ATLANTIS PRESS

Advances in Engineering Research, volume 208

Proceedings of the International Conference on Innovation in Science and Technology (ICIST 2020)

Model Sea Levels Prediction With ARIMA for Coastal Area in Semarang

A. Musbikhin*
Master's in information system
Universitas Diponegoro
Semarang, Indonesia
*musbikhin.ahmad@gmail.com

Eko Sediyono
Faculty of Information Technology
Universitas Kristen Satya Wacana
Salatiga, Indonesia

Catur Edi Widodo
Faculty of Science and Mathematics
Universitas Diponegoro
Semarang, Indonesia

***Abstract*—Coastal flood is one of the most frequent disasters in coastal cities in Indonesia. Semarang the largest city in Central Java is one of the cities most affected by rising sea levels. The low land surface along the coast and decreasing every year accompanied by rising sea levels makes the city prone to tidal flooding. To predict the sea level rise, which is the main cause, this study offers data analysis using the Auto-Regressive Integrated Moving Average (ARIMA) method. The ARIMA model is divided into 3 elements, namely: Auto-Regressive (AR) model, the Moving Average (MA), and the Integrated (I) model. These three elements can be modified to form a new model in general form, it will be ARIMA (p, d, q) and this study found that model (1,1,1) is more suitable. The results of this study produce predictions that are quite accurate than can provide accurate data and information for stakeholders to prepare for disaster management and mitigation in affected areas. Besides, it can be used as a reference for making prevention for the short and medium-term in areas prone to coastal flooding.**

Keywords—ARIMA, prediction, time series, coastal flood, sea levels

I. INTRODUCTION

Flood is one of the most dangerous natural disasters with a widespread distribution in developed cities and cities with high population density throughout the region globally [1]. Global warming and climate change are causing more frequent and intense flooding in various regions around the world [2]. IPCC (Intergovernmental Panel on Climate Change) research found that there is clear evidence that climate change can be felt by increasing earth's surface temperature, melting polar ice caps, and rising sea levels. The IPCC also predicts that in the next two decades the earth's surface temperature will increase by about 0.2° C per decade which will trigger melting of ice on the earth's surface [3].

Semarang has experienced different natural disasters several times, but is relatively safe from earthquakes because it is far from friction between the earth's crust and volcanoes [4]. Disasters that often occur are landslides and floods, both tidal floods and floods due to river overflows Tidal flooding is caused by rising sea levels and land subsidence [5]. Land subsidence is mostly caused by neglect of environmental assessments or environmental assessments in policies, planning, and programs such as failed city planning and excessive groundwater extraction so that flooding is a major problem in Semarang City [4].

Flood prediction has long been used both to reduce the impact of flooding and as an early warning. Prediction is carried out by various methods such as using machine learning techniques [2], multiple regression for decision makers [6] for prediction, even using radar satellites for Geographic Information System (GIS) modeling in estimating potential flooding in future [7]. Tidal flood is a unique type of flood because it is a type of flood that has a fixed factor, namely rising sea levels which can be caused by various things. Therefore, the optimal prediction is needed to be able to handle this problem [8].

Tide predictions for sea level can be useful for knowing the potential for tidal flooding in the future so that stakeholders can prepare themselves for disaster management such as disaster mitigation, regional planning, and anticipation of tidal flooding to minimize casualties and both material and non-material losses. From existing historical data, it can be used as a reference for prediction calculations using various models and methods to produce fairly accurate predictions. One model that is suitable for analyzing time series data is ARIMA (Autoregressive Integrated Moving Average). The data modeling can be accurately used for short-term predictions so that it can be used as an early warning model with an

appropriate information system. An accurate prediction of the tide of sea level will be very useful for mitigating the potential for tidal flooding in coastal areas where the land surface is lower than sea level.

ARIMA modeling has often been used for various predictions such as prediction of Seer and Mullet fish catch based on data from previous annual catches. With data-deficient fisheries situations, this method can support the evaluation of potential fisheries production for decision-making and management [9]. Meanwhile, ARIMA modeling is also used to monitor rice price movements so that rice trading can be supervised and controlled [10]. ARIMA is also applied in predicting energy consumption and greenhouse gas (GHG) emissions to determine trends in energy use and their effects on the environment [11].

II. Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) or also known as the Box-Jenkins time series is a derivative of the time series algorithm which completely ignores the independence of variables in forecasting [12]. ARIMA uses the past and present values of the dependent variable to produce accurate short-term predictions, but for long-term predictions the accuracy of the prediction is not good. The purpose of ARIMA is to determine a good statistical relationship between the predicted variables and the historical value of these variables so that predictions can be made with this model [11].

ARIMA is a combination statistical method and the development of the time series method. This method is quite effective for making predictions using time series variables as the main analysis data such as natural disaster data, stocks, food prices, sales, and demand data. In the short-medium term prediction, the accuracy of the prediction results using ARIMA is one that is quite accurate but for long-term accuracy the predictions will tend to be flat and inaccurate [9].

ARIMA model is divided into 3 elements, namely: autoregressive model (AR), moving average (MA), and Integrated (I). these three elements can be modified to form a new model. for example, the autoregressive model and moving average (ARMA). however, if you want to make it in its general form, it will be ARIMA (p, d, q). p represents the order AR, d represents the order I and q represents the order MA [13].

A. Autoregressive (AR)

The general form of the autoregressive model with the order p (AR (p)) or the ARIMA model (P, 0,0) is stated as follows:

$$X_t = \mu' + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + e_t \tag{1}$$

which:

μ' : Constant

$\emptyset_p$: autoregressive parameters to-p

e_t : Error value t

The meaning of autoregressive is that the x value is influenced by the x value of the previous period to the p-period. So, what matters here is the variable itself.

B. Integrated (I)

The general form of the integrated model with the order d (I (d)) or the ARIMA model (0, d, 0), integrated here is to express the difference from the data. it means that in making ARIMA models the mandatory requirements that must be met are stationary data. if the data is stationary at the level, then the order is equal to 0, but if it is stationary at the first difference then the order is 1, and so on.

C. Moving Average (MA)

The general form of the moving average model with the order q (MA (q)) or the ARIMA model (0.0, q) is stated as follows:

$$X_t = \mu' + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-k} \tag{2}$$

which:

μ' : Constant

θ_1 until θ_t are moving average parameter

e_{t-k} : Error value t-k

The purpose of the moving average is that the value of the x variable is affected by the error of the x variable.

Apart from the elements forming the third ARIMA above, this method is also divided into two models, namely, ARIMA model without seasonality and seasonal ARIMA model. ARIMA model without seasonality is not affected by season time factor. The general form of ARIMA (p, d, q) can be expressed in the following equation:

$$(1 - B)(1 - \emptyset_1 B)X_t = \mu'(1 - \theta_1 B)e_t \tag{3}$$

or

$$Y_t = \mu + Y_{t-1} + \emptyset(Y_{t-1} - Y_{t-2}) - \theta e_{t-1} \tag{4}$$

Meanwhile, seasonal ARIMA is an ARIMA model which is influenced by the time of season factor. this model is usually called the Season ARIMA (SARIMA). The general form is expressed in the following equation:

$$(1 - B)(1 - \emptyset_1 B)X_t = \mu' + (1 - \theta_1 B)e_t \tag{5}$$

III. Data Analysis

Sea level prediction is a complex one and requires a lot of data in the process. In this research, the data used in this study are sea level data which is the premiere data and the data on the height of the sub-districts in Semarang City which are in direct contact with the coastline. The data is in the form of the

average sea level taken over a period of seven years from 2013 to 2019 with a total of 84 data. The data is taken from the intergovernmental Oceanographic Commission (IOC) database of the tidal wave measuring sensor located at the Port of Tanjung Mas Semarang.

After the data is collected, it is analyzed whether the data can be used for the prediction process using ARIMA remote by normalizing the data. The requirement for data to be used with the time series method is to check whether the data is stationary or not. There are several ways to do this, among others, with differencing. The data analysis process uses Minitab19 software because the program has easy-to-use navigation and can analyze data with various statistical analyzes and data management well.

In prediction using the ARIMA model, there are four stages that must be passed, namely, (1) Identification (2) Estimation (3) Diagnostic Check (4) Forecasting.

A. Identification

The process of checking the data is whether the data is stationary or not stationary, the first step is to change the data table into a diagram as in Figure. 1, the time series plot for data for the period 2013 - 2019 shows that the data is not stationary, shown by graphical data that is not consistent and random.

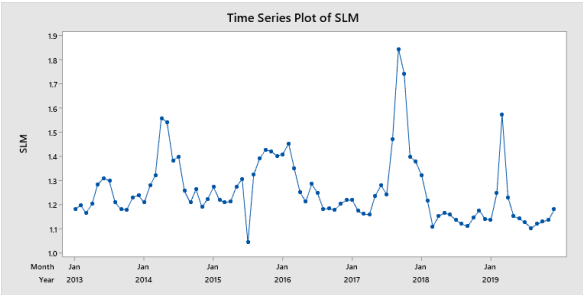

Fig. 1. Time series plot sea level data.

To make non-stationary data into stationary data, differencing is carried out. The process is carried out using the backward shift operator B, with the following equation.

$$BY_t = Y_{t-1} \tag{6}$$

The B notation attached to Y_t has the effect of sliding data back one period. If the plot results show the data is still not stationary, the differencing process is repeated until the plot results show stationary. Next, look for stationary data with respect to the range and average with the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) with the following equation.

ACF value in the lag-k function.

$$r_k = \frac{\sum_{t=1}^{n-k} Y_t - Y\, Y_{t+k} - Y}{\sum_{t=1}^{n} Y_t - Y^2} \tag{7}$$

PACF value in the lag-k function.

$$\emptyset_{kk} = \mathrm{Corr}\left(Z_t, Z_{t-k} | Z_{t-1}, Z_{t-2}, \ldots, Z_{t-k+1}\right) \tag{8}$$

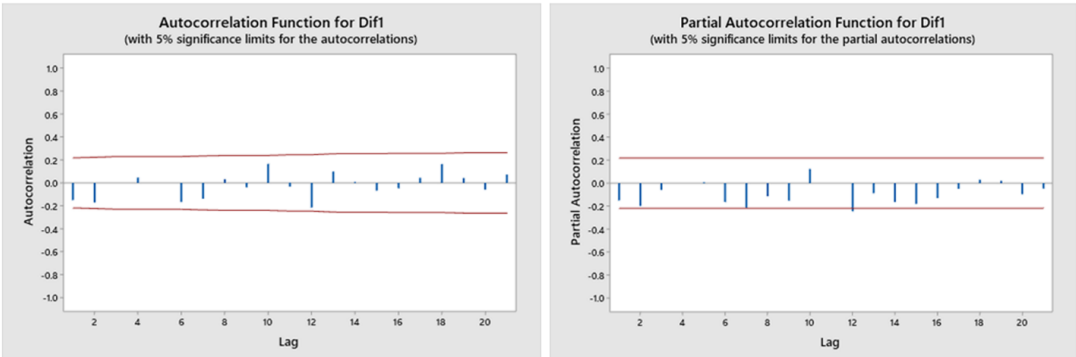

Fig. 2. ACF and PACF diagram after differencing data.

In figure 2 the data is stationary with the mean and variance because there is no lag out of the confident interval (red lines). After obtaining the ACF and PACF plots, no more lags come out, then the next step is to analyze the model. An ARIMA model cannot contain *i* so then it can fill the order *p* and *q* with a value of 1, then the tentative models obtained are models (1,1,0), (0,1,1) and (1,1,1) of three tentative models above can then be tested to find the best model that can be used. The best model is a model with high significance and has the smallest Mean Square Error (MSE) value.

B. Estimation and Diagnostic Check

From the results of the calculation of the three ARIMA models (1,1,0), (0,1,1) and (1,1,1) it can be made in the comparison table which can be seen in Table.1 seen from the p

value will show a significant number if the P-Value <Alpha 5% (0.05).

TABLE I. COMPARISON MODEL

Model	Parameter	P-Value	Description	MSE
AR (1)	(1,1,0)	0.951	Insignificant	0.0116510
MA (1)	(0,1,1)	0.900	Insignificant	0.0116504
ARIMA (1,1,1)	AR 1	0.000	Significant	0.0102542
	MA 1	0.000	Significant	

Seen in Table 1, it can be concluded that the ARIMA model (1,1,1) is the most appropriate model to use because it sees the p value which shows the number 0 which indicates a high significant value, while the two models show a number that is not significant to the alpha value. In addition, the Mean Square Error (MSE) value of the ARIMA model (1,1,1) shows the lowest value of the three models above. So, it can be concluded that this model is the most appropriate model for calculating sea level prediction.

C. Forecast Model

From the results of data analysis, the ARIMA model (1,1,1) is the most appropriate model to use for prediction of sea level. Experimental data calculations using sea level data from 2013-2019 show that the prediction accuracy reaches more than 98% for the prediction of the first period, while the prediction for the next period has decreased the percentage of accuracy as can be seen in Table 2 and Table 3 regarding the results of predictions and comparisons with real data.

TABLE II. COMPARISON OF PREDICTION RESULTS WITH REAL DATA FOR 2017

No	Forecast	Real	Period	Change	Error %	Accuracy %
1	1.2336	1.220	1/1/2017	0.0136	1.1	98.9
2	1.2489	1.174	2/1/2017	0.0749	6.4	93.6
3	1.2649	1.162	3/1/2017	0.1029	8.9	91.1
4	1.2817	1.159	4/1/2017	0.1227	10.6	89.4
5	1.2993	1.237	5/1/2017	0.0623	5.0	95.0
6	1.3177	1.280	6/1/2017	0.0377	2.9	97.1

TABLE III. COMPARISON OF PREDICTION RESULTS WITH REAL DATA FOR 2019

No	Forecast	Real	Period	Change	Error %	Accuracy %
1	1.1054	1.138	1/1/2019	0.0326	2.9	97.1
2	1.071	1.248	2/1/2019	0.177	14.2	85.8
3	1.0378	1.573	3/1/2019	0.5352	34.0	66.0
4	1.0057	1.231	4/1/2019	0.2253	18.3	81.7
5	0.9747	1.152	5/1/2019	0.1773	15.4	84.6
6	0.9448	1.142	6/1/2019	0.1972	17.3	82.7

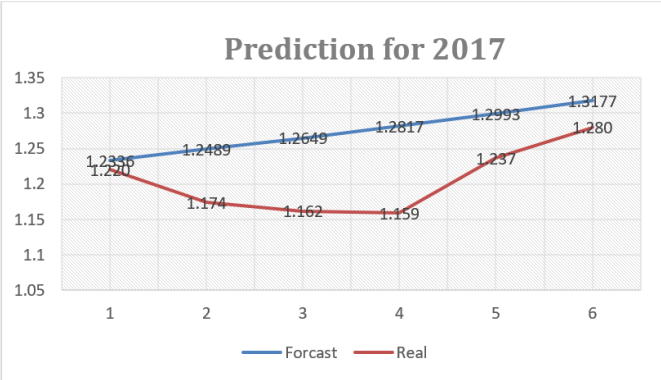

Fig. 3. Comparison between prediction and real data.

From the prediction experiment using data from 2013-2016 to predict the 2017 monthly average height and the results can be seen in Table 2 and Figure 3, it can be concluded that the prediction results show good accuracy although not evenly distributed in each period. The best accuracy results are achieved in the first period with an accuracy of up to 98.9% with an error of 1.1%. Meanwhile, in the second experiment for predicting 2019 using data from 2013-2018, the highest accuracy value reached 97.1% in the first period and decreased in the next period. The worst accuracy occurs in a period where the accuracy only reaches 66% with an error of 34% from the analysis that the poor prediction results are due to the lack of supporting data variables in the prediction process.

IV. CONCLUSION

Sea level prediction is a complex prediction, many factors cause sea level tides such as global warming, extreme weather, the revolution of the moon to the earth. In this case study studied only uses single data variable and single mathematic method, namely the monthly average altitude data so that some cases in certain periods experience bad accuracy values, but in general the results of these predictions can help stakeholders in make policies on coastal area zoning and tidal flood mitigation. To produce the maximum accuracy value, even to predict future tidal flooding, further and complex research is needed, such as research on land subsidence, weather prediction and global warming.

REFERENCES

[1] Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae, "Global flood risk under climate change," Nat. Clim. Chang., vol. 3, pp. 816–821, 2013.

[2] J. Noymanee, N.O. Nikitin, and A.V. Kalyuzhnaya, "Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin," Procedia Comput. Sci., vol. 119, pp. 288–97, 2017.

[3] O. Edenhofer, R. Pich-sMadruga, and Y. Sokona, Climate Change 2014: Mitigation of Climate Change Assess. Cambridge: Cambridge University Press, 2015.

[4] A. Ismanto, A. Wirasatriya, M. Helmi, A. Hartoko and P. Prayogi, "Model Sebaran Penurunan Tanah di Wilayah Pesisir Semarang,"

International Journal of Marine Sciences (IJMS), vol. 14, no. 4, pp. 189-196, 2009.

[5] A. Susilowardhani, "The Potential of Strategic Environmental Assessment to Address the Challenges of Climate Change to Reduce the Risks of Disasters: A Case Study from Semarang, Indonesia," Procedia - Soc. Behav. Sci., vol. 135, pp. 3–9, 2014.

[6] J.T. De Castro, G.M. Salistre, Y.C. Byun, and B.D. Gerardo, "Flash flood prediction model based on multiple regression analysis for decision support system," Lect. Notes Eng. Comput. Sci., vol. 2 802–807, 2013.

[7] M.S.Chubey and S. Hathout, "Integration of RADARSAT and GIS modelling for estimating future Red River flood risk," GeoJournal, vol. 59, pp. 237–246, 2004.

[8] I. Buchori, A. Pramitasari, A. Sugiri, M. Maryono and Y. Basuki, "Adaptation to coastal fl ooding and inundation : Mitigations and migration pattern in Semarang City, Indonesia," Ocean Coast. Manag., vol. 163, pp. 445–455, 2018.

[9] J.J. Selvaraj, V. Arunachalam, K.V. Coronado-Franco, L.V. Romero-Orjuela and Y.N. Ramírez-Yara, "Time-series modeling of fishery landings in the Colombian Pacific Ocean using an ARIMA model," Reg. Stud. Mar. Sci., vol. 39, pp. 101477, 2020.

[10] M. Ohyver and H. Pudjihastuti, "Arima Model for Forecasting the Price of Medium Quality Rice to Anticipate Price Fluctuations," Procedia Comput. Sci., vol. 135, pp. 707–711, 2018.

[11] P. Sen, M. Roy and P. Pal, "Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization," Energy, vol. 116, pp. 1031–1038, 2016.

[12] M.F. Akhter, D. Hassan and S. Abbas, "Predictive ARIMA Model for coronal index solar cyclic data," Astron. Comput., vol. 32, pp. 100403, 2020.

[13] Q. Wang, X. Song, and R. Li, "A novel hybridization of nonlinear grey model and linear ARIMA residual correction for forecasting U.S. shale oil production," Energy, vol. 165, pp. 1320–1331, 2018.