BAREKENG: Journal of Mathematics and Its Applications
September 2026 Volume 20 Issue 3 Page 2045-2062
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss3pp2045-2062

FORECASTING SEA LEVEL CHANGES USING HYBRID ARIMA-RADIAL BASIS FUNCTION NEURAL NETWORK METHODS

Soehardjoepri [1*], Ulil Azmi [2], Prilyandari Dina Saputri [3], Moch. Taufik Hakiki [4], Denisha A. E. Ananda [5], Roslinazairimah Zakaria [6]

[1,2,3,4,5]*Department of Actuarial Science, Faculty of Science and Analytical Data, Institut Teknologi Sepuluh Nopember*
Jln. Raya ITS, Surabaya, 60111, Indonesia

[6]*Center for Mathematical Sciences, Pusat Sains Matematik, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan Pahang, Malaysia*

*Corresponding author's e-mail: * djoepri.its@gmail.com*

Article Info

Article History:

Received: 17th July 2025
Revised: 22nd January 2026
Accepted: 16th March 2026
Available online: 8th April 2026

Keywords:

ARIMA;
Climate change;
Marina Ancol beach;
Radial basis function neural network;
Sea level.

ABSTRACT

Understanding sea level variability is crucial for ensuring the safety of tourists, particularly in marine tourism areas like Marina Ancol Beach in North Jakarta. Climate change has led to rising sea levels, significantly impacting coastal regions. Accurate predictions of sea level are essential for anticipating tidal flooding, which occurs when seawater inundates these areas. Short-term sea level fluctuations are influenced by both linear tidal patterns and nonlinear local effects, making accurate forecasting challenging when using a single modeling approach. This study proposes a hybrid forecasting method that combines the Autoregressive Integrated Moving Average (ARIMA) model to capture linear temporal structures and a Radial Basis Function Neural Network (RBFNN) to model nonlinear patterns present in the residuals. Hourly sea level data consisting of 17,520 observations collected from January 2021 to December 2022 were analyzed. The proposed hybrid ARIMA–RBFNN model achieved a Mean Absolute Percentage Error (MAPE) of 2.74%, slightly outperforming the ARIMA model, which yielded a MAPE of 2.76%. The model provides accurate 24-hour sea level forecasts for Marina Ancol Beach, offering timely information that can support local authorities in anticipating and mitigating tidal flooding events.

How to cite this article:

Soehardjoepri, U. Azmi, P. D. Saputri, M. T. Hakiki, D. A. E. Ananda and R. Zakaria., "FORECASTING SEA LEVEL CHANGES USING HYBRID ARIMA-RADIAL BASIS FUNCTION NEURAL NETWORK METHODS", *BAREKENG: J. Math. & App.*, vol. 20, no. 3, pp. 2045-2062, Sep, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

As climate change accelerates, rising sea levels pose significant challenges for coastal regions worldwide, particularly in areas reliant on marine tourism. Marina Ancol Beach in North Jakarta, Indonesia, serves as a prime example, attracting numerous visitors while facing the threat of tidal flooding due to increased sea levels. This phenomenon not only endangers the coastal ecosystem but also impacts the economic viability of tourism-dependent communities [1], [2]. According to the Department of Water Resources of DKI Jakarta Province, sea levels are categorized into four levels, with Alert Level 1 representing the most dangerous conditions and Alert Level 4 indicating normal levels [3]. This classification is crucial for understanding the potential risks and impacts on coastal ecosystems and tourism-dependent communities.

Sea level is used in storm surge forecasting and can provide a clear signal to increase the warning status in disaster mitigation, thereby strengthening and supporting disaster emergency management in coastal areas [4]. Understanding sea level dynamics is crucial for effective coastal management and disaster preparedness. Accurate predictions can help mitigate risks associated with tidal flooding, allowing local governments and stakeholders to implement timely interventions. The integration of advanced modeling techniques, such as ARIMA and Radial Basis Function Neural Networks (RBFNN), offers promising avenues for enhancing prediction accuracy [5]. ARIMA has been used in previous research [6], [7], for example sea level prediction, yielding predictions that are quite accurate and can serve as a reference for short- and medium-term disaster prevention [8]. Research comparing ARIMA with exponential smoothing state space models (ESMs) shows that ARIMA is more effective in predicting short-term sea level rise in the Arabian Sea for the next 3 years, every month [9]. There is research that shows that RBFNN modeling has very good performance in predicting heavy metal concentrations in lakes, especially in tropical and dry environments, and is able to identify the most correlated input parameters in polluted and unpolluted lake scenarios [10]. A study also proves that the RBFNN method is an alternative that is able to predict the inflation rate in Batam City effectively and efficiently [11].

Previous studies on sea level rise have employed a variety of methods. Setiawan (2016) utilized a Support Vector Machine approach optimized by Particle Swarm Optimization to analyze water level data in the Marabahan area of Barito Kuala Regency, South Kalimantan, covering the period from April 2008 to December 2012. This study found a mean prediction error, indicated by a Root Mean Square Error (RMSE) value of 37.685 [12].

Alsuwaylimi (2023) conducted a study to forecast monthly gold prices using ARIMA, Artificial Neural Networks (ANN) and hybrid models. The findings indicated that the best model for monthly gold prices forecasting was the ARIMA-ANN hybrid model [13]. In a related study, Baluk (2020) applied a hybrid approach to predict sea wave levels in Pekalongan, Rembang, and Semarang. This research utilized a Hybrid Vector Autoregressive-Radial Basis Function Neural Network (VAR-RBFNN) method, demonstrating that the dataset exhibited non-linearity, which the RBFNN effectively addressed [14]. Additionally, Soebroto (2015) employed Time Variant Inertia Weight Particle Swarm Optimization (TVIWPSO) in conjunction with Support Vector Regression (SVR) to predict river water levels. The results revealed a Mean Absolute Percentage Error (MAPE) of 0.00755 using the SVR-TVWPO method [15].

While previous studies have reported that hybrid forecasting methods can improve predictive performance, their application to high-frequency sea level forecasting remains limited, particularly for operational coastal management. Unlike most existing studies that rely on daily or monthly data and focus on long-term trends, this study introduces a residual-based hybrid ARIMA–Radial Basis Function Neural Network (RBFNN) framework designed for short-term, hourly sea level forecasting. The model explicitly decomposes the sea level time series into linear components captured by ARIMA and nonlinear dynamics learned by RBFNN, enhancing its ability to represent complex local fluctuations. Using 17,520 hourly observations from the Marina Ancol Floodgate between January 2021 and December 2022, this study provides one of the first high-resolution forecasting analyses for Marina Ancol Beach. The resulting 24-hour forecasts offer operationally relevant and actionable information to support local government decision-making and improve coastal safety for residents and tourists under increasingly dynamic coastal conditions.

2. RESEARCH METHODS

2.1 Data Source and Research Variable

This research utilizes secondary data obtained from the Jakarta Water Resources Agency. The primary variable analyzed in this study is sea level, as detailed in Table 1.

Table 1. Research Variable

Variable	Description	Scale
Y	Sea level	Ratio
X_1	Sea level at 1 hour before observation (t-1)	Ratio
X_2	Sea level at 2 hours before observation (t-2)	Ratio
X_3	Sea level at 3 hours before observation (t-3)	Ratio
⋮	⋮	⋮
Xn	Sea level at n hours before observation (t-n)	Ratio

The unit of observation for this research is the hourly sea level data from the Marina Ancol floodgate, covering the period from January 1, 2021, to December 31, 2022. In total, 17,520 observations were collected, with data updated every hour throughout the year. The structure of the data is presented in Table 2.

Table 2. Data Structure of Research Variables

th-observation	date	time	Y	Sea level (cm)				
				X_1	X_2	X_3	…	X_n
1	1 January 2021	00.00	$Yn+1$	$X_{1,1}$	$X_{2,1}$	$X_{3,1}$	…	$X_{n,1}$
2	1 January 2021	01.00	$Yn+_2$	$X_{1,2}$	$X_{2,2}$	$X_{3,2}$	…	$X_{n,2}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
17.520	31 December 2022	23.00	$Y_{17.520}$	$X_{1,17.520}$	$X_{2,17.520}$	$X_{3,17.520}$	…	$X_{n,17.520}$

Table 2 describes the structure of the sea level time-series data used in this study. Each observation represents an hourly sea level measurement. For example, $X_{1,1}$ denotes the sea level recorded at 00:00 on 1 January 2021, while Y_2 represents the sea level at the next hour (01:00 on the same date).

2.2 Water Level and Marina Ancol Flood Gate

Floodgates are used to manage water flow and control the flow of sea, river, and dam water, allowing for effective management during sudden increases in water levels. The Jakarta Water Resources Agency (2023) states that floodgates are vital for regulating water flow. Observing water levels helps determine the best times to open or close the floodgates [3].

In DKI Jakarta, several floodgates exist, including the Marina Ancol Floodgate, located in North Jakarta near the popular tourist destination Taman Impian Jaya Ancol. The SDA categorizes water levels at the floodgates into four groups, with Alert Level 4 representing normal levels and Alert Level 1 indicating the highest levels. Detailed information about these categories is provided in Table 3.

Table 3. Water Level Category at Marina Jaya Ancol Floodgate

Water Level Category	Status	Water level (cm)
Alert Level 1	Danger	>250
Alert Level 2	Alert	200 – 250
Alert Level 3	Watch	170 – 200
Alert Level 4	Normal	<170

2.3 Time Series Analysis

Time series data consists of observations where each variable is arranged chronologically. Researchers often use time series data to identify historical patterns, which can then be applied to forecast future outcomes. When analyzing time series patterns, it is helpful to consider several key components, including trends, cycles, seasonal variations, and irregular fluctuations [16].

When studying time series data, researchers typically aim for two main outcomes: to understand or model the underlying stochastic processes that generate the observed data and to predict future values based on past observations [17].

2.3.1 Autoregressive Integrated Moving Average (ARIMA)

ARIMA, also known as the Box-Jenkins method, is a forecasting technique developed by George Edward Pelham Box and Gwilym Meirion Jenkins, and it is widely used for analyzing time series data. The ARIMA Box-Jenkins analysis involves five key steps: model identification, parameter estimation, significance testing of parameters, model diagnostic testing, and selection of the best model [18].

The ARIMA (p, d, q) model can generally be expressed in the following Eq. (1) [19]:

$$\phi_p(B)(1-B)^d Z_t = \mu + \theta_q(B)a_t, \tag{1}$$

where,

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_P B^P$,
$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$,

with B is the backshift operator $B^j x_t = x_{t-j}$.

2.3.2 Radial Basis Function Neural Network (RBFNN)

Radial Basis Function Neural Network (RBFNN) is a type of Artificial Neural Network (ANN) characterized as a feedforward network capable of short-term forecasting [20]. Among various activation functions, the Gaussian function is the most commonly used due to its local properties and its ability to measure the proximity of inputs to the mean (or cluster center). RBFNN consists of a multi-layer architecture, which includes three distinct layers: an input layer, a hidden layer that employs a non-linear Gaussian activation function, and a linear output layer. For this research, the architecture of the RBFNN with three input neurons and four output neurons is illustrated in Figure 1 below.

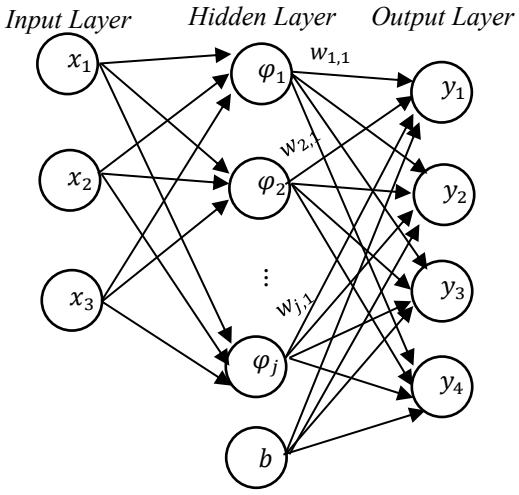


Figure 1. Architecture of Radial Basis Function Neural Network

To implement the RBFNN, the first step is to initialize the centroid values, which can be done through trial and error. Following this, the spread value (neuron width) is calculated using the formula below.

$$\sigma_j = \frac{d_{max}}{\sqrt{2K}}, \tag{2}$$

where σ_j is the spread value, d_{max} is the maximum distance between each neuron center, and K is the total number of neurons formed. Next, we calculate the Gaussian activation function using the formula below.

$$\varphi_{ij} = \exp\left(-\frac{\|x_i - c_j\|^2}{2{\sigma_j}^2}\right), \tag{3}$$

where φ_{ij} the *Gaussian* activation function of *j-th hidden neuron*, x_i is the *i-th* input data, c_j is the *centroid* value of the *j-th hidden neuron,* $\|x_i - c_j\|^2$ is the distance between the *i-th* data point and the centroid of the

j-th hidden layer, and $\sigma_j{}^2$ is the spread value of the *j-th* hidden neuron. Once the calculations are complete, we will construct the Gaussian matrix, represented by the notation shown below.

$$G = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1j} & b \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2j} & b \\ \vdots & \vdots & \ddots & \vdots & b \\ \varphi_{i1} & \varphi_{i2} & \cdots & \varphi_{ij} & b \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_j \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_j \end{bmatrix}, \tag{4}$$

where G is the *Gaussian* matrix, j is the index of the hidden neuron, and b is the bias, with the bias value set to 1. Gaussian Matrix can be expressed as follows:

$$G \cdot w = d, \tag{5}$$

where,

$$d = [d_1 \quad d_2 \quad \cdots \quad d_j]^T, \tag{6}$$

$$w = [w_1 \quad w_2 \quad \cdots \quad w_j]^T, \tag{7}$$

$$G = \varphi_{ij}. \tag{8}$$

The next step involves updating the weight values and the bias value using the Gaussian matrix calculation, as illustrated in formula below.

$$w = (G^T G)^{-1} G^T d, \tag{9}$$

where w is the weight value matrix from the hidden neurons to the output neurons, G^T is the transpose of the Gaussian matrix, and $(G^T G)^{-1}$ is the inverse of the product of G^T and G, with d representing the target matrix. The next step is to calculate the output using formula below.

$$y_k = \sum_{k=1}^{j} \varphi_{ij} w_{jk} + b_k\,, \tag{10}$$

where y_k is the value of the *k-th* output, w_{jk} is the weight from the *j-th* hidden neuron to the *k-th* output neuron, φ_{ij} is the *i-th* Gaussian activation function for the *j-th* hidden neuron, and b_k is the bias for the *k-th* output neuron. The final step is to determine the class of each data point based on the output index that has the maximum y_k.

2.3.3 Hybrid ARIMA and Radial Basis Function Neural Network Model (ARIMA-RBFNN)

The ARIMA-RBFNN model combines the strengths of Autoregressive Integrated Moving Average (ARIMA) for time series analysis and Radial Basis Function Neural Networks (RBFNN) for capturing non-linear relationships in the data. This hybrid approach aims to improve forecasting accuracy by leveraging the linear modeling capabilities of ARIMA alongside the flexibility of RBFNN to model complex patterns.

To obtain a hybrid model from ARIMA and an Artificial Neural Network, the results from both models must be summed . The ARIMA model is constructed using the observation data, while the Artificial Neural Network model is developed based on the residuals (errors) from the ARIMA model. The relationship can be expressed as:

$$\hat{Y}_t = L_t + N_t, \tag{11}$$

where Lt is the fitted value or prediction from the ARIMA model and N_t is the fitted value or prediction from the Artificial Neural Network.

The process typically involves the following steps:

1. **Model Identification**: Identify the appropriate ARIMA parameters (*p, d, q*) for the time series data.
2. **ARIMA Forecasting**: Use the ARIMA model to generate initial forecasts.
3. **Residual Analysis**: Analyze the residuals from the ARIMA model to identify non-linear patterns that the ARIMA model may not have captured.
4. **RBFNN Training**: Train the RBFNN using the residuals as the target variable, employing the inputs from the original time series data.

5. **Final Forecasting**: Combine the forecasts from the ARIMA model with the outputs from the RBFNN to produce the final predictions.

This hybrid model is particularly useful in scenarios where time series data exhibits both linear trends and complex non-linear behaviors, enhancing the overall forecasting performance.

2.3.4 Evaluation of Forecasting

When evaluating the predictive or forecasting ability of the developed model, it is essential to calculate the model's accuracy. Common methods for measuring model performance include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE).

1. MAE indicates the average absolute difference between the actual and predicted values.

$$MAE = \frac{1}{m}\sum_{t=1}^{m}|R_t - P_t|. \tag{12}$$

2. MAPE provides a percentage-based measure of accuracy by calculating the error percentage, offering a more intuitive understanding of the model's performance.

$$MAPE = \frac{1}{m}\sum_{t=1}^{m}\left|\frac{R_t - P_t}{R_t}\right|. \tag{13}$$

3. RMSE quantifies the average squared error, giving greater weight to larger errors in the predictions.

$$RMSE = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(R_t - P_t)^2}. \tag{14}$$

where m represents the total number of data points, R_t is the actual value, and P_t is the predicted value [22].

3. RESULTS AND DISCUSSION

This research focuses on forecasting sea levels at the Marina Ancol floodgate in Jakarta using conventional forecasting methods, specifically Autoregressive Integrated Moving Average (ARIMA). It also compares the Radial Basis Function Neural Network (RBFNN) method with the hybrid ARIMA-RBFNN approach. Before delving into the forecasting process, a descriptive analysis is conducted to characterize the data. This analysis is essential for accurately comparing the distributions of the training and testing datasets, ensuring reliable results for forecasting the upcoming period.

3.1 Characteristic of the Sea Level

After completing the data preprocessing, the next step is to perform a descriptive analysis of the sea level data. The plot below illustrates the sea levels at Marina Ancol from January 2021 to December 2022, presented on an hourly basis.

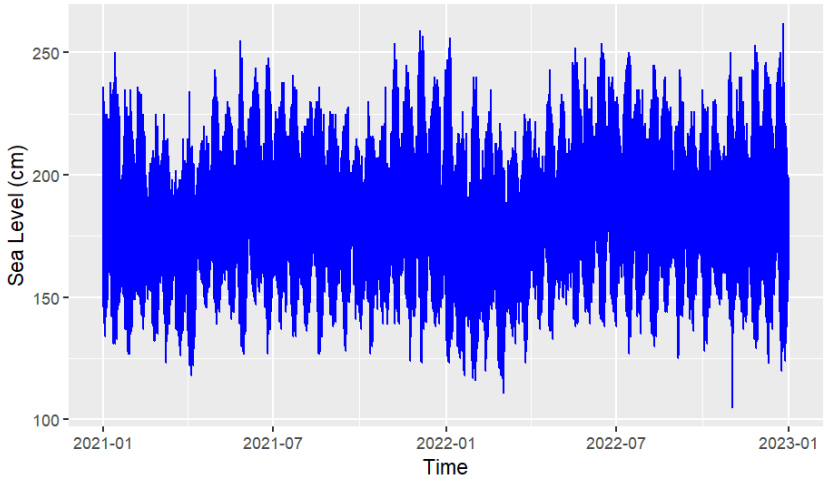


Figure 2. Time Series Plot of Hourly Sea Levels
(Source: Output R)

Figure 2 illustrates the overall data pattern, revealing a seasonal trend characterized by a repetitive up-and-down movement. To gain a clearer understanding of the distribution pattern, a more detailed plot is generated for the data from January 1 to January 13, 2021, as shown in Figure 3 below.

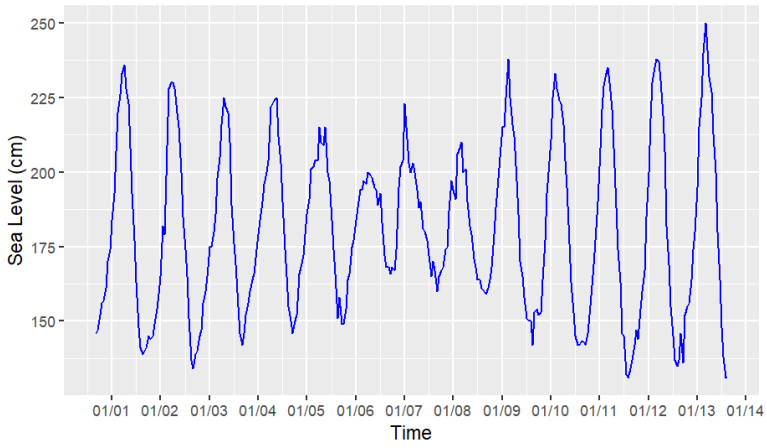


Figure 3. Time Series Plot of Sea Level for Period January 1st – January 13th, 2021
(Source: Output R)

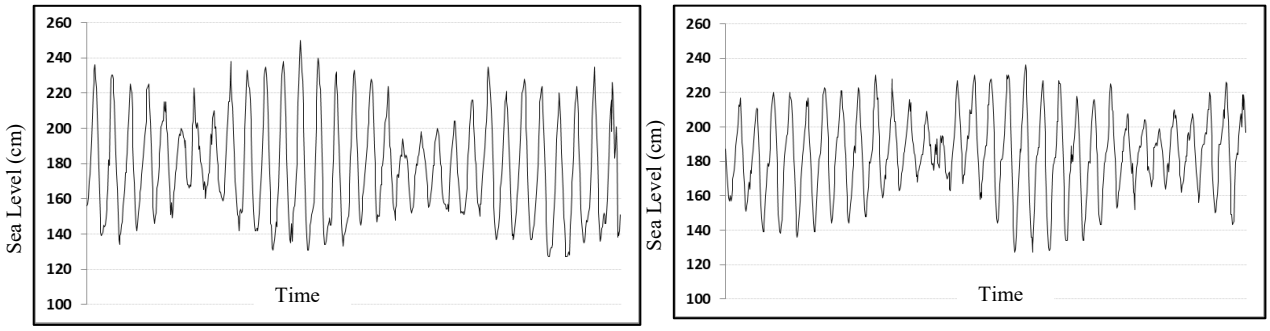
The seasonal pattern is clearly evident in Figure 3, which indicates that an upward trend typically occurs from 10:00 a.m. to 1:00 p.m., while a downward trend is observed during the nighttime. In addition to the time series plot, a summary of the data is provided through a descriptive analysis, as shown in Table 4 below.

Table 4. Descriptive Statistics of Sea Level 2021-2022

	Water Level (cm)
Minimum	105
1st *Quartile*	164
Median	183
3rd *Quartile*	202
Maximum	262
Mean	183

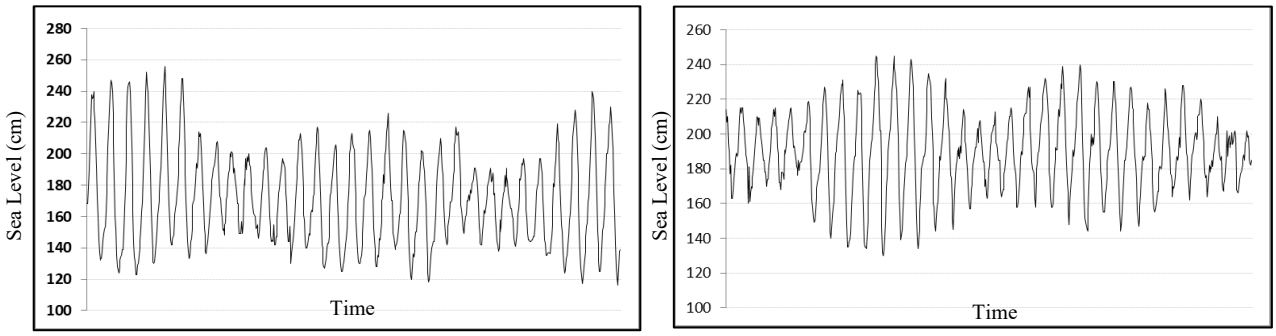
Table 4 shows that the mean sea level at the Marina Ancol floodgate for the period of 2021-2022 is 183 cm, which is categorized as Alert Level 3. The minimum sea level recorded was 105 cm (Alert Level 1), occurring on November 1, 2022, at 4 a.m., while the maximum sea level reached 262 cm on December 25, 2022, at 11 a.m.

Since the time series plot in Figure 2 does not effectively display the data pattern, a re-plotting has been conducted, dividing the data by month to enhance clarity. The results are shown in Figure 4 below.



(a) (b)

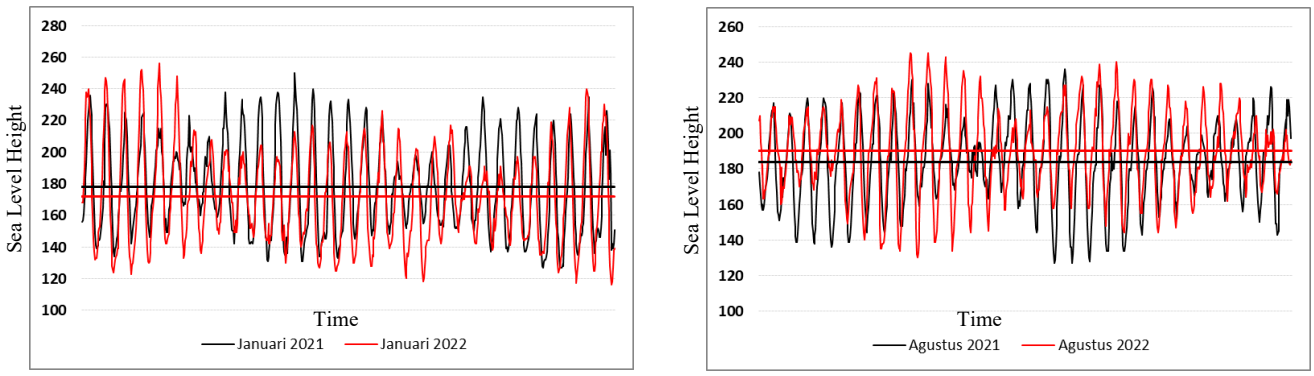
Figure 4. Time Series Plot of Sea Level at Marina Ancol Floodgate
(a) January 2021, (b) August 2021
(Source: Output R)



(a) (b)

Figure 5. Time Series Plot of Sea Level at Marina Ancol
(a) January 2022, (b) August 2022
(Source: Output R)

Since the overall time series plot in Fig. 2 does not clearly reveal short-term seasonal patterns, the data were re-plotted by month to improve interpretability. Figs. 4 and 5 present representative monthly sea level variations during the rainy season (January) and dry season (August) for two different years. These plots highlight consistent diurnal tidal cycles as well as seasonal differences in sea level amplitude. Specifically, January exhibits higher variability and wider ranges of sea level fluctuations compared to August, reflecting seasonal hydrodynamic conditions. The similarity of patterns across years indicates temporal stability in short-term sea level behavior, which supports the use of time-series-based forecasting models in this study. A comparison of the sea levels for January and August in both 2021 and 2022 is presented in Fig. 6.



(a) (b)

Figure 6. Time Series Plot of Sea Level at Marina Ancol
(a) January 2021 dan January 2022, (b) August 2021 dan August 2022
(Source: Output R)

Fig. 6 compares sea levels in January and August for the years 2021 and 2022. The black vertical line indicates the mean sea level for 2021, while the red vertical line represents the mean for 2022. In January, the mean sea level in 2021 is higher than that of 2022. Conversely, in August, the mean sea level for 2021 is lower than for 2022. The numerical data for sea levels can effectively represent the corresponding sea level categories. The details of the total data in each category are shown below.

Table 5. Detailed Total Data of Sea Level per Category

Water Level Category	Status	Total Data
Alert Level 1 (>250 cm)	Danger	35
Alert Level 2 (200-250 cm)	Alert	4640
Alert Level 3 (170-200 cm)	Watch	7012
Alert Level 4 (<170 cm)	Normal	5833

Table 5 indicates that the Alert Level 3 category was the most frequently observed over the past two years (2021-2022), occurring 7,012 times. In contrast, the Alert Level 1 category was the least common, with only 36 occurrences during this period. Figure 7 below provides a visual representation of the data presented in Table 5.

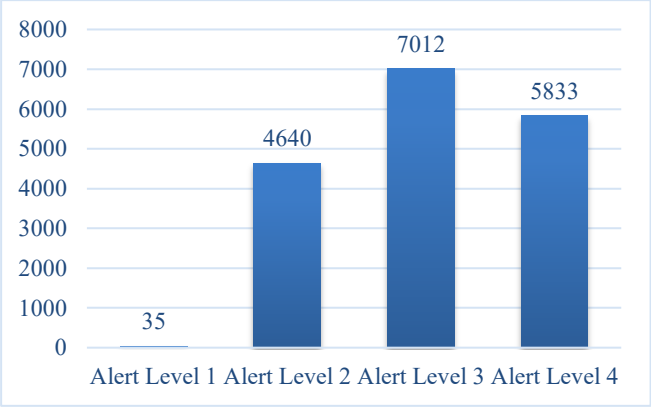


Figure 7. Column Chart Total Data per Sea Level Category

To visualize the distribution pattern by hour, a box plot has been created, and the results are shown in Figure 7 below.

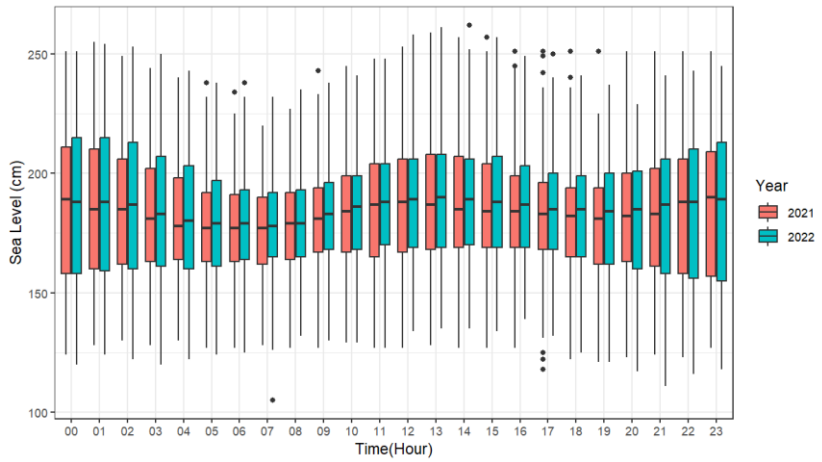


Figure 8. Box Plot Data Distribution per Hour
(Source: Output R)

Figure 8 illustrates that the sea level in 2022 is higher than in 2021, indicating a rise in sea levels over the course of one year. Additionally, the data shows that sea levels tend to be higher at noon compared to the morning, with a decrease occurring later in the day, followed by a subsequent rise at dawn. The monthly distribution of sea levels is presented in Figure 9.

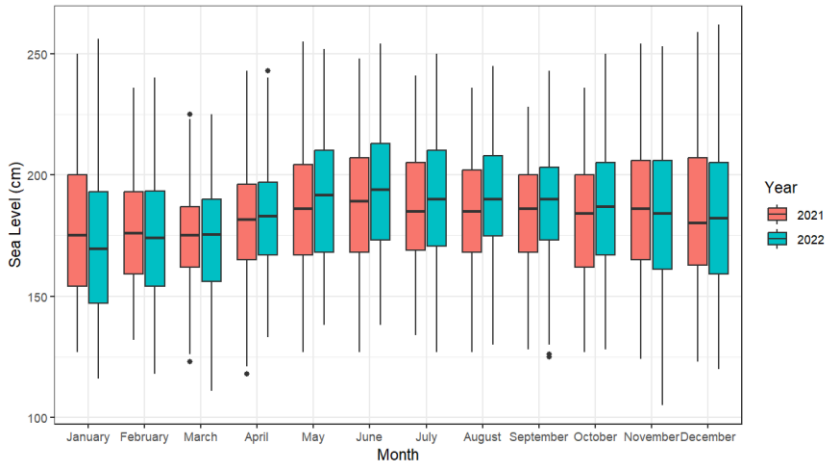


Figure 9. Box Plot of Data Distribution per Hour
(Source: Output R)

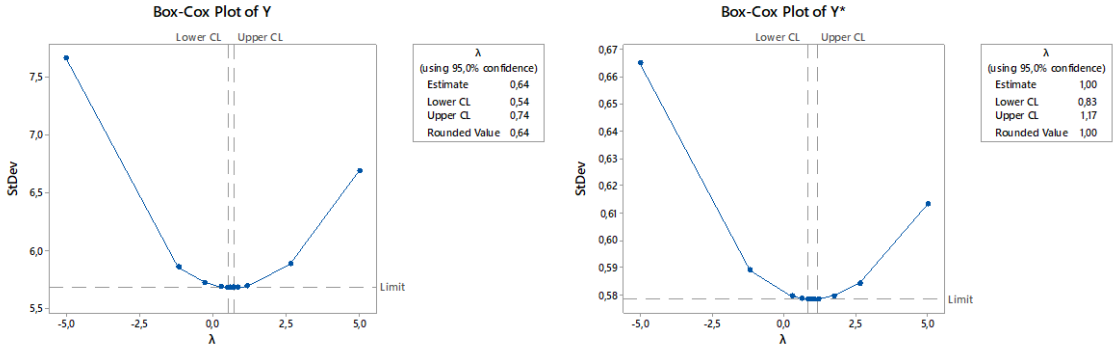
Figure 9 indicates that the highest sea level category occurred between May and June, while the lowest sea level was observed in March.

3.2 Modeling using ARIMA

To ensure the robustness of the model, the dataset will be divided into training and testing subsets. The training set will encompass data from January 2021 to November 2022, while the testing set will consist of observations from December 2022. Prior to modeling, it is essential to confirm that the data meets the ARIMA requirements, specifically that it is stationary in both variance and mean. By addressing these prerequisites, the study seeks to develop a reliable forecasting model that can provide valuable insights into future sea level trends.

The time series plot indicates that the data is not stationary in variance. To address this issue, data transformation is necessary to achieve stationarity. One effective method for this is the Box-Cox transformation. This transformation helps stabilize the variance of the data, making it more suitable for further analysis, including the application of the ARIMA model. By using the Box-Cox transformation, it is anticipated that the data will be controlled and rendered stationary, enabling more accurate and effective modeling.

The data Y is transformed into $Y^* = Y^{\lambda}$ with λ=0.64, and the data is stationary with respect to variance by observing the rounded value of 1 on the Box-Cox plot for Y^*.



(a) (b)

Figure 10. Box-Cox Plot of a) Actual Data (Y), (b) Transformed Data (Y^*)
(Source: Output R)

After satisfying the stationarity requirement, the next criterion to address is stationarity in terms of means. The Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) plots are presented below.

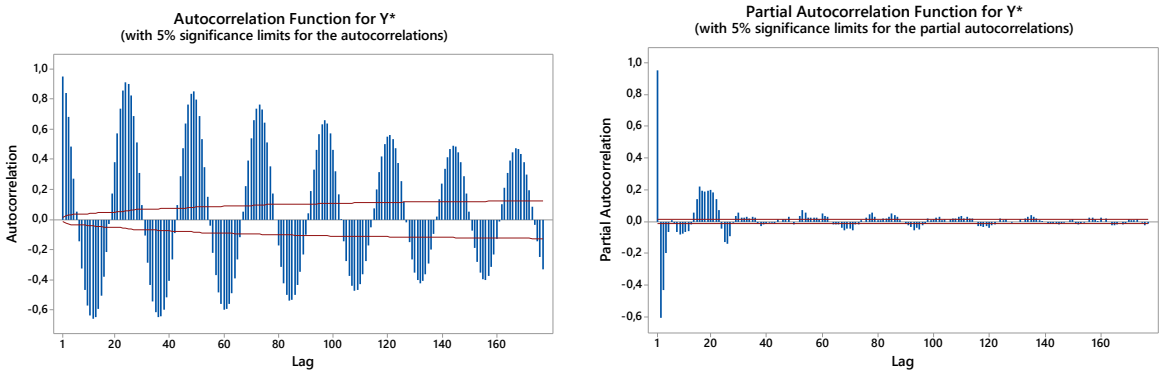


Figure 11. Plot ACF and PACF on Transformed Data
(Source: Output R)

Figure 11 illustrates that the data is not yet stationary, as evidenced by the ACF plot, which displays a discernible pattern. In the PACF plot, the first lag value is nearly 1, indicating that differencing is necessary at the first lag. The results following the differencing are presented in Figure 12 and Figure 13.

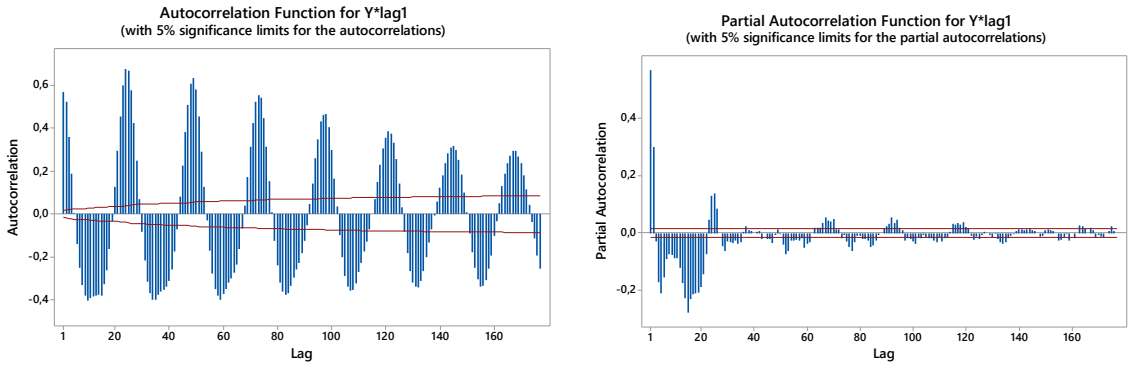


Figure 12. ACF and PACF Plots of Transformed Data after First Differencing
(Source: Output R)

Figure 12 reveals that the ACF plot still exhibits a seasonal pattern, particularly evident at the lag of 12, which is more pronounced than the others. This indicates that the seasonal pattern remains non-stationary.

The next step is to perform seasonal differencing at the lag of 24, given that the sea level data is recorded hourly. The results of this differencing are shown in Figure 13.

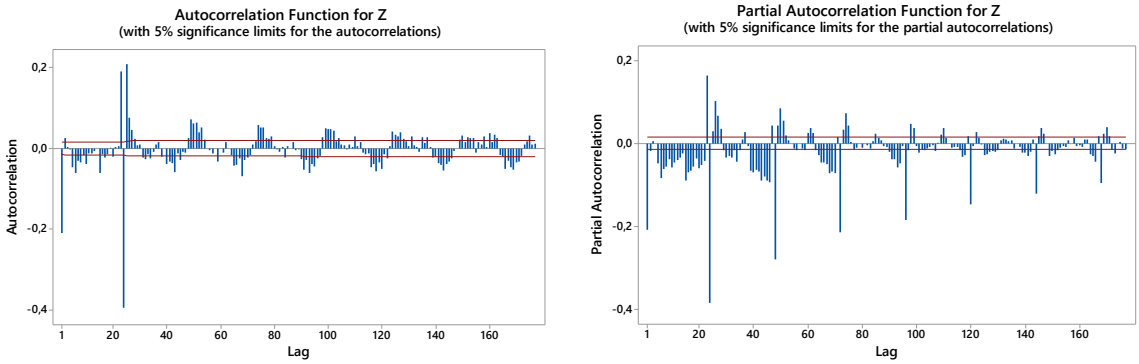


Figure 13. ACF and PACF Plots of Differenced Data for Seasonal Period of 24
(Source: Output R)

Figure 13 indicates that the data is now stationary with respect to mean, demonstrating that the ARIMA model exhibits a seasonal effect with a period of 24. Once all requirements are met, model estimation is carried out using the ACF and PACF plots presented in Figure 13. Several model estimations for Seasonal ARIMA have been derived using diagnostic techniques, with detailed information provided in Table 6.

Table 6. Comparison of Prediction Seasonal ARIMA Model

No	Estimated *Seasonal* ARIMA Model	Parameter	RMSE	MAPE
1	SARIMA $(1,1,1)(1,1,0)^{24}$	significant	0.6068132	1.599858
2	SARIMA $(1,1,0)(1,1,0)^{24}$	significant	0.6073556	1.602251
3	SARIMA $(0,1,1)(1,1,0)^{24}$	significant	0.6077408	1.602967
4	SARIMA $(2,1,0)(1,1,0)^{24}$	significant	0.6063576	1.596974
5	SARIMA $(2,1,1)(1,1,0)^{24}$	significant	0.6060967	1.595634
6	SARIMA $(0,1,2)(1,1,0)^{24}$	significant	0.6060296	1.595561
7	**SARIMA $(1,1,2)(1,1,0)^{24}$**	**significant**	**0.6057199**	**1.593918**

The estimation and testing of the estimated model parameters presented in Table 6 were conducted using a 95% confidence interval (α=0.05), confirming that the models are statistically significant. Notably, the Root Mean Squared Error (RMSE) of 0.606 and the Mean Absolute Percentage Error (MAPE) of 1.59% indicates that the ARIMA model $(1,1,2)(1,1,0)^{24}$ produces the smallest MSE value, demonstrating its effectiveness. Detailed information is provided in Table 7.

Table 7. The Estimated Parameter Values for The Best ARIMA Model

Model	Parameter	Coefficient	S.E Coefficient	P-Value
$(1,1,2)(1,1,0)^{24}$	AR 1	0.2614486	0.0499643	1.670e-07
	MA 1	-0.3614227	0.0496345	3.297e-13
	MA 2	0.1095503	0.0094495	< 2.2e-16
	SAR 24	-0.3823873	0.0074701	< 2.2e-16

The model indicates that the Marina Ancol Sea level at time t is dependent on its previous values as well as those from 23 and 24 hours prior. Additionally, the SARIMA $(1,1,2)(1,1,0)^{24}$ model in Table 7 demonstrates that all p-values for the parameters are less than 0.05, indicating that all model parameter are statistically significant. The SARIMA $(1,1,2)(1,1,0)^{24}$ model was then fitted using a testing dataset to evaluate its goodness of fit, with detailed information presented in

Table 8.

Table 8. Goodness of ARIMA Model on Data Testing

Model ARIMA	RMSE	MAPE
SARIMA $(1,1,2)\ (1,1,0)^{24}$	3.072573	8.829044

Table 8 presents the MAPE value of 8.8% obtained from fitting the SARIMA $(1,1,2)(1,1,0)^{24}$ model on the testing dataset, indicating that this model is sufficiently effective for forecasting sea level at Marina Ancol.

Following the estimation and significance testing of the parameters, diagnostic testing for the SARIMA $(1,1,2)(1,1,0)^{24}$ model was performed. This included residual analysis, which involved white noise testing using the Ljung-Box-Pierce test and normality testing using the Kolmogorov-Smirnov test. The results of the Ljung-Box-Pierce test, shown in Table 9, reveal that the residuals from the model are not white noise, as indicated by a p-value less than 0.05. This may be attributed to the inherent complexity of sea level variations.

Table 9. White Noise Residual Assumption Test on Estimated Seasonal ARIMA Model

Lag	**12**	**24**	**36**	**48**
Chi-Square	492,29	1283,26	2923,37	3792,65
DF	7	19	31	43
P-Value	0,000	0,000	0,000	0,000

The residual ACF plot for Model 7, shown in Figure 14, further illustrates that the residuals do not satisfy the white noise assumption, as significant correlations persist across multiple lags, extending up to lag 150.

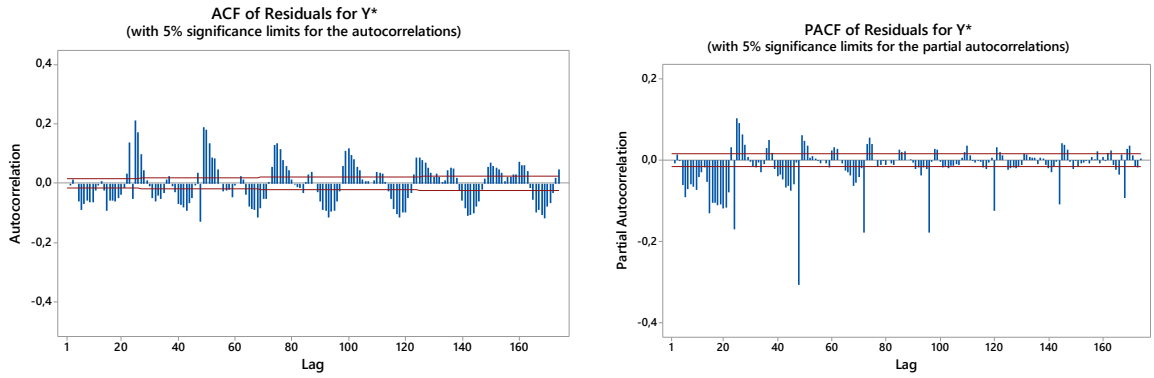


Figure 14. ACF and PACF Plot of Residuals for Model 7
(Source: Output R)

The results of the normality test on the ARIMA residuals, conducted using the Jarque-Bera test, show a chi-square value is 58477 and the P-Value is less than < 2.2e-16. This indicates that the model residuals do not conform to a normal distribution, as evidenced by a p-value less than 0.05. This non-normality may be attributed to the substantial volume of sea level data at Marina Ancol, as illustrated in Figure 2.

3.3 Modeling using Hybrid ARIMA-RBFNN Models

After obtaining the residual data from the optimal ARIMA model, the next step is to perform Hybrid modeling using ARIMA and Radial Basis Function Neural Networks (RBFNN). The first phase involves modeling the data with the ARIMA method. Following the identification of the best ARIMA model, the subsequent step is to apply the RBFNN method, utilizing inputs derived from the significant lags identified in the PACF plot and the residuals from the best ARIMA model. The PACF of the ARIMA model residuals is depicted in Figure 15.

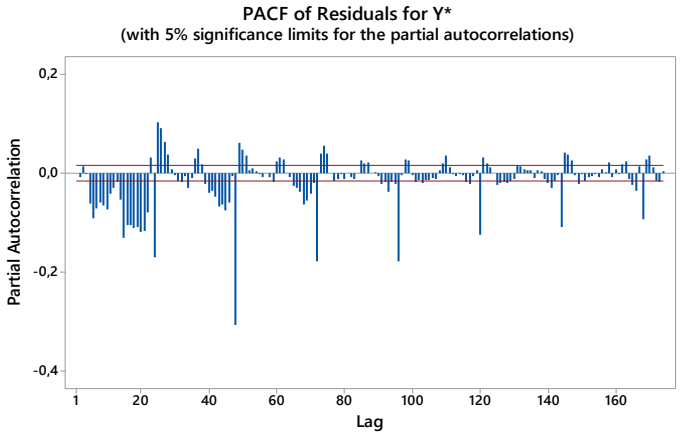


Figure 15. PACF Plot of Residuals from the Optimal ARIMA Model
(Source: Output R)

Figure 15 shows that the residuals from the optimal ARIMA model still exhibit significant partial autocorrelations, indicating the presence of remaining temporal dependence not captured by the linear model. The most prominent and consistent significance occurs around the 24-hour lag, which corresponds to the dominant diurnal tidal cycle. Based on this, the RBFNN input structure was designed to include residuals from lag 1 to lag 24 in order to capture short-term nonlinear dependencies within one complete tidal cycle. This input selection is not only supported by the PACF pattern but also aligns with the physical characteristics of sea level dynamics and ensures that the RBFNN effectively learns intra-day nonlinear behavior without introducing excessive model complexity. Consequently, the input variables for the RBFNN modeling consist of 24 variables, specifically the residuals from the 1st lag to the 24th lag, denoted as $e_{t-1}, e_{t-2}, \ldots, e_{t-24}$, .

After completing the data preprocessing, the training and testing datasets for the RBFNN analysis are defined: data from January 2021 to October 2022 is used for training, while November 2022 is designated for testing. The Radial Basis Function Neural Network (RBFNN) employed in this study adopts a three-layer architecture consisting of an input layer, a hidden layer, and an output layer. The input layer receives lagged residual values as predictor variables, while the hidden layer is composed of radial basis neurons using a Gaussian (exponential) activation function to capture nonlinear patterns in the data. The output layer generates the final forecast through a linear combination of the hidden layer outputs.

The number of hidden nodes determines the model complexity and is selected through a trial-and-error process by testing 1 to 10 nodes. The optimal architecture is chosen based on the smallest RMSE obtained from the testing dataset. Both input and output variables are normalized prior to training to improve numerical stability and learning efficiency. The detailed RMSE results for different network configurations are presented in Table 10.

Table 10. RMSE of RBFNN Models According to Total Node in Hidden Layer

Total Node	RMSE Training	RMSE Testing
1	0.999	1.002
2	0.998	1.002
3	0.993	0.994
4	0.991	0.986
5	**0.989**	**0.985**
6	0.973	1.003
7	0.971	1.004
8	0.968	1.017
9	**0.962**	**1.015**
10	0.962	1.018

Table 10 indicates that the smallest RMSE value for the training dataset is achieved by the RBFNN model with a total of 8 nodes, yielding an RMSE of 0.962. Conversely, the smallest RMSE value for the testing dataset is produced by the model with 7 nodes, resulting in an RMSE of 0.975. The selection of the number of nodes in the hidden layer is based on the smallest RMSE observed in the testing dataset; thus, the optimal RBFNN model is identified as RBFNN (24,7,1) represents 24 nodes in the input layer, 7 nodes in the

hidden layer, and 1 node in the output layer. Compared to conventional linear models such as ARIMA, RBFNN is more flexible in capturing nonlinear relationships in the data. Moreover, unlike more complex deep learning models, such as LSTM or deep neural networks, the selected RBFNN structure maintains a simple architecture with lower computational cost while still delivering competitive accuracy. These characteristics make the chosen RBFNN configuration suitable for short-term sea level forecasting using high-frequency data.

3.4 Forecasting Sea Level of Marina Ancol using Hybrid ARIMA-RBFNN Model

In alignment with the primary objective of this research, which is to forecast the sea level at the Marina Ancol floodgate, both ARIMA modeling and RBFNN modeling for the residuals have been conducted. A comparison of the MAPE values for the training and testing datasets is presented in Table 11.

Table 11. Comparison of MAPE value in ARIMA and Hybrid Model

	MAPE ARIMA	MAPE Hybrid
Training	2.47%	**2.44%**
Testing	2.76%	**2.74%**

Table 11 shows that the hybrid ARIMA–RBFNN model achieves a slightly lower MAPE than the ARIMA model for both training and testing datasets. Although the numerical improvement appears modest, this result is meaningful in the context of short-term sea level forecasting, where residual errors are already small and further reductions are inherently difficult to obtain. The improvement indicates that the hybrid model is able to capture nonlinear residual patterns that remain unmodeled by ARIMA, leading to more stable and reliable forecasts. Moreover, the consistent performance gain observed in both training and testing phases suggests that the hybrid approach enhances generalization rather than merely fitting noise, which is particularly important for operational forecasting applications such as tidal flooding preparedness.

To forecast the next 24 hours, starting from January 1, 2023, at 1 a.m. to 11 p.m., the model employed is Hybrid SARIMA $(1,1,2)(1,1,0)^{24}$–RBFNN (24,7,1). Detailed forecasting results are provided in Table 12.

Table 12. Sea Level forecasting at Marina Ancol for January 1st 2023 (per hour)

No	Time	ARIMA	Residual	Sea Level Rise
17521	01:00	151.200	0.058	151.26
17522	02:00	146.251	0.095	146.35
17523	03:00	151.643	0.072	151.71
17524	04:00	151.462	0.080	151.54
17525	05:00	153.868	0.069	153.94
17526	06:00	159.548	0.063	159.61
17527	07:00	171.858	0.037	171.90
17528	08:00	169.970	0.030	170.00
17529	09:00	177.642	0.046	177.69
17530	10:00	180.217	-0.061	180.16
17531	11:00	188.850	-0.076	188.77
17532	12:00	192.809	-0.106	192.70
17533	13:00	197.528	-0.134	197.39
17534	14:00	198.067	-0.163	197.90
17535	15:00	199.984	-0.198	199.79
17536	16:00	190.049	-0.216	189.83
17537	17:00	185.332	-0.149	185.18
17538	18:00	180.026	0.081	180.11
17539	19:00	185.886	0.195	186.08
17540	20:00	183.109	0.130	183.24
17541	21:00	171.919	0.223	172.14
17542	22:00	161.364	0.856	162.22
17543	23:00	159.977	0.851	160.83
17544	00:00	158.356	0.786	159.14

Table 12 presents examples of 24-hour sea level forecasts generated by the hybrid model, showing short-term fluctuations that are mainly driven by tidal dynamics. The 24-hour sea level forecasts mainly reflect short-term tidal behavior at Marina Ancol Beach.

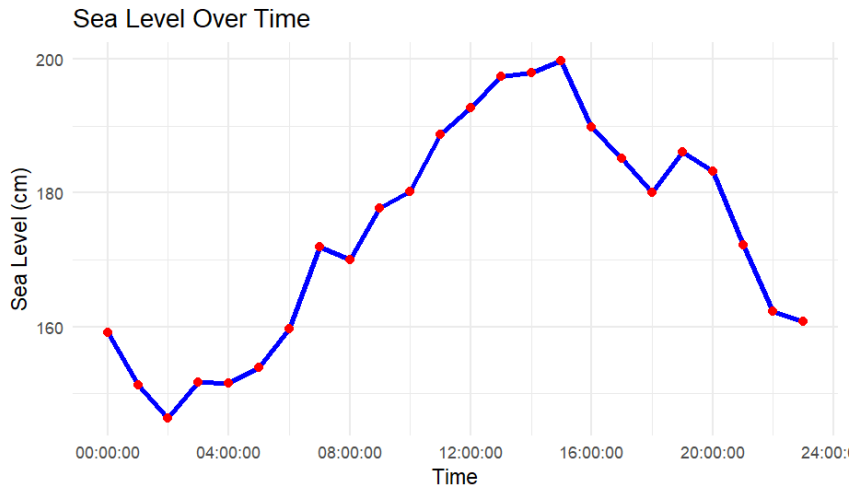


Figure 16. Forecasted Sea Level Values for 24-hour

Fig. 16 shows predicted patterns with clear daily fluctuations consistent with coastal tidal cycles. In addition, short-term changes in sea level are influenced by local weather conditions such as wind, air pressure, and rainfall. These factors introduce nonlinear effects that cannot be fully captured by the ARIMA model alone, which explains why the hybrid ARIMA–RBFN model provides slightly improved forecasts.

Previous studies on sea level forecasting generally use either statistical models or machine learning methods with daily or monthly data. Many of them focus on long-term trends rather than short-term predictions. In contrast, this study uses hourly sea level data and a hybrid ARIMA–RBFNN approach to improve short-term forecasting. Although the improvement in accuracy is modest, the results are comparable to those reported in earlier hybrid modeling studies and demonstrate the benefit of combining linear and nonlinear models for high-frequency sea level data.

The 24-hour forecast horizon is chosen to support short-term decision-making and early warning for tidal flooding. Short-term forecasts are useful for local authorities to prepare mitigation actions, manage coastal activities, and improve public safety in tourism areas. Extending the forecast horizon may increase uncertainty due to rapidly changing weather conditions. Therefore, a 24-hour horizon provides a practical balance between accuracy and usefulness.

4. CONCLUSION

In conclusion, this study successfully forecasts the sea level at the Marina Ancol floodgate using a Hybrid ARIMA-RBFNN model. The results indicate that the Hybrid method outperforms the traditional ARIMA approach, as evidenced by lower MAPE values. The Hybrid model effectively addresses residual issues, which are not classified as white noise in the ARIMA model. The forecast reveals a seasonal pattern in sea level, with elevated levels expected at the beginning of the year, fluctuations through mid-year, and another peak at year-end. These findings align with observed rainfall patterns, underscoring the model's relevance and accuracy in predicting sea level changes. Overall, the Hybrid ARIMA-RBFNN approach proves to be a robust tool for effective sea level forecasting, contributing valuable insights for flood management and planning in the Marina Ancol area.

Despite these promising results, this study has several limitations. First, the forecasting horizon is limited to 24 hours, which restricts its applicability for longer-term planning. Second, the model relies solely on historical sea level data and does not explicitly incorporate external physical factors such as wind, atmospheric pressure, or rainfall, which may influence sea level variability. In addition, the improvement in forecasting accuracy over the ARIMA model is relatively modest, indicating that further enhancement is possible.

Future research may extend this work by exploring longer forecast horizons, integrating meteorological and oceanographic variables, and comparing the proposed hybrid model with other advanced machine learning or deep learning approaches. Such extensions could improve forecasting accuracy and broaden the practical usefulness of sea level prediction models for coastal risk management.

Author Contributions

Soehardjoepri: Conceptualization, Supervision, Funding Acquisition, and Resources. Ulil Azmi: Conceptualizations, Methodology, Data Analysis, Software, Validation, and Writing of the original draft. Prilyandari Dina Saputri: Data Curation, Analysis, Validation and Review. Moch. Taufik Hakiki: Draft Preparation, Writing-Original Draft, and Software Development. Denisha A. E. Ananda: Validation, Visualization, and Writing – Review and Editing. Roslinazairimah Zakaria: Literature Review, Validation, Manuscript Refinement. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This work is funded by the Directorate of Research and Community Service, Institut Teknologi Sepuluh Nopember (ITS), Surabaya for the programme of Scientific Research in ITS.

Acknowledgment

We would like to extend our sincere gratitude to the Directorate of Research and Community Service (DRPM) at Institut Teknologi Sepuluh Nopember (ITS) and Dinas Sumber Daya Air DKI Jakarta for their invaluable support and guidance throughout this research. Their encouragement and resources were instrumental in the successful completion of this study.

Declarations

The authors declare that they have no competing interests.

Declaration of Generative AI and AI-assisted Technologies

AI-assisted technology (ChatGPT) was used to support sentence restructuring and clarity improvements. The authors confirm that the underlying ideas, arguments, data analyses, and conclusions are original and were not generated by AI. All AI-assisted edits were critically reviewed and validated by the authors.

REFERENCES

[1] IPCC, "CLIMATE CHANGE 2021: THE PHYSICAL SCIENCE BASIS," in *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, 2021. [Online]. Available: www.ipcc.ch

[2] D. Khojasteh, W. Glamore, V. Heimhuber, and S. Felder, "SEA LEVEL RISE IMPACTS ON ESTUARINE DYNAMICS: A REVIEW," *Science of the Total Environment*, vol. 780, Aug. 2021, doi: https://doi.org/10.1016/j.scitotenv.2021.146470

[3] Dinas Sumber Daya Air, "PINTU AIR," DSDA Jakarta.

[4] S. Liu *et al.*, "WARNING WATER LEVEL DETERMINATION AND ITS SPATIAL DISTRIBUTION IN COASTAL AREAS OF CHINA," *Natural Hazards and Earth System Sciences*, vol. 23, pp. 127–138, Jan. 2023, doi: https://doi.org/10.5194/nhess-23-127-2023

[5] R. J. Hyndman and G. Athanasopoulos, *FORECASTING: PRINCIPLES AND PRACTICE* , 2nd ed. Melbourne: OTexts, 2018, doi: https://doi.org/10.32614/CRAN.package.fpp2

[6] I. Syahzaqi, Sediono, M. K. Dyaksa, A. T. Vionita, and A. N. Ghasani, "PREDICTION OF AVERAGE TEMPERATURE IN BANYUWANGI REGENCY USING SARIMA," *Barekeng*, vol. 19, no. 3, pp. 2207–2218, 2025, doi: https://doi.org/10.30598/barekengvol19iss3pp2207-2218

[7] W. A. Pratiwi, A. F. Rizki, K. A. Notodiputro, Y. Angraini, and L. N. A. Mualifah, "THE COMPARISON OF ARIMA AND RNN FOR FORECASTING GOLD FUTURES CLOSING PRICES," *Barekeng*, vol. 19, no. 1, pp. 397–406, 2025, doi: https://doi.org/10.30598/barekengvol19iss1pp397-406

[8] A. Musbikhin, E. Sediyono, and C. E. Widodo, "MODEL SEA LEVELS PREDICTION WITH ARIMA FOR COASTAL AREA IN SEMARANG," *Proceedings of the International Conference on Innovation in Science and Technology (ICIST 2020)*, 2021, doi: 10.2991/aer.k.211129.001.

[9] P. K. Srivastava, T. Islam, S. K. Singh, G. P. Petropoulos, M. Gupta, and Q. Dai, "FORECASTING ARABIAN SEA LEVEL RISE USING EXPONENTIAL SMOOTHING STATE SPACE MODELS AND ARIMA FROM TOPEX AND SASON SATELLITE RADAR ALTIMETER DATA," *Meteorological Applications*, vol. 23, pp. 633–639, 2016, doi: https://doi.org/10.1002/met.1585

[10] A. Elzwayie, A. El-shafie, Z. M. Yaseen, H. A. Afan, and M. F. Allawi, "RBFNN-BASED MODEL FOR HEAVY METAL PREDICTION FOR DIFFERENT CLIMATIC AND POLLUTION CONDITIONS," *Neural Comput Appl*, vol. 28, no. 8, pp. 1991–2003, 2017, doi: https://doi.org/10.1007/s00521-015-2174-7

[11] W. Reza, F. C. H. Buan, and S. N. Abdussamad, "COMPARISON OF RADIAL BASIS FUNCTION NEURAL NETWORKS (RBFNN) AND AUTOREGRESSIVE MOVING AVERAGE (ARMA) ALGORITHMS ON INFLATION RATE PREDICTION MODELS IN BATAM CITY," *Jurnal Info Sains : Informatika dan Sains*, vol. 14, no. 02, p. 2024, doi: https://doi.org/10.58471/infosains.v14i02

[12] A. Setiawan, "PREDIKSI TINGGI MUKA AIR MENGGUNAKAN SUPPORT VECTOR MACHINE BERBASIS PARTICLE SWARM OPTIMIZATION," *Jurnal Ilmiah Fakultas Technologia*, vol. 7, no. 2, 2016, doi: https://doi.org/10.31602/tji.v7i2.616

[13] A. A. Alsuwaylimi, "COMPARISON OF ARIMA, ANN AND HYBRID ARIMA-ANN MODELS FOR TIME SERIES FORECASTING," *Information Sciences Letters*, vol. 12, no. 2, pp. 1003–1016, Feb. 2023, doi: https://doi.org/10.18576/isl/120238

[14] A. P. Baluk, H. Yasin, and Sugito, "PERAMALAN TINGGI GELOMBANG LAUT DENGAN METODE VECTOR AUTOREGRESSIVE-RADIAL BASIS FUNCTION NETWORK (VAR-RBFN)," *J Statistika*, vol. 13, no. 1, pp. 39–46, 2020, doi: https://doi.org/10.36456/jstat.vol13.no2.a3270.

[15] A. A. Soebroto, I. Cholissodin, R. C. Wihandika, M. T. Frestantiya, and Z. El Arief, "PREDIKSI TINGGI MUKA AIR (TMA) UNTUK DETEKSI DINI BENCANA BANJIR MENGGUNAKAN SVR-TVIWPSO," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, p. 79, Jul. 2015, doi: https://doi.org/10.25126/jtiik.201522126

[16] B. L. Bowerman and R. O'Connel, *FORECASTING AND TIME SERIES: AN APPLIED APPROACH*, 3rd ed. California: Duxburry Press, 1993.

[17] J. D. Cryer and K. S. Chan, *TIME SERIES ANALYSIS WITH APPLICATIONS IN R*. Springer, 2008, doi: https://doi.org/10.1007/978-0-387-75959-3

[18] K. Ramasubramanian and A. Singh, *MACHINE LEARNING USING R* . New Delhi: Apress, 2017, doi: https://doi.org/10.1007/978-1-4842-2334-5

[19] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *TIME SERIES ANALYSIS: FORECASTING AND CONTROL*. New Jersey: John Wiley and Sons Inc., 2015.

[20] V. Wahyuningrum, "PENERAPAN RADIAL BASIS FUNCTION NEURAL NETWORK DALAM PENGKLASIFIKASIAN DAERAH TERTINGGAL DI INDONESIA," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 12, no. 1, 2020, doi: https://doi.org/10.34123/jurnalasks.v12i1.250.

[21] G. P. Zhang, "TIME SERIES FORECASTING USING A HYBRID ARIMA AND NEURAL NETWORK MODEL," *Neurocomputing*, vol. 50, pp. 159–175, 2003, doi: https://doi.org/10.1016/S0925-2312(01)00702-0.

[22] Y. Liu, C. Yang, K. Huang, and W. GUI, "NON-FERROUS METALS PRICE FORECASTING BASED ON VARIATIONAL MODE DECOMPOSITION AND LSTM NETWORK," *Knowl Based Syst*, vol. 188, p. 105006, 2020, doi: https://doi.org/10.1016/j.knosys.2019.105006.