Open Access, August 2024
p-ISSN : 2087-9423
e-ISSN : 2620-309X

Jurnal Ilmu dan Teknologi Kelautan Tropis, 16(2):163-178
DOI: https://doi.org/10.29244/jitkt.v16i2.52818
Available online at https://journal.ipb.ac.id/index.php/jurnalikt

COMPARISON OF ARIMA AND LSTM METHODS IN PREDICTING JAKARTA SEA LEVEL

Yanuar Nurul Hilal, Gibson Daniel Andrianto Nainggolan, Sabilla Hamda Syahputri, and Fitri Kartiasih*
Statistical Computing Study Program, Politeknik Statistika STIS, Jakarta, 13330, Indonesia
Statistics Study Program, Politeknik Statistika STIS, Jakarta, 13330, Indonesia
*Correspondence: fkartiasih@stis.ac.id
(Received 31-12-2024; Revised 07-05-2024; Accepted 28-07-2024)

ABSTRACT

As a coastal city, Jakarta faces enormous risks from sea level rise brought on by climate change, and it is critical to create efficient plans to anticipate and minimize any potential negative effects. Predictive modeling is essential in addressing this challenge in order to anticipate and mitigate any potential negative effects of sea level rise. Therefore, research was conducted with the aim of comparing the performance of two prediction methods, namely Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM). Sea level was predicted using both techniques up to the end of 2023. Performance indicators, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE), were employed to assess the quality of both prediction models. The result shows that the ARIMA (1,1,4) model is more effective in predicting sea level than the LSTM. The MAE, MAPE, and RMSE values for ARIMA (1,1,4) are 7.19, 4.86%, and 10.35, respectively. In the meantime, the sea level in Jakarta is predicted to remain reasonably steady, according to the forecasted findings from both models. This study is expected to make a significant contribution to understanding and mitigating the potential impacts of sea level rise in Jakarta as a result of climate change.
Keywords: ARIMA, forecasting, Jakarta, LSTM, sea level

Perbandingan Metode ARIMA dan LSTM untuk Peramalan Tinggi Permukaan Air Laut Jakarta

ABSTRAK

Dalam menghadapi risiko yang signifikan akibat kenaikan permukaan air laut yang dipicu oleh perubahan iklim, Jakarta sebagai kota pesisir memiliki kebutuhan mendesak untuk mengembangkan strategi yang efektif guna mengantisipasi dan memitigasi potensi dampak negatif. Dalam menghadapi tantangan ini, prediksi menjadi kunci untuk mengantisipasi dan meminimalkan dampak negatif yang mungkin timbul dari kenaikan permukaan air laut. Oleh karena itu, penelitian dilakukan dengan tujuan memperbandingkan kinerja dua metode prediksi, yaitu Autoregressive Integrated Moving Average (ARIMA) dan Long Short-Term Memory (LSTM). Kedua metode ini diaplikasikan untuk meramalkan tinggi permukaan air laut hingga akhir tahun 2023. Dalam mengevaluasi kualitas kedua model prediksi, digunakan metrik kinerja seperti Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan Root Mean Square Error (RMSE). Hasil analisis menunjukkan bahwa model ARIMA (1,1,4) lebih efektif dalam memprediksi tinggi permukaan air laut dibandingkan LSTM. ARIMA (1,1,4) memiliki nilai MAE 7,19, MAPE 4,86%, dan RMSE sebesar 10,35. Sementara itu, hasil forecasting dari kedua model didapatkan bahwa ketinggian permukaan air laut Jakarta diprediksi relatif stabil. Studi ini diharapkan dapat memberikan kontribusi yang signifikan dalam pemahaman serta mitigasi potensi dampak kenaikan permukaan air laut di Jakarta sebagai hasil dari perubahan iklim.
***Kata kunci**: ARIMA, Jakarta, LSTM, peramalan, tinggi air laut*

INTRODUCTION

High oscillations in rainfall and sea level rise are two major effects of climate change caused by global warming events, driven by an increase in greenhouse gases, primarily carbon dioxide (CO_2) and methane (CH_4). Global temperature records dating back to the 19th century have revealed variations in average temperatures that serve as markers for climate change. The average global temperature increased steadily between 1906 and 2005, indicating this shift in global temperature. It is estimated that the average global temperature will rise by an additional 1.8 – 4.0°C this century, and some IPCC estimates have the estimate as high as 1.1 – 6.4°C (Leontinus, 2022).

Furthermore, it is currently the sixth year that the Sustainable Development Goals (SDGs), established in September 2015, are being implemented. One SDG target focuses on climate change, and that is the 13th goal, which addresses tackling climate change. In response to this discovery, the government incorporated it into the National Medium-Term Development Plan IV (RPJMN), which includes elements like: (1) enhanced resilience to disasters and climate change; and (2) low-carbon development. By enhancing the convergence of disaster risk reduction and climate change adaptation through disaster management techniques and boosting climate resilience, the overall resilience to disasters and climate change can be enhanced.

Sea level rise is one of the most important effects of climate change. Sea levels regularly rise and fall due to tides, which are resulting from the gravitational pull of the earth, sun, and moon. The primary cause of sea level change is tides. Flooding in low-lying areas is a phenomenon that can happen when sea levels rise significantly, which is a direct result of sea level rise.

One of the environmental phenomena that happens frequently and is highly highlighted by its development is Jakarta, the capital city of Indonesia. In the province of DKI Jakarta, there were 20 instances of flash floods and 14 incidents of flooding in 2021, according to BPS data on natural disasters. For both the local administration and the society at large, this is undoubtedly a concerning situation. A dictionary known as 'Jakarta Tenggelam' emerged as a result of this circumstance. Rising sea levels and excessive groundwater use, which results in land subsidence, are the main causes of Jakarta's sinking.

The National Research and Innovation Agency (BRIN) estimates that Jakarta's sea level is rising by roughly 4.3 mm year, although land subsidence also plays a part in Jakarta's possible sinking. The softness of the rocks, groundwater extraction, and extensive development on Jakarta's north coast are the main causes of land subsidence. A major threat is posed by the combination of land subsidence and sea level rise, particularly to coastal areas like city along the north coast of java, including Jakarta (Widodo, 2017).

Experts anticipate that Jakarta will submerge by 2050. The findings of a simulation of sea level rise and its effects on Jakarta, which is covering approximately 160.4 km^2 area of its area were given by the principal researcher of the National Research and Innovation Agency (BRIN). There have been long-standing predictions that Jakarta will sink. This has to do with the increasing water levels and ground subsidence in Jakarta. In fact, in ten years, the earth has sunk by 2.5 meters in portions of North Jakarta that are extremely vulnerable to flooding (Bennett *et al.*, 2023).

Jakarta's collapse can be caused by a multitude of factors. These include Jakarta's thirteen rivers, its closeness to the ocean, its excessive and progressively use of groundwater extraction that resulted in the subsidence of its land (Sumartapraja and Christianti, 2022). The government is planning to build a massive sea wall spanning 35 kilometers when it is finished in 2025 as one of the measures to stop Jakarta from slipping under the waves. But this solution will only be effective temporarily. Therefore, sea level forecasting needs to be done in order to see conditions that may occur in the future and can be used as a basis for policy making for potentially dangerous things in coastal areas.

The Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA) techniques will be utilized to forecast sea level. A strong statistical association between the anticipated variable and the variable's historical value can be found using the ARIMA model. One of the advancements in neural network techniques that can be applied to the modeling of time series data is called long short-term memory networks, or LSTM.

There have been various attempts to predict sea level at several locations using several method approaches. As was done by Ramadhan *et al.* (2021) in short-term forecasting of sea level in the Sunda Strait. The research was conducted by comparing two machine learning methods, namely RNN and LSTM with the result that LSTM is better at analyzing sea level patterns. In addition, comparisons of classical statistical methods such as sarima with machine learning are also often carried out, such as research by Sun *et al.* (2020) in predicting variations in sea level in the Chinese sea. this study also uses the SARIMA + LSTM combination method which is also the best fitting model compared to other methods.

Another research was conducted by Balogun and Adebisi (2021) in predicting sea level along West Peninsular Malaysia coastline by creating several scenarios of different variable combinations that affect sea level by comparing ARIMA, SVR, and LSTM methods. The results showed that the ARIMA method performed best due to the dominating influence of tides. This suggests that the performance and suitability of prediction models vary across the region and selecting the optimal prediction model depends on the dominant physical processes governing sea level variability in the region of investigation.

The purpose of this research is to evaluate the effectiveness of the LSTM and ARIMA techniques and to decide which approach will provide the most accurate sea level values forecast for the Jakarta region based on the computation results. From November 19, 2023, to December 31, 2023, sea level is predicted using both techniques. Naturally, the goal's expectation is that it will yield results that will be helpful for future research of a similar nature and that the local government can use to inform policy decisions aimed at curbing sea level rise and achieving the SDG's thirteenth target.

RESEARCH METHODS

Collecting Data

The data used in this study are observations from monitoring station facilities run by the Intergovernmental Oceanographic Commission of UNESCO (IOC/UNESCO), which are accessible via the organization's official website at https://www.ioc-sealevelmonitoring.org. Users can find comprehensive information about monitoring stations on this page, along with data and other characteristics pertaining to sea level monitoring at particular monitoring locations.

Monitoring station is located in Tanjung Priuk, North Jakarta (Figure 1), at coordinates -6° 6' 24" S and 106° 53' 26.988" E. Sea level data taken from November 1, 2021, at 00.00 WIB to November 17, 2023, at 23.59 WIB with prs (pressure signals) sensors that record absolute sea level data every minute, yielding approximately one million observations. Some values, meanwhile, are missing because of potential mistakes made during the recording procedure. The insufficient data available for this location prevents us from adequately accounting for tides, winds, or other elements that may influence sea level in our model. Preprocessing is done to get consistent values by taking the maximum value of m for each observation time and the majority of these values appear at night time. The final dataset contains 748 observations for daily prediction.

Descriptive Analysis

Given that the data in this study are quantitative, a descriptive analysis can be employed to methodically characterize the data and present a summary of the connections among the phenomena under investigation. Since sea level height is the

Figure 1. Location of sea level monitoring station in North Jakarta.

research phenomenon, a number of data concentration measures pertaining to sea level height in the Jakarta region will be discussed. Furthermore, the descriptive analysis employed in this study is primarily concerned with comparing descriptions of sea level data collected in the Jakarta region over time, as demonstrated by the creation of graphs that illustrate the trend and the computation of the average sea level height range, as well as the middle, lowest, and maximum values of sea level height for each year. It is anticipated that the outcomes of many data concentration measure descriptions would be examined to offer a quick first summary of sea level data in the Jakarta region prior to the data being further processed using further data modeling and forecasting techniques.

Autoregressive Integreated Moving Average (ARIMA)

The AR, MA, and ARMA models are often combined into one model called the Autoregressive Integrated Moving Average (ARIMA) model. With *p* representing the AR (autoregressive) order, q the MA (moving average) order, and *d* the differential order, the ARIMA model order is expressed as ARIMA (*p*,*d*,*q*). The following equation (Dobre and Alexandru, 2008) generally illustrates the ARIMA model (*p*,*d*,*q*).

$$\left(1-\emptyset_1 B-\emptyset_2 B^2-\cdots-\emptyset_p B^p\right)(1-B)^d \tilde{Z}_t = \left(1-\theta_1 B-\cdots-\theta_q B^q\right)\alpha_t \quad \text{..........................} (1)$$

Information: Z_t is a forecast for period *t*; $\theta_1 \ldots \theta_q$. is a parameter of MA; α_t is a residual from period *t*; and $\emptyset_1 \ldots \emptyset_p$. is a parameter of AR; and $(1-B)^d$ is a differencing level.

For AR, MA, and ARMA models, the stationary time series data that needs to be modeled is the primary prerequisite. It is therefore required to distinguish in order to make the data that needs to be modeled stationary when it is not yet. The process of differencing will proceed to the next level if the initial one fails to yield stationary data. Details about how much agar stationary data differencing is listed in order d of the ARIMA model. One illustration would be if the sea level data (LEVEL) is not yet stationary, in which case distinguishing between :

$$\Delta LEVEL_t = LEVEL_t - LEVEL_{t-1} \quad \text{................} (2)$$

The *ARIMA(p, 1, q)* model of LEVEL is written as follows:

$$\Delta LEVEL_t = \mu + \alpha_1 \Delta LEVEL_t + \alpha_2 \Delta LEVEL_{t-1} + \cdots + \alpha_p \Delta LEVEL_{t-p} + u_t \beta_1 u_{t-1} + \beta_2 u_{t-2} + \beta_q u_{t-q} \quad (3)$$

Additionally, under the following assumptions, the patterns of the autocorrelation function (ACF) and partial autocorrelation function (PACF) can be used to determine the tentative ARIMA model if the data is stationary (Table 1).

To find the optimal model based on goodness of fit using the t test, F test, and R^2 score, mean square error (MSE), Schwarz criteria (SC), and Akaike information criterion (AIC), Subsequently, a diagnostic test and evaluation of the model are conducted based on the following criteria: The model has a small AIC, SC, and MSE; if the residual or error is not random or white noise, the steps must be repeated from the beginning; the model adheres to the concept of parsimony; and the estimated parameters are significant.

In addition, based on the model's evaluation, the following action might be to forecast for the upcoming few periods (short term). In general, long-term predictions made with the ARIMA model will be less accurate and more flat. Root Mean Squares Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are three error evaluation values that are used to determine how accurate forecasting values are. The predicting outcomes are better the smaller these numbers are.

Long Short-Term Memory (LSTM)

Modeling data using traditional methods such as ARIMA often requires the use of many parameters, classical assumptions, and complex equations. (LeCun *et al.*, 2015). For this reason, the development of the ability to learn, analyze and make a conclusion is done through the machine learning process until now continues to grow. LSTM is a type of Recurrent Neural Network (RNN) that is capable of learning long-term dependencies or interrelationships and is relatively popular for application to sequential data types such as time series data. (Yamak *et al.*, 2019). The LSTM model provides the benefit of not only overcoming long lags but also solving the gradient loss issue that occurs when learning long-term links in the RNN method. (Hochreiter and Schmidhuber, 1997). Gates additionally helps the network to understand what information should be retained, ignored, remembered, focused on, and excluded. (Yamak *et al.*, 2019). The LSTM architecture, which is based on research Van Houdt *et al.* (2020), can be seen in Figure 2.

Table 1. Patterns of ACF and PACF

Model	Patterns of ACF	Patterns of PACF
AR(p)	Exponential, exponential-oscillation or sinewave	Drastically decreased at certain lags (cut off)
MA(q)	Drastically decreased at certain lags (cut off)	Exponential, exponential-oscillation or sinewave
ARMA(p, q)	Exponential, exponential-oscillation or sinewave	Exponential, exponential-oscillation or sinewave

Information: Autoregressive (AR), Moving Average (MA), and Autoregressive Moving Average (ARMA) are models to make predictions based on historical data. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) help in identifying patterns and determining the appropriate order (p and q) for the time series model. Source: Makridakis *et al.* (1992).

Processing through the LSTM structure will be described in detail in the following context. The first structure is Block input, this section is used to update the component input block that combines the current input $x^{(t)}$ and the output of the LSTM observation $y^{(t-1)}$ in the previous iteration.

$$z^{(t)} = g(W_z x^{(t)} + R_z\, y^{(t-1)} + b_z) \text{ ...... (4)}$$

Information: W_z *and* R_z are constant weights while b_z is the bias and g is the hyperbolic tangent function $g(x) = h(x) = tanh(x)$. The hyperbolic tangent function is a nonlinear function that transforms an input in the form of a real number with a range between -1 and 1.

The next part is Input gate, this part determines how the process updates the long-term memory by performing operations on the current input $x^{(t)}$ and the outputs of the and $c^{(t-1)}$.

$$i^{(t)} = \sigma(W_i x^{(t)} + R_i\, y^{(t-1)} + p_i \cdot c^{(t-1)} + b_i) \text{ ...... (5)}$$

Information: $W_i, R_i, and\ p_i$ are constant weights and b_i is the bias while σ is sigmoid logistic function $\sigma(x) = \frac{1}{1+e^{1-x}}$. The sigmoid function is an activation function that transforms a real number into a number that ranges between 0 and 1.

As for the Forget gate, this part determines how much of the long-term memory ($c^{(t-1)}$) from the previous iteration to keep.

$$f^{(t)} = \sigma(W_f x^{(t)} + R_f\, y^{(t-1)} + p_f \cdot c^{(t-1)} + b_f) \text{ ...... (6)}$$

Information: $W_f, R_f, dan\ p_f$ are constant weights and b_f is the bias while σ is sigmoid logistic function $\sigma(x) = \frac{1}{1+e^{1-x}}$.

Meanwhile, the Output gate updates the short-term memory and outputs the result of the operation.

$$o^{(t)} = \sigma(W_o x^{(t)} + R_o\, y^{(t-1)} + p_o \cdot c^{(t-1)} + b_o) \text{ ...... (7)}$$

Information: $W_o, R_o, dan\ p_o$ are constant weights and b_o is the bias while σ is sigmoid logistic function $\sigma(x) = \frac{1}{1+e^{1-x}}$..

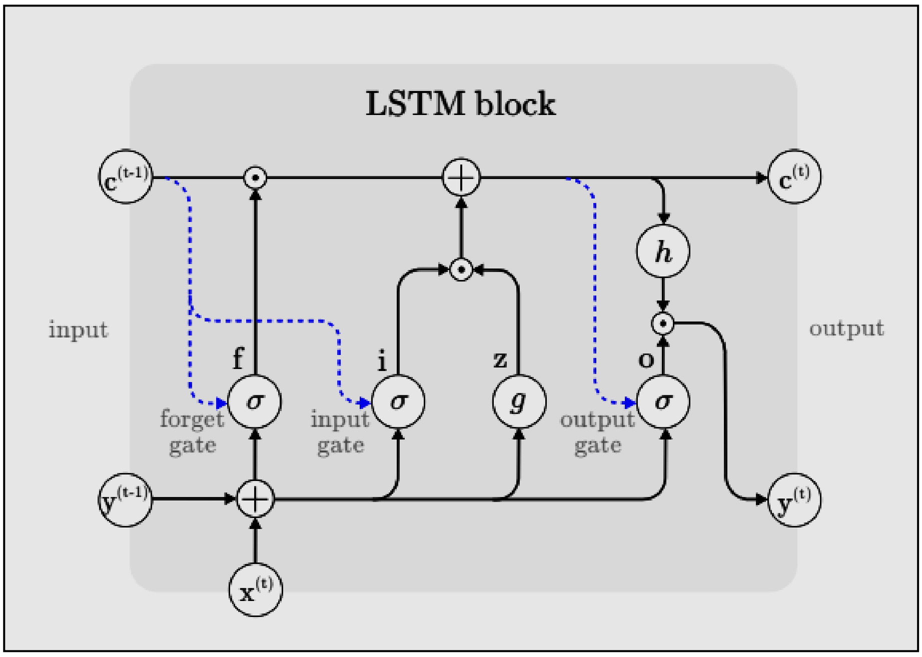


Figure 2. The Long Short-Term Memory architecture.

Finally Block output generates the final output of the current LSTM loop by operating $c^{(t)}$ or long-term memory with $o^{(t)}$ or short-term memory.

$$y^{(t)} = g(c^{(t)}) \cdot o^{(t)} \quad (8)$$

In this study, the LSTM method will be measured for accuracy by focusing on the root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) criteria. RMSE and MAE will give the degree of error in a unit value, while MAPE will give the degree of error in percentage. As explained in Hansun and Young (2021) all these measures can be represented as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2} \quad (9)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - F_t}{Y_t}\right| \quad (10)$$

$$MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - F_t}{Y_t}\right|\right) \times 100\% \quad (11)$$

Information: n is the number of data, Y_t is the true value, dan F_t is the forecasted value of the model.

RESULTS AND DISCUSSION

Descriptive Analysis

Figure 3 shows a descriptive study of sea level data collected in Jakarta between November 1, 2021, and November 18, 2023. Figure 3 illustrates that the data on the graph is largely steady, with very few obvious outliers. The period from January 2021 to August 2022 sees relatively constant sea levels. Nonetheless, there will be notable variations between September 2022 and December 2022 in the form of notable rises and falls in sea level. Increases and declines in sea level height between January 2023 and April 2023 will likely be more frequent and unstable. The sea surface height increases to a value of 2.64 meters, indicating a very notable shift. Given that the peak occurs in February 2023, the ongoing rainy season in Indonesia may possibly have an impact on this oddity. But from May 2023 to November 2023, the sea surface height returns to a fairly consistent level. Sea level height ranges from 0.85 meters at the minimum to 2.46 meters at the maximum.

Table 2 further suggests that in 2021, the sea level will average 1.61 meters year-round, with a minimum point of 1.31 meters and a maximum point of 1,988 meters. Furthermore, the sea level height is 1.60 meters, which is the middle number for this year. The data does not fluctuate significantly, as indicated by the 0.15 meters distribution of the data. The sea level height range for 2021 is 0.68 meters.
As of 2022, it can be said that the sea level is 1.11 meters at its lowest point and 2.27 meters at its highest point, with an average of 1.54 meters throughout the course of the year.

Furthermore, the sea level height this year is 1.51 meters, which is the medium value. The data does not change significantly, as: indicated by the distribution of the data, which is 0.21 meters. In 2023, the difference in sea level height is 1.16 meters.

It may be inferred that in 2023, the average annual sea level height is 1.40 meters, with a minimum point of 0.85 meters and a maximum point of up to 2.64 meters. Furthermore, the sea level height is 1.40 meters, which is the midway value for this year. The values do not fluctuate significantly, as indicated by the distribution of the data, which is 0.20 meters. With a value of 1.78 meters, the sea level height range in 2023 is fairly wide.

Autoregressive Integreated Moving Average (ARIMA)

The ADF test is utilized in the first phase to find non-stationarity non-stationary. P-values above the threshold denote non-stationarity non-stationary. First-order differentiation techniques must be used to

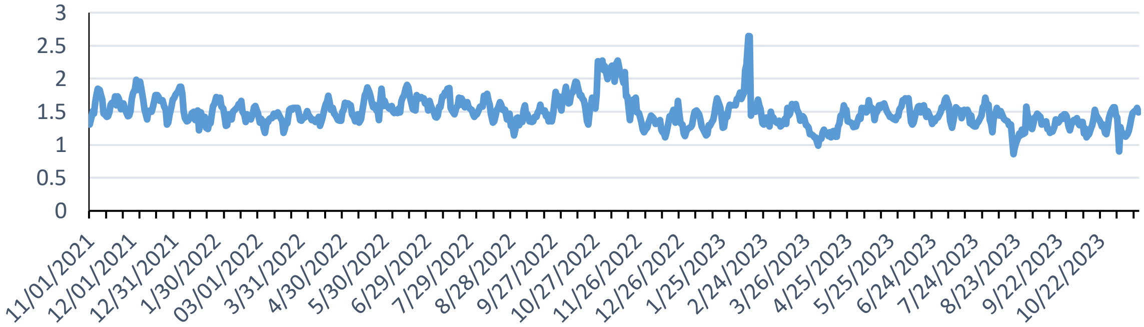


Figure 3. Time series Jakarta sea level processing results with x-axis as observation time and y-axis as sea level in meters.

Table 2. Statistical summary of time series data (sea level in meters)

Year	Mean	Median	Standard Deviation	Minimum	Maximum
2021	1.61	1.61	0.16	1.31	1.99
2022	1.55	1.52	0.22	1.11	2.28
2023	1.41	1.40	0.20	0.86	2.65

Information: Five summaries of statistics in order to measure a sea level in meters.

transform time series that are not stationary, such as the daily maximum sea level height. Developing an ARIMA model based on the ACF and PACF visual plots is the second stage. Significant correlations between observations at various time lags were demonstrated by significant values on the correlogram that were beyond confidence bounds. Through the correlogram plot in Figure 4, it can be shown that the actual data is not stationary as shown in the ACF correlogram pattern which is not significantly approaching zero, so it is necessary to perform differentiation at the first order and the result is that the data is stationary as shown in Figure 4b.

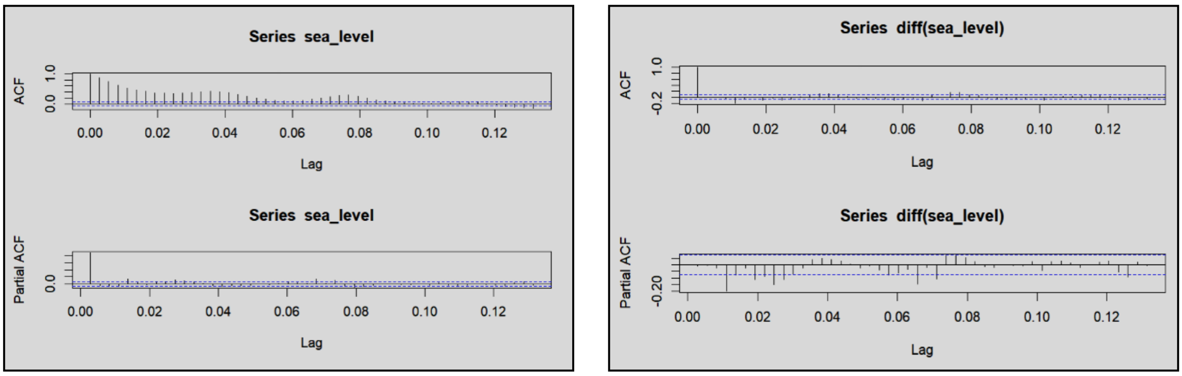
Finding the optimal model based on the minimum AICc criteria is the third stage. The ability to automatically choose the optimal ARIMA model in RStudio based on the lowest AIC value, the auto ARIMA result can be seen in Table 3. ARIMA(1,1,4) is choosen because it has the lowest AICc value and further evaluation is also performed on ARIMA(1,1,4), with an RMSE value of 0.1035, a MAPE value of 4.868, and an MAE value of 0.0719.

The chosen model is then checked for correct specification using Ljung-Box statistics, a p-value larger than the significance level (0.05) suggests that the model is deemed sufficiently specified in terms of residual autocorrelation (white noise). The test results can be seen in the Table 4, based on the p-value which is more than 0.05, it can be ascertained that the model is white noise and passes the Ljung-Box test.

Figure 5 displays the forecast results using ARIMA(1,1,4) for a period of 43 days. The average sea surface height is 1.45 meters, the maximum height is 1.45 meters, and the minimum height is 1.36 meters.

Long Short-Term Memory (LSTM)

Training and testing datasets compose the two sets of datasets used in this study. The last 30 data are the training data, whilst the first 718 data make up the training dataset. Eighty



(a) (b)

Figure 4. Correlogram: (a) Actual data ;(b) Data on first difference.

Table 3. Selection of the best Model and AICc Value

ARIMA Model	AICc
ARIMA(2,1,2) (with drift)	-1225.32
ARIMA(0,1,0) (with drift)	-1192.89
ARIMA(1,1,0) (with drift)	-1190.79
ARIMA(0,1,1) (with drift)	-1190.93
ARIMA(0,1,0)	-1194.90
ARIMA(1,1,2) (with drift)	-1231.71
ARIMA(0,1,2) (with drift)	-1188.94
ARIMA(1,1,1) (with drift)	-1228.88
ARIMA(1,1,3) (with drift)	-1236.59
ARIMA(0,1,3) (with drift)	-1187.68
ARIMA(2,1,3) (with drift)	-1238.90
ARIMA(2,1,4) (with drift)	-1245.59
ARIMA(0,1,4) (with drift)	-1230.72
ARIMA(1,1,5) (with drift)	-1247.13
ARIMA(0,1,5) (with drift)	-1240.12
ARIMA(2,1,5) (with drift)	-1243.79
ARIMA(1,1,4)	**-1249.59**
ARIMA(0,1,4)	-1232.75
ARIMA(1,1,3)	-1237.97
ARIMA(2,1,4)	-1247.51
ARIMA(1,1,5)	-1248.90
ARIMA(0,1,3)	-1189.70
ARIMA(2,1,3)	-1240.93
ARIMA(2,1,5)	-1245.77

Table 4. Ljung-Box test statistics

Model	x-Squared	*p-value*
ARIMA(1,1,4)	0.0032	0.9546

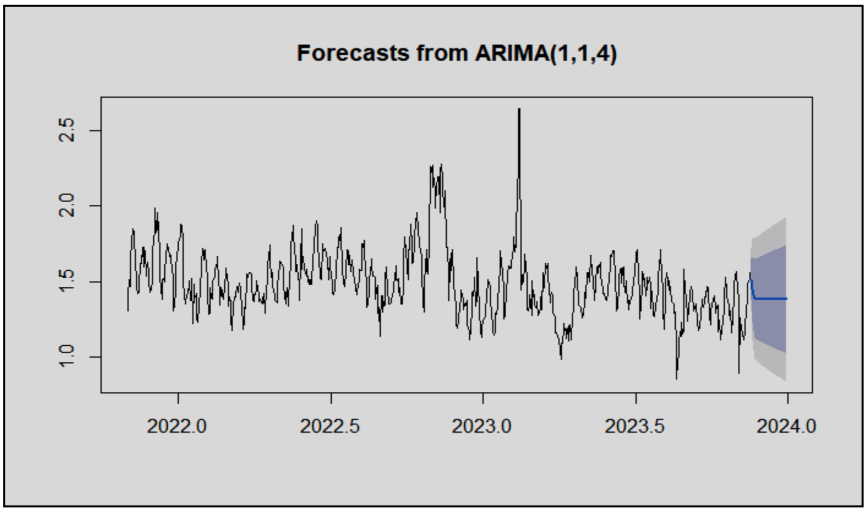


Figure 5. Forecasting with model ARIMA(1,1,4).

percent training data and twenty percent testing data are employed in certain other investigations. (Pramesti *et al.*, 2022; Gülmez, 2023; Pierre *et al.*, 2023), then there are also studies that use a composition of 70% training data and 30% testing data (Faishol *et al.*, 2020). However, this research does not use a combination of these compositions with the intention that the more training data, the more accurate the model will be in capturing patterns in the data, so it is expected that the model formed will be better. Processing is done using the Python programming language along with the Keras package, tensorflow, and others. In LSTM, processing is done on a certain amount of data to make predictions on the next data. Meanwhile, in this research, processing is carried out on 30 data to get one next data with a total of 50 batches and Epoches. Batch size indicates the number of training data samples to be included before updating the weights of an artificial neural network. In other words, it determines how much data is processed in one iteration during the model training process, before updating the model parameters. (Goodfellow *et al.*, 2016). While epoch represents one full iteration through the entire training data. Any modeling using any method always provides uncertainty which is reflected in the residual value. The LSTM method also has such a thing called a loss value. Ideally the loss value should be below 0.01 (Syahram *et al.*, 2021). Figure 6 shows the loss value graph of the LSTM model.

The loss value reflects the extent of the difference between the predicted value generated by the model and the actual value. According to Figure 6, it can be seen that the loss value decreases as the number of iterations increases, indicating an improvement in the model's performance in making predictions. This suggests that the model progressively improves its accuracy over time and iterations, resulting in predictions that are closer to the true value.

With a number that is comparatively close to the pattern seen in the testing data, the model appears to be able to predict it at first glance. The LSTM model is one of the machine learning techniques intended to learn temporal patterns and capture nonlinear relationships and store them as a memory to create accurate output. In Figure 7 can be seen that the prediction results have a relatively close pattern with the actual data, even though the the predicted value are not exactly the same as the original values or not very accurate. This inaccuracy may be due in part to parts of the data containing more noise than other parts of the data, such as the pattern of sea levels that tend to be higher and more extreme in late October 2022 and mid-February 2023. This kind of pattern can cause

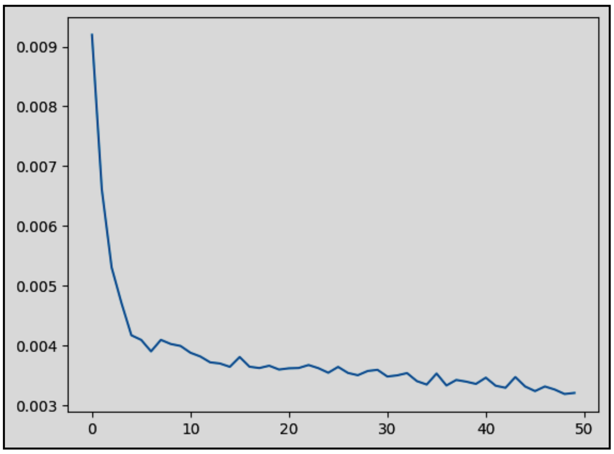


Figure 6. Loss value of the LSTM model along the y-axis and x-axis as Epoch.

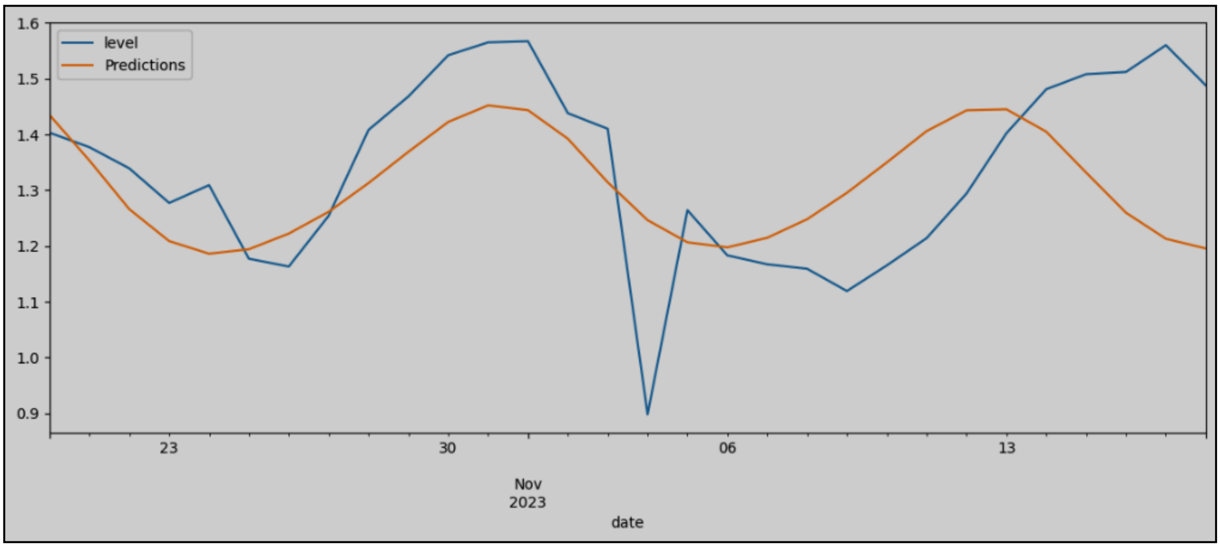


Figure 7. Comparison between actual data (blue) and LSTM predicted values (orange).

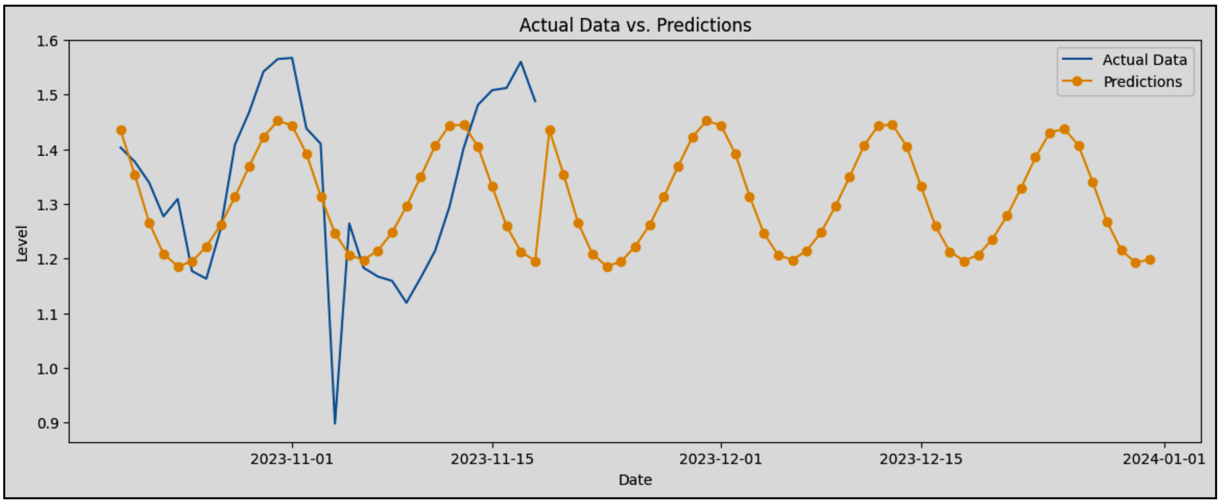


Figure 8. Forecasting results with LSTM.

the LSTM model to learn the wrong pattern and produce inaccurate predictions. In addition, the amount of data also affects the prediction results. More data will make the model learn more and produce more accurate values. Nevertheless, the LSTM model has been able to capture the patterns in the data. (Gu *et al.*, 2019). Furthermore, accuracy metrics like MSE, RMSE, MAE, and MAPE are used to assess how well the model performs in estimation. The model's MSE, RMSE, MAE, and MAPE values are 2.23%, 14.95%, 11.80, and 9.04%, respectively.

A visual representation of the forecasting performed by the model can be seen in Figure 8. The forecasting results of the LSTM model until the end of 2023 tend to have a pattern that repeats for every few days in predicting sea level. This is corresponding to the study that Gu *et al.* (2019) which states that LSTMs will tend to capture temporal patterns in data. There is also a kind of prediction that does not follow the previous pattern in the middle of the graph, this occurs because the model makes further predictions to re-form a new pattern by adjusting and continuing the pattern shape of the last few actual data values. In addition, it was also found that the sea level in Jakarta showed relative stability until the end of 2023, with an average height of about 1.39 meters and the highest value of 1.45 meters which occurred on November 30, 2023. This Figure 8 also contains prediction results that do not follow the forecasting pattern.

Comparison Between ARIMA and LSTM

A comparative analysis of the evaluation of the outcomes from the LSTM and ARIMA methods is conducted in order to accomplish the research goal, which is to demonstrate the suitable and efficient short-term sea level forecasting technique in Jakarta.

The results show that ARIMA is a more effective method because it has a lower error evaluation value as shown in the Table 5. Both forecasting methods basically have the same principle, which is to study past data to build a trend and pattern to predict future data. The ARIMA model filters out random patterns in the data and predicts trends, while LSTM studies patterns to identify nonlinear dependencies and predicts the next observation (Katambire *et al.*, 2023). Overall, both models actually predicted the sea level pattern well. However, based on these accuracy measures, ARIMA is more effective because it excels in all model performance evaluations. This may be due to the relatively small amount of data used, whereas nonlinear models in some literature will outperform ARIMA when the number of datasets is large. However, research conducted by Paliari *et al.*, (2021) shows that ARIMA still outperforms LSTM even when performed on a large number of datasets. Meanwhile, in forecasting over the next period of time, visually the LSTM can relatively predict the same recurring pattern as the trend pattern that occurred in the past, while the actual ARIMA forecasting also looks relatively good, but there is no pattern like the LSTM. In general, both models provide fairly stable sea level forecasts over the next few days. According to the prediction results, until the end of 2023, Jakarta's coastal areas are at least unlikely to experience tidal flooding due to extreme sea

Table 5. Summary statistics of evaluation comparison between models

Model	MAE	RMSE	MAPE (%)
ARIMA	7.19	10.35	4.86
LSTM	11.80	14.95	9.04

Information: MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and MAPE (Mean Absolute Percentage Error).

level rise because if referring to older historical data, every year-end period, Jakarta's sea level conditions are always at a point that tends to be high compared to other months.

CONCLUSION

Jakarta, as a city geographically adjacent to the sea, is threatened by rising sea levels due to climate change. Historical data from November 2021 to November 2023 shows that during December to January, Jakarta's sea level always experiences a significant increase so that this will certainly pose a threat of flooding for people who live near the sea. In this research, ARIMA and LSTM methods are used to analyze sea level patterns and predict sea level until the end of 2023. ARIMA (1,1,4) provides more effective prediction performance than LSTM. In predicting sea level, both models provide relatively stable forecasts, but the LSTM method is slightly better because it manages to capture the pattern for every few times. However, overall, the results from both methods implicitly state that at least Jakarta's sea level until the end of 2023 is still within safe limits with an average height of 1.39 meters so the chance of tidal flooding will be relatively small. Further studies are possible by adding exogenoues variables to create a better forecasting model.

REFERENCES

Balogun, A.L., & Adebisi, N. (2021). Sea level prediction using ARIMA, SVR and LSTM neural network: assessing the impact of ensemble Ocean-Atmospheric processes on models' accuracy. *Geomatics, Natural Hazards and Risk*, *12*(1), 653–674. https://doi.org/10.1080/19475705.2021.1887372

Bennett, W.G., Karunarathna, H., Xuan, Y., Kusuma, M.S.B., Farid, M., Kuntoro, A. A., Rahayu, H.P., Kombaitan, B., Septiadi, D., Kesuma, T.N.A., Haigh, R., & Amaratunga, D. (2023). Modelling compound flooding: a case study from Jakarta, Indonesia. *Natural Hazards*, *118*(1), 277–305. https://doi.org/10.1007/s11069-023-06001-1

Dobre, I., & Alexandru, A.A. (2008). Modelling unemployment rate using Box-jenkins procedure. *Journal of Applied Quantitative Methods*, *3*(Quantitative Models), 156–166.

Faishol, M.A., Endroyono, E., & Irfansyah, A. N. (2020). Predict Urban Air Pollution in Surabaya Using Recurrent Neural Network – Long Short Term Memory. *JUTI: Jurnal Ilmiah Teknologi Informasi*, *18*(2), 102. https://doi.org/10.12962/j24068535.v18i2.a988

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. In *Adaptive Computatuion and Machine Learning series*. Massachusetts Institute of Technology. https://lccn.loc.gov/2016022992

Gu, J., Liang, L., Song, H., Kong, Y., Ma, R., Hou, Y., Zhao, J., Liu, J., He, N., & Zhang, Y. (2019). A method for hand-foot-mouth disease prediction using GeoDetector and LSTM model in Guangxi, China. *Scientific Reports*, *9*(1), 17928. https://doi.org/10.1038/s41598-019-54495-2

Gülmez, B. (2023). Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. *Expert Systems with Applications*, *227*(January), 120346. https://doi.org/10.1016/j.eswa.2023.120346

Hansun, S., & Young, J.C. (2021). Predicting LQ45 financial sector indices using RNN-LSTM. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00495-x

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

https://doi.org/10.1162/neco.1997.9.8.1735

Katambire, V.N., Musabe, R., Uwitonze, A., & Mukanyiligira, D. (2023). Forecasting the Traffic Flow by Using ARIMA and LSTM Models: Case of Muhima Junction. *Forecasting*, *5*(4), 616–628. https://doi.org/10.3390/forecast5040034

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Leontinus, G. (2022). Program dalam Pelaksanaan Tujuan Pembangunan Berkelanjutan (SDGs) Dalam Hal Masalah Perubahan Iklim di Indonesia. *Jurnal Samudra Geografi*, *5*(1), 43–52. https://doi.org/10.33059/jsg.v5i1.4652

Makridakis, S. 1992. Metode dan Aplikasi Peramalan (terjemahan). Jakarta: Erlangga.

Paliari, I., Karanikola, A., & Kotsiantis, S. (2021). A comparison of the optimized LSTM, XGBOOST and ARIMA in Time Series forecasting. *2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 1–7. https://doi.org/10.1109/IISA52424.2021.9555520

Pierre, A.A., Akim, S.A., Semenyo, A.K., & Babiga, B. (2023). Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches. *Energies*, *16*(12). https://doi.org/10.3390/en16124739

Pramesti, D.D., Novitasari, D.C.R., Setiawan, F., & Khaulasari, H. (2022). Long-Short Term Memory (Lstm) for Predicting Velocity and Direction Sea Surface Current on Bali Strait. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *16*(2), 451–462. https://doi.org/10.30598/barekengvol16iss2pp451-462

Ramadhan, A.W., Adytia, D., Saepudin, D., Husrin, S., & Adiwijaya, A. (2021). Forecasting of Sea Level Time Series using RNN and LSTM Case Study in Sunda Strait. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, *12*(3), 130. https://doi.org/10.24843/lkjiti.2021.v12.i03.p01

Sugiyono. (2014). *Metodologi Penelitian Kuantitatif, Kualitatif dan R & D*. Alfabeta.

Sumartapraja, A.R., & Christianti, D.W. (2022). The Right to Water in Jakarta: Limitation in a Sinking City. *Padjadjaran Jurnal Ilmu Hukum*, *9*(1), 67–88. https://doi.org/10.22304/pjih.v9n1.a3

Sun, Q., Wan, J., & Liu, S. (2020). Estimation of Sea Level Variability in the China Sea and Its Vicinity Using the SARIMA and LSTM Models. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 3317–3326. https://doi.org/10.1109/JSTARS.2020.2997817

Syahram, E. F., Effendy, M. M., & Setyawan, N. (2021). Sun Position Forecasting Menggunakan Metode RNN – LSTM Sebagai Referensi Pengendalian Daya Solar Cell. *Jurnal JEETech*, *2*(2), 65–77. https://doi.org/10.48056/jeetech.v2i2.169

Widodo, A. (2017). Analyzing Indonesia's NCICD Project to Stop the Capital City Sinking. *Otoritas : Jurnal Ilmu Pemerintahan*, *7*(2), 54–66. https://doi.org/10.26618/ojip.v7i2.769

Yamak, P.T., Yujian, L., & Gadosey, P.K. (2019). A comparison between ARIMA, LSTM, and GRU for time series forecasting. *ACM International Conference Proceeding Series*, 49–55. https://doi.org/10.1145/3377713.3377722