <10% is categorized as excellent, 10-20% as good, 20-50% as fair, and >50% as poor. The MAPE formula is written in Equation (17).

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{17}$$

2.6. Data Sources and Research Variables

This research uses earth skin temperature and cloud amount as case studies, obtained through scraping NASA MERRA-2 satellite data from January 2015 to June 2025. This research focused on two cities in East Kalimantan Province: Samarinda and Balikpapan. Data were obtained using a purposive sampling technique, selected based on the novelty and relevance of the data received. This technique was chosen to ensure the data used in the study was up to date and aligned with the research objectives. The variables used in this study are further presented in Table 1.

Table 1. Research Variables

Notation	Variable	Unit	Operational Definition
y_{1t}	Earth Skin Temperature	Degree Celcius (C)	The average temperature at the earth skin.
y_{2t}	Cloud Amount	Percentage (%)	The average percent of cloud amount during the temporal period.

2.7. Framework Conceptual Idea

The stages of this research are systematically carried out to ensure that the forecasting process is accurate and reliable. The research begins with data collection and exploration, followed by data partitioning, model development, evaluation, and forecasting. The detailed stages are as follows:

1. Collect data from the MERRA 2 satellite for the East Kalimantan region, specifically Samarinda and Balikpapan, covering earth skin temperature and cloud amount variables.
2. Explore the data using time series plot and box plot to identify seasonal patterns in the earth skin temperature and cloud amount data.
3. Divide the data into training and testing data in a 90:10 ratio.

Table 2. Splitting Data into Training and Testing Sets

Splitting Data	Time Period
Training	January 2015 – June 2024
Testing	July 2024 – June 2025

4. Modeling training data using Fs, DLFFNN, and Hybrid Fs – DLFFNN.
5. Forecasting testing data, which is then evaluated using MSE, RMSE, and MAPE criteria.
6. Determining the best model based on MSE, RMSE, and MAPE criteria on the training and testing data.
7. Forecasting the earth skin temperature and cloud amount data for Samarinda and Balikpapan cities for the following 18 periods.
8. Evaluating the model and developing mitigation strategies based on the forecast results.

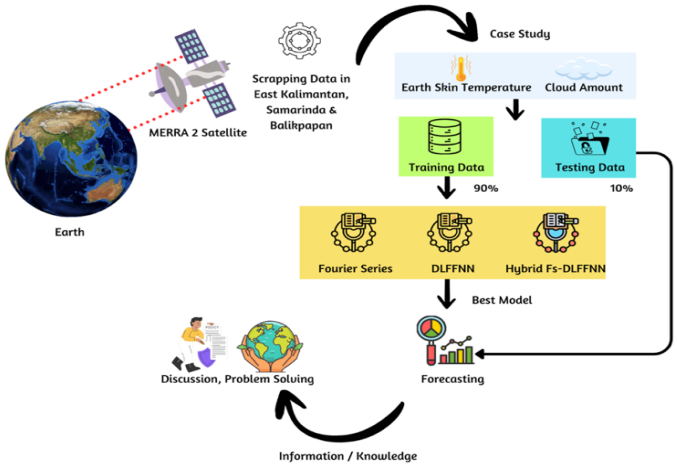


Figure 3. Research Framework

3. RESULTS AND DISCUSSION

3.1. Data Exploration

The data exploration stage includes time series plots, boxplots, and correlation heatmaps, which are used to provide an initial understanding of the data characteristics. The time series plots are utilized to observe temporal patterns, trends, and potential seasonality in the variables. Boxplots are employed to examine the distribution of the data and to identify the presence of outliers. Meanwhile, correlation heatmaps are used as an exploratory tool to assess the general relationships between variables.

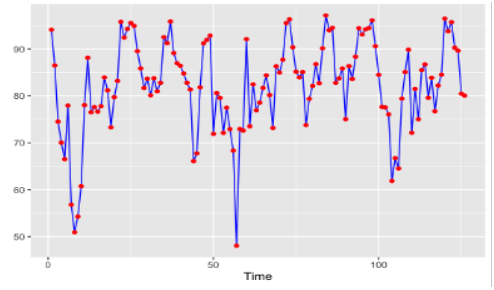


Figure 4. Time series plot of Cloud Amount in Samarinda

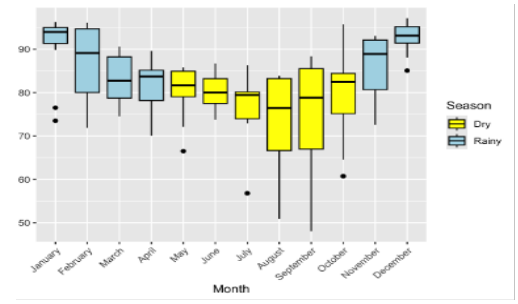


Figure 5. Box plot of Cloud Amount in Samarinda

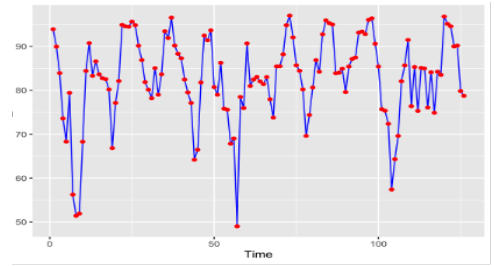


Figure 6. Time series plot of Cloud Amount in Balikpapan

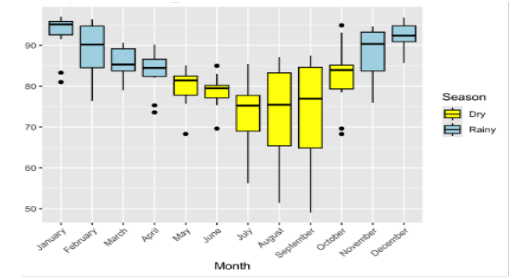


Figure 7. Box Plot of Cloud Amount in Balikpapan

Based on the analysis of time series plots (Figure 4 & Figure 6) and box plots (Figure 5 & Figure 7) related to cloud amount in Samarinda and Balikpapan from January 2015 to June 2025, a clear seasonal pattern is observed, influenced by the rainy and dry seasons, which are key characteristics of Indonesia's tropical climate. During the rainy season (November to March), cloud amount tends to be higher and more stable, reflecting higher rainfall and higher humidity. This is consistent with the natural phenomenon where increased rainfall leads to more consistent cloud formation. Conversely, during the dry season (June to October), cloud amount is lower with greater variability, as rainfall decreases and a shift towards drier weather occurs. These more significant fluctuations illustrate the highly characteristic seasonal changes in Indonesia's tropical climate, where the transition between the rainy and dry seasons influences cloud formation patterns and distribution. Furthermore, outliers seen in the box plots indicate occasional extreme weather events, both in unexpected heavy rains and droughts during both seasons.

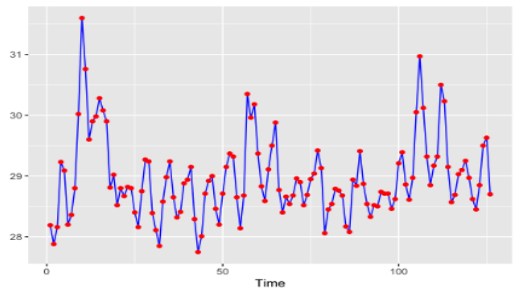


Figure 8. Time series plot of Earth Skin Temperature in Samarinda

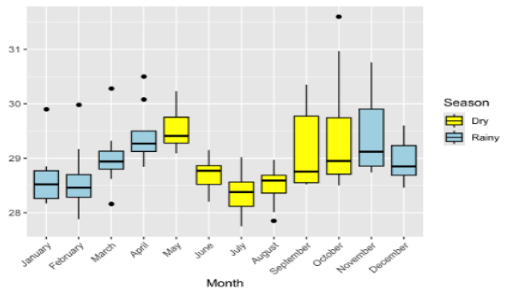


Figure 9. Box plot of Earth Skin Temperature in Samarinda

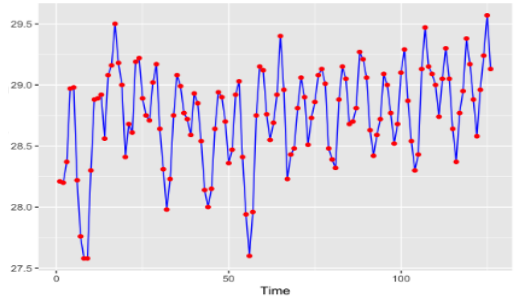


Figure 10. Time series plot of Earth Skin Temperature in Balikpapan

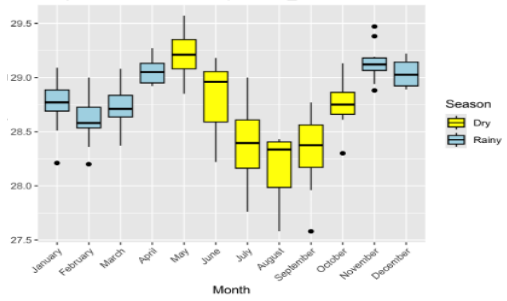


Figure 11. Box plot of Earth Skin Temperature in Balikpapan

Based on the analysis of time series plots (Figure 8 & Figure 10) and box plots (Figure 9 & Figure 11) of Earth skin temperature in Samarinda and Balikpapan, a significant seasonal pattern is observed. During the dry season (June to October), Earth skin temperature tends to be higher and more stable, with Samarinda reaching a peak temperature of around 31°C and Balikpapan slightly lower at around 29°C. This is due to dry weather conditions with few clouds blocking direct sunlight. Conversely, during the rainy season (November to March), temperatures decrease, with greater fluctuations, especially in Samarinda, reflecting increased humidity and cloud formation that reduce solar heating. The box plot shows that the temperature range is narrower and more stable during the dry season, while during the rainy season, temperatures tend to be lower with greater variations, especially in Samarinda. Overall, this pattern illustrates the influence of seasons on Earth skin temperature, which is higher and more stable during the dry season, and lower and more variable during the rainy season.

Based on the correlation matrix in Figure 12, the relationship between cloud amount and earth skin temperature in two cities, Balikpapan and Samarinda, can be interpreted.

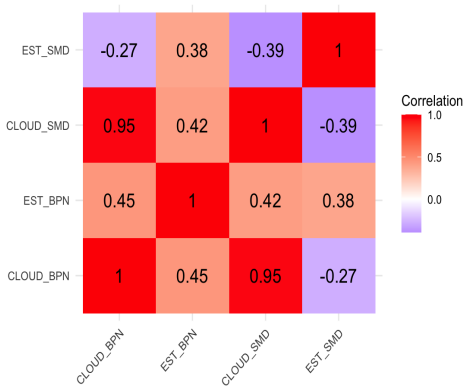


Figure 12. Heatmap matrix correlation

The cloud amount in Balikpapan shows a strong correlation with the cloud amount in Samarinda, with a correlation value of 0.954, reflecting a positive relationship. This indicates that changes in cloud amount in Balikpapan are consistently directly proportional to changes in Samarinda. Conversely, the relationship between cloud amount in Balikpapan and earth skin temperature in Samarinda is relatively weak and negative, with a correlation value of -0.272, indicating that increases in cloud amount in Balikpapan are not directly proportional to increases in earth skin temperature in Samarinda.

Earth skin temperature in Balikpapan has a moderate positive correlation with cloud amount in Balikpapan, as reflected in a correlation value of 0.449. Although this relationship is not as strong as the one between cloud amount in the two cities, it indicates a tendency for temperature in

Balikpapan to increase with increasing cloud amount. This suggests that cloud amount can influence the earth skin temperature in Balikpapan.

Cloud amount in Samarinda has a moderate negative correlation with the earth skin temperature in Samarinda, with a correlation value of -0.388. This indicates that the Earth skin temperature tends to decrease as more clouds form in Samarinda. This phenomenon can be explained by cloud formation blocking direct sunlight, thereby reducing heating of the Earth skin temperature.

Earth skin temperature in Samarinda shows a moderate positive correlation with the earth skin temperature in Balikpapan, with a correlation value of 0.384. This indicates that Earth skin temperature in both cities tends to fluctuate, although this relationship is not particularly strong. Overall, these correlation results suggest that the relationship between cloud amount and earth skin temperature tends to be stronger in each city's context.

3.2. Modeling with Fs, DLFFNN, and Hybrid Fs-DLFFNN

The data modeling was conducted using three different approaches: the Fourier series (Fs) estimator, the Double Layer Feed Forward Neural Network (DLFFNN), and the hybrid Fs-DLFFNN approach. In the first approach, the Fourier series model was tested with varying oscillations ranging from 1 to 5. The optimal oscillation count was selected based on Generalized Cross-Validation (GCV), where the oscillation that produced the smallest GCV value was chosen as the best for capturing the periodic patterns in the data. This approach identifies periodic components or cycles within the data, such as seasonal patterns or cyclic variations that could influence the analyzed variables.

Next, the Double Layer Feed Forward Neural Network (DLFFNN) was applied using the lagged variables y_{t-1} and y_{t-2}, which were identified through Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analysis. The DLFFNN model consists of two hidden layers, with the number of neurons in each layer tested from 1 to 9 neurons. The training process used the backpropagation algorithm, with the data first undergoing min-max standardization to ensure consistent scaling of the input variables. The best architecture for this model was selected based on Mean Absolute Percentage Error (MAPE), where the model with the smallest MAPE value was considered the most accurate in forecasting future values.

The hybrid Fs-DLFFNN approach was then tested to improve prediction accuracy by combining the previous methods. The optimal Fourier series model residuals were re-modeled using the DLFFNN in this approach. The input variables for the DLFFNN model were also y_{t-1} and y_{t-2}, derived from the ACF and PACF analysis of the residuals. This hybrid approach aims to leverage the strengths of both methods, Fourier Series ability to capture periodic patterns and DLFFNN ability to predict non-linear relationships, thereby improving the overall prediction results. The modeling results for earth skin temperature and cloud amount are presented in Table 3 for Balikpapan City and Table 4 for Samarinda City. These tables compare the three approaches in modeling the data for both cities, focusing on the accuracy of the predictions generated by each method.

Table 3. Modeling Earth Skin Temperature and Cloud Amount for Balikpapan

Variable	Method	Training			Testing		
		MSE	RMSE	MAPE	MSE	RMSE	MAPE
Cloud Amount	Fs	84.45	9.18	9.48	141.32	11.88	11.12
	DLFFNN (4,3)	49.26	7.02	6.65	79.50	8.91	8.63
	Hybrid Fs-DLFFNN (5,3)	**40.87**	**6.39**	**6.23**	102.22	10.11	11.87
Earth Skin Temperature	Fs	0.13	0.36	1.03	0.43	0.65	1.94
	DLFFNN (4,3)	0.04	0.21	0.59	0.11	0.33	0.84
	Hybrid Fs-DLFFNN (8,5)	**0.03**	**0.17**	**0.48**	0.49	0.70	1.88

Noted: DLFFNN (a,b) indicates a network architecture with a neurons in the first hidden layer and b neurons in the second hidden layer.

Table 3 presents a comparison of the performance of three modeling approaches, namely Fs, DLFFNN, and Hybrid Fs-DLFFNN, in modeling Cloud Amount and Earth Skin Temperature in Balikpapan. Based on the training data results, the Hybrid Fs-DLFFNN model consistently demonstrates the best performance, as indicated by the lowest MSE, RMSE, and MAPE values for both variables. This suggests that the hybrid approach is more flexible in capturing data patterns through the combination of Fourier series components and neural network modeling.
However, on the testing data, the performance of the hybrid model is not consistently good. For the Cloud Amount variable, the DLFFNN (4,3) method yields the best performance, with lower MSE, RMSE, and MAPE compared to both Fs and Hybrid Fs-DLFFNN (5,3). Meanwhile, for Earth Skin Temperature, the DLFFNN (4,3) method provides the most optimal results, with lower error values than the other models. In contrast, the hybrid model produces higher prediction errors on the testing data for both variables.

Table 4. Modeling Earth Skin Temperature and Cloud Amount for Samarinda

Variable	Method	Training			Testing		
		MSE	RMSE	MAPE	MSE	RMSE	MAPE
Cloud Amount	Fs	62.35	7.90	7.92	262.55	16.20	16.46
	DLFFNN (6,9)	48.52	6.96	6.35	79.71	8.92	8.65
	Hybrid Fs-DLFFNN (6,4)	**40.44**	**6.36**	**6.32**	203.17	14.25	13.66
Earth Skin Temperature	Fs	0.31	0.56	1.48	0.36	0.59	1.88
	DLFFNN (9,5)	0.15	0.38	0.99	0.34	0.58	1.42
	Hybrid Fs-DLFFNN (6,8)	**0.13**	**0.36**	**0.95**	0.23	0.48	1.44

Noted: DLFFNN (*a,b*) indicates a network architecture with *a* neurons in the first hidden layer and *b* neurons in the second hidden layer.

Table 4 presents a comparison of the performance of three modeling approaches, namely Fs, DLFFNN, and Hybrid Fs-DLFFNN, in modeling Cloud Amount and Earth Skin Temperature in Samarinda. Based on the training data, the Hybrid Fs-DLFFNN model again demonstrates the best performance, as indicated by the lowest MSE, RMSE, and MAPE values for both variables. This finding suggests that the hybrid approach has a high degree of flexibility in capturing data patterns during the training phase.

However, in contrast to the training results, the performance on the testing data exhibits more varied patterns. For the Cloud Amount variable, the DLFFNN (6,9) method provides the best results, with an MSE of 79.71, which is substantially lower than that of Fs (262.55) and Hybrid Fs-DLFFNN (203.17). This indicates that DLFFNN (6,9) has better generalization capability for this variable. Meanwhile, for Earth Skin Temperature, the Hybrid Fs-DLFFNN model achieves the best performance on the testing data, with an MSE of 0.23, RMSE of 0.48, and MAPE of 1.44, outperforming both Fs and DLFFNN.

Consistent findings from Table 3 and Table 4 show that the Hybrid Fs-DLFFNN model systematically delivers the best performance on the training data across both variables and locations. This indicates that the hybrid approach is more capable of capturing complex data characteristics, particularly through the integration of periodic components (Fourier series) and nonlinear patterns (DLFFNN). Therefore, this model can be viewed as a methodological innovation that enhances flexibility and explanatory power in representing data structures. Although the testing performance varies across cases, the overall pattern suggests that the hybrid model provides a more comprehensive modeling capacity compared to single-method approaches. In this context, the discrepancy between training and testing performance does not necessarily indicate a limitation, but rather highlights opportunities for further improvement, particularly in enhancing model generalization.

Although the Hybrid Fs-DLFFNN model showed indications of overfitting and did not consistently outperform the competing models on the testing data, it remains the primary focus of this study because the objective is to investigate the potential of integrating Fourier Series and DLFFNN within a hybrid framework. Therefore, the forecasting results are presented as an

evaluation of the proposed hybrid methodology rather than as evidence of superior predictive performance. Furthermore, the comparison between actual and predicted values on the training data is illustrated in Figures 13 to 16.

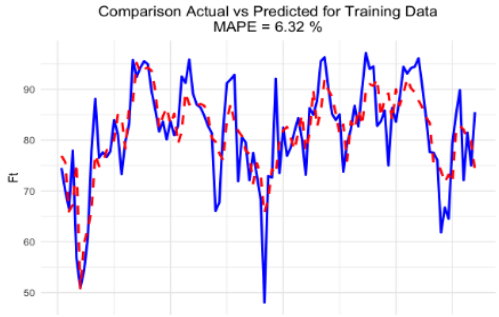


Figure 13. Comparison Actual vs Predicted cloud amount in Samarinda using the Hybrid Fs-DLFFNN model

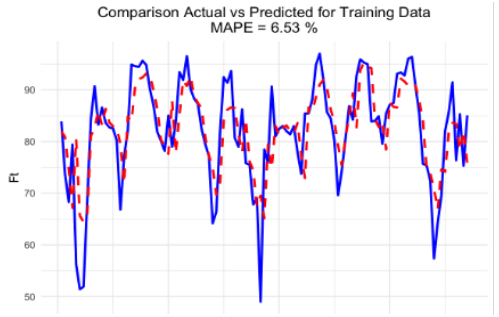


Figure 14. Comparison Actual vs Predicted cloud amount in Balikpapan using the Hybrid Fs-DLFFNN model

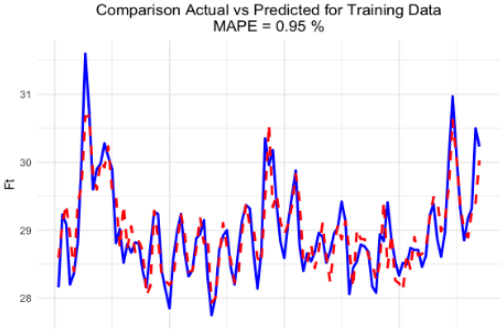


Figure 15. Comparison Actual vs Predicted earth skin temperature in Samarinda using the Hybrid Fs-DLFFNN model

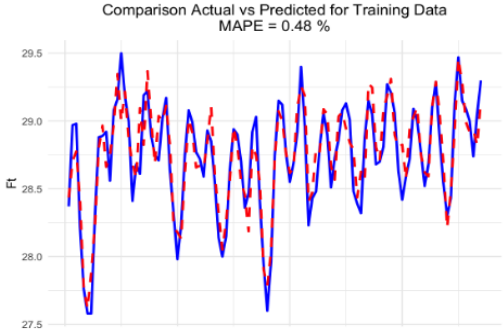


Figure 16. Comparison Actual vs Predicted earth skin temperature in Balikpapan using the Hybrid Fs-DLFFNN model

Figures 13 to 16 present a comparison between the actual and predicted values on the training data using the Hybrid Fs-DLFFNN model for Cloud Amount and Earth Skin Temperature in Samarinda and Balikpapan. In general, the predicted values (dashed lines) closely follow the actual data (solid lines), effectively capturing both short-term fluctuations and underlying periodic patterns. Overall, these visualizations reinforce previous findings that the Hybrid Fs-DLFFNN model has excellent fitting capability. However, it is important to note that strong performance on the training data does not necessarily guarantee that the hybrid model will consistently outperform single models on the testing data. This phenomenon is evident in the earlier results, where single models such as Fs or DLFFNN occasionally provide better performance on testing data. This condition indicates a potential overfitting issue, where the model fits the training data too closely, leading to limited generalization ability. Therefore, although the Hybrid Fs-DLFFNN model demonstrates strong capability in capturing data patterns, model evaluation should carefully consider the balance between fitting and generalization. This is essential to ensure that the selected model is not only optimal on training data but also reliable and robust in predicting new data.

3.3. Best Model Forecasting and Discussion

In this section, the Hybrid Fs-DLFFNN model is retained for forecasting purposes because the primary objective of this study is methodological development and evaluation of the proposed hybrid framework. This combined framework enhances the model's capability to capture complex and non-stationary data characteristics. Accordingly, despite not consistently outperforming single models in all testing scenarios, the Hybrid Fs-DLFFNN is selected for forecasting, given its flexibility and potential for further methodological development.

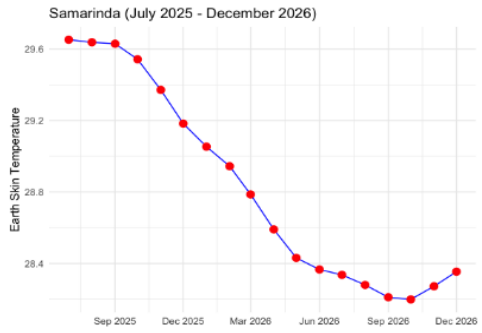


Figure 17. 18-period ahead forecast of earth skin temperature in Samarinda

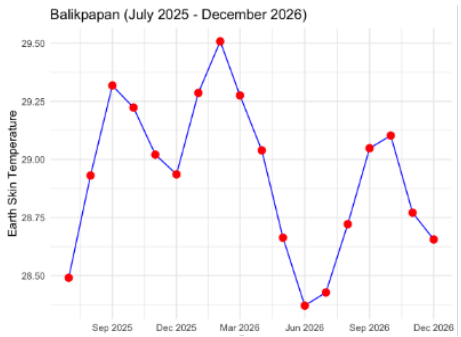


Figure 18. 18-period ahead forecast of earth skin temperature in Balikpapan

Based on Figure 17 and Figure 18, the earth skin temperature forecast graph for Balikpapan and Samarinda shows the predicted results for the next periods. In Balikpapan, the model predicts temperature fluctuations with a significant peak in March 2026 and a sharp decline thereafter. Meanwhile, in Samarinda, temperatures are predicted to decrease drastically from September 2025 to December 2026, indicating a more consistent downward trend throughout the forecast period.

Cloud Amount forecasts for Samarinda and Balikpapan show different trends in Figure 19 and Figure 20. In Samarinda, there is a sharp increase in cloud amount from September 2025 to March 2026, followed by a steadier decline until December 2026. Meanwhile, in Balikpapan, the pattern is a sharp decline from September 2025 to March 2026, followed by a slower recovery thereafter.

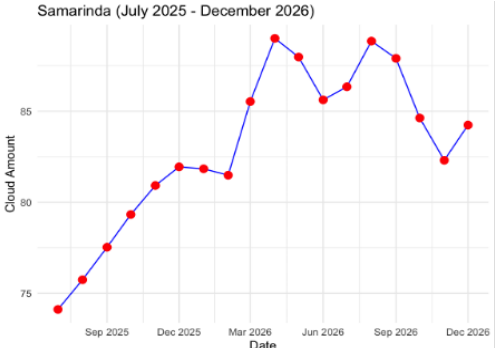


Figure 19. 18-period ahead forecast of cloud amount in Samarinda

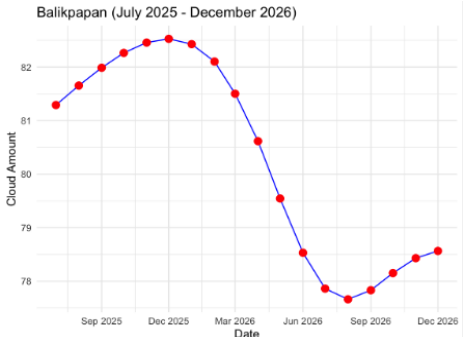


Figure 20. 18-period ahead forecast of cloud amount in Samarinda

The forecast results for earth skin temperature and cloud amount in Samarinda indicate a significant relationship between the two variables. The predicted decrease in land surface temperature from September 2025 to December 2026 is followed by a sharp increase in Cloud Amount until March 2026, before declining steadily. This phenomenon may indicate increased atmospheric humidity associated with decreased land surface temperature, potentially leading to increased water vapor condensation and cloud amount. This relationship suggests potential changes in rainfall patterns that could influence disasters such as flooding. Therefore, mitigation policies in Samarinda should focus on climate-based natural resource management, including efficient land-use planning and water management.

On the other hand, Balikpapan shows a different trend, where the decrease in Earth skin temperature is followed by a sharp reduction in Cloud Amount during the same period, potentially reducing atmospheric humidity and increasing the risk of drought. This decrease in cloud amount can impact rainfall patterns and water availability, indicating a potential risk of future drought. Therefore, mitigation policies in Balikpapan should focus on efficient water management, water

conservation, and strengthening early warning systems to anticipate the impact of drought on the agricultural sector and water availability.

4. CONCLUSION

The modeling of Earth Skin Temperature and Cloud Amount using Fourier Series (Fs), Double Layer Feed Forward Neural Network (DLFFNN), and Hybrid Fs-DLFFNN models has been successfully implemented for Balikpapan and Samarinda. The results reveal that the Hybrid Fs-DLFFNN model consistently demonstrates good performance in the training phase, indicated by lower MSE, RMSE, and MAPE values. This highlights the model's strength in capturing both periodic and nonlinear patterns through the integration of Fourier series and neural network approaches. However, the evaluation on testing data shows that the superiority of the hybrid model is not consistently observed across all cases. In several instances, single models such as Fs or DLFFNN provide more competitive or even better predictive performance. This phenomenon suggests the presence of varying data characteristics and indicates that model performance is highly dependent on the balance between fitting capability and generalization.

Despite these variations, the Hybrid Fs-DLFFNN model remains an important methodological contribution, particularly in its ability to represent complex environmental dynamics. The differences observed between training and testing performance, as well as across variables and locations, constitute interesting findings that warrant further investigation. These phenomena open opportunities for deeper exploration, such as improving model generalization, refining hybrid architectures, and incorporating additional regularization or validation strategies.

Forecasting results for the next 18 periods indicate distinct temporal patterns in both cities, reflecting different environmental dynamics. Rather than making definitive claims, these projections provide indicative insights that can support further analysis and discussion. In this context, the study contributes to the broader understanding of environmental modeling and offers a foundation for developing more robust predictive frameworks, which are relevant to climate-related studies and aligned with the objectives of SDG 13 (Climate Action).

ACKNOWLEDGEMENTS

Authors gratefully acknowledge the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan/LPDP), under the Ministry of Finance of the Republic of Indonesia, for the financial support provided through the Doctoral Program Scholarship, as specified in Decree No. SKPB-10349/LPDP/LPDP.3/2024.

REFERENCES

[1] M. Akhtaruzzaman, S. Boubaker, P. Gamage, and V. Obeng, "Environmental innovation and climate risk awareness: The moderating role of SDG13," *Res Int Bus Finance*, vol. 75, Mar. 2025, doi: 10.1016/j.ribaf.2024.102713.

[2] K. E. Trenberth and D. J. Shea, "Relationships between precipitation and surface temperature," *Geophys Res Lett*, vol. 32, no. 14, pp. 1–4, Jul. 2005, doi: 10.1029/2005GL022760.

[3] J. A. Sobrino, Y. Julien, and S. García-Monteiro, "Surface temperature of the planet earth from satellite data," *Remote Sens (Basel)*, vol. 12, no. 2, Jan. 2020, doi: 10.3390/rs12020218.

[4] H. Iftikhar, M. Khan, J. Żywiołek, M. Khan, and J. Linkolk López-Gonzales, "Modeling and forecasting carbon dioxide emission in Pakistan using a hybrid combination of regression and time series models," *Heliyon*, vol. 10, no. 13, Jul. 2024, doi: 10.1016/j.heliyon.2024.e33148.

[5] A. Radovan, V. Šunde, D. Kučak, and Ž. Ban, "Solar irradiance forecast based on cloud movement prediction," *Energies (Basel)*, vol. 14, no. 13, Jul. 2021, doi: 10.3390/en14133775.

[6] S. Sam and M. Özger, "A multi-method and multi-duration trend analysis of temperature and precipitation in Istanbul, Turkey, by using meteorological records, MERRA-2 reanalysis, and IMERG estimations," *HydroResearch*, vol. 8, pp. 209–222, Jan. 2025, doi: 10.1016/j.hydres.2024.11.005.

[7] M. T. Nakkazi, J. I. Sempewo, M. D. Tumutungire, and J. Byakatonda, "Performance evaluation of CFSR, MERRA-2 and TRMM3B42 data sets in simulating river discharge of data-scarce tropical catchments: a case study of Manafwa, Uganda," *Journal of Water and Climate Change*, vol. 13, no. 2, pp. 522–541, Feb. 2022, doi: 10.2166/wcc.2021.174.

[8] A. T. R. Dani, N. Chamidah, I. N. Budiantara, B. Lestari, R. Kusuma, and D. Aydin, "Semiparametric Time Series Regression Using Exponential Complex Fourier Series for Temperature Prediction in the Tropical Rainforest of Samarinda," *Innovaciencia*, vol. 13, no. 1, Dec. 2025, doi: 10.15649/2346075x.5364.

[9] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed. New York: Pearson Education, Inc, 2006.

[10] T. W. Kim and M. Jehanzaib, "Drought risk analysis, forecasting and assessment under climate change," Jul. 01, 2020, *MDPI AG*. doi: 10.3390/W12071862.

[11] B. Billah, M. L. King, R. D. Snyder, and A. B. Koehler, "Exponential smoothing model selection for forecasting," *Int J Forecast*, vol. 22, no. 2, pp. 239–247, Apr. 2006, doi: 10.1016/j.ijforecast.2005.08.002.

[12] J. Yu, S. B. Kim, J. Bai, and S. W. Han, "Comparative study on exponentially weighted moving average approaches for the self-starting forecasting," *Applied Sciences (Switzerland)*, vol. 10, no. 20, pp. 1–18, Oct. 2020, doi: 10.3390/app10207351.

[13] D. Zhao, R. Zhang, H. Zhang, and S. He, "Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models," *Sci Rep*, vol. 12, no. 1, pp. 1–13, 2022, doi: 10.1038/s41598-022-23154-4.

[14] Suhartono, E. Setyowati, N. A. Salehah, M. H. Lee, S. P. Rahayu, and B. S. S. Ulama, "A hybrid singular spectrum analysis and neural networks for forecasting inflow and outflow currency of bank Indonesia," in *Communications in Computer and Information Science*, Springer Verlag, 2019, pp. 3–18. doi: 10.1007/978-981-13-3441-2_1.

[15] M. Bilodeau, "Fourier smoother and additive models," *The Canadian Journal of Statistics*, vol. 20, no. 3, pp. 257–269, 1992.

[16] D. Aydin, "A Comparative Study of Neural Networks and Non-Parametric Regression Models for Trend and Seasonal Time Series," *Trends Appl Sci Res*, vol. 4, no. 3, pp. 126–137, Mar. 2009, doi: 10.3923/tasr.2009.126.137.

[17] N. Chamidah, S. D. Febriana, R. A. Ariyanto, and R. Sahawaly, "Fourier series estimator for predicting international market price of white sugar," in *AIP Conference Proceedings*, American Institute of Physics Inc., Feb. 2021. doi: 10.1063/5.0042287.

[18] N. Afifah, I. N. Budiantara, and I. N. Latra, "Mixed Estimator of Kernel and Fourier Series in Semiparametric Regression," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2017. doi: 10.1088/1742-6596/855/1/012002.

[19] I. E. Livieris, E. Pintelas, S. Stavroyiannis, and P. Pintelas, "Ensemble Deep learning models for forecasting cryptocurrency time-series," *Algorithms*, vol. 13, no. 5, May 2020, doi: 10.3390/A13050121.

[20] A. T. R. Dani and F. B. Putra, "Time Series Modeling with Intervention Analysis to Evaluate of COVID-19 Impact on the Stock Markets in Indonesia and Global," *Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 1, pp. 113–126, Apr. 2025, doi: 10.37905/euler.v13i1.31081.

[21] R. Recylia, T. Saifudin, and N. Chamidah, "Forecasting the Number of Search and Rescue Operations for Ship Accidents in Indonesia using Fourier Series Analysis," *BAREKENG: Journal of Mathematics and Its Applications*, vol. 19, no. 2, pp. 1023–1036, Jun. 2025, doi: 10.30598/barekengvol19iss2pp1023-1036.

[22] A. K. Rahma, Q. Z. Abidin, J. K. Prasetyo, B. Larasati, D. Amelia, and N. Chamidah, "Prediction of Crude Oil Prices in Indonesia using Fourier Series Estimator and ARIMA Method," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 3, pp. 1673–1682, Jul. 2024, doi: 10.30598/barekengvol18iss3pp1673-1682.

[23] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Information*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/info15090517.

[24] Suhartono, D. E. Ashari, D. D. Prastyo, H. Kuswanto, and M. H. Lee, "Deep neural network for forecasting inflow and outflow in Indonesia," *Sains Malays*, vol. 48, no. 8, pp. 1787–1798, 2019, doi: 10.17576/jsm-2019-4808-26.

[25] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," in *European Journal of Operational Research*, Jan. 2005, pp. 501–514. doi: 10.1016/j.ejor.2003.08.037.