Wind Wave Prediction by using Autoregressive Integrated Moving Average model : Case Study in Jakarta Bay

M. Alif Rizal Yonanta AR[#1], Didit Adytia[#2], Nugrahinggil Subasita [*3]

 # School of Computing, Telkom University
Jl. Telekomunikasi 1 Bandung 40257, Indonesia

[1] alifyonanta@student.telkomuniversity.ac.id
[2] adytia@telkomuniversity.ac.id

[*]*Next Waves-ID*
Jl. Karapitan 1 Bandung 40261, Indonesia
[2] nunu@nextwaves-id.com

Abstract

Prediction of wind wave is highly needed to support safe navigation, especially for ship. Besides that, loading and unloading activities in a harbour, as well as for design purpose of coastal and offshore structures, data of prediction of wave height are needed. Based on its nature, the wind wave has random behaviour that is highly depending on behaviour of wind as the main driving force. In this paper, we propose a prediction method for wind wave by using Autoregressive Integrated Moving Average or ARIMA. To obtain historical data of wind wave, we perform wave simulation by using a phase-averaged wave model SWAN (Simulating Wave Near Shore). From the simulation, time series of wind wave is obtained. The prediction of wind wave is performed to calculate forecast of 24 hours ahead. Here, we perform wind wave prediction in a location in Jakarta Bay, Indonesia. We perform several combination of ARIMA model to obtain best fit model for wind wave prediction in the location in Jakarta Bay. Results of prediction show that ARIMA model give an accurate prediction especially for short term prediction.

Keywords: wind wave prediction, wave forecating, ARIMA, time series, SWAN.

I. INTRODUCTION

WIND wave is a type of wave that is generated by wind forcing. The behaviour and the generation of of the wind wave are highly depending on the wind as the main driving force. Besides that, geometry of islands and bathymetry of the ocean affect significantly the propagation of the wind wave. Prediction of wind wave is needed to provide safer navigation for ships. Besides that, the information is needed for loading and unloading activities from ships in coastal zone such as in a harbour, or in offshore such as in offshore platform. Moreover the prediction is needed for the process of structural design both in coastal as well as in offshore zone (Sorensen, 1993), (Weggel & Sorensen, 1986). A real time prediction of wave, for few hours or several days ahead, becomes necessary for engineers for planning and building coastal and offshore structures, as well as for ship navigation.

The wind wave prediction was firstly used in the early 1940s. At the time, it was used for planning the operation of naval fleets during the second world war (Mandal & Prabaharan, 2010). Currently, wind wave prediction is approached by using numerical simulation of phase-averaged wave models. In the phase-

averaged wave models, the wave energy is modelled and propagated with the wind field as the main driving force. The model takes into account the nonlinear wave-wave interaction, dissipation by wave breaking and bottom roughness, refraction by bathymetry as well as wave diffraction. The popular phase-averaged wave models that are used for wave predictions are WAVE WATCH III (Tolman et al. 1999), WAM model (WAMDI group, 1988) and SWAN model (Booij et al. 1999). Other approaches for calculating wind wave prediction by using numerical simulation with SWAN model are done by (Akpınar, van Vledder, Kömürcü, & Özger, 2012) (Moeini & Etemad-Shahidi, 2007), and by using radar that are performed by (Weissman, 1973) (Salcedo-Sanz et al., 2015).

Depending on the size of the areas to be predicted, usually the models require a high performance computing for running the simulation. Therefore, the computation effort for building the prediction system by using these models is relative high. In this paper, we propose another approach for building wind wave prediction system, i.e. the prediction is built by using a time series method called Autoregressive Integrated Moving Average or ARIMA. The method requires historical time series data as basis for calculating the wave prediction. The historical time series of wind wave is obtained by simulating wind wave by using SWAN model. To study the implementation of the method in a real situation, we choose a location in offshore of Jakarta Bay, Indonesia.

The structure of this paper is as follows. In the second section, background of wind wave prediction by using phase-averaged wave model is discussed. It is then followed by the basic idea of the method of Autoregressive Integrated Moving Average (ARIMA). In the Section 3 the building of historical wind wave time series at Jakarta Bay by using SWAN model is described, and then is followed the application of ARIMA for wind wave prediction. The Section 4 describes results and analysis of the implementation of the new method. Finally we conclude the paper in Section 5.

II. Literature Review

Prediction of wind wave traditionally is performed by firstly simulating wind as the main driving force to generate wind waves. The wind is simulated by solving a climate model. The resulting wind fields are then used as an input for wave model for simulating the propagation of wave spectra or energy. The type wave model that is commonly used for simulating wave prediction is so-called phase-averaged wave model.

The development of method for prediction of wave has been quite significant during the last four decades, from the first, the second, and the third generation wave model. The first generation numerical model can describe parametric condition of the ocean and uses wind as input (Sverdrup, 1947). This first generation exclude nonlinear wave-wave interactions, such that this generation of wave model is a relatively very simple model. An improvement is introduced in the second generation wave model, that uses wind and incorporate a simple nonlinear wave-wave interactions (Thomas & Dwarakish, 2015). In this model it is introduced a type of wave model that can be categorized as decoupled propagation model (DP), coupled hybrid (CH) and coupled discrete (CD). The second generation wave model is included in the couple hybrid (CH) and coupled discrete (CD), whereas the first generation model is the decoupled propagation (DP) model (Mandal & Prabaharan, 2010).

The third generation wave model is developed to improve mathematical model of the previous generation, such as to include refraction, shoaling, force by wind, triad wave-wave interactions, dissipation due to bottom and breaking (Thomas & Dwarakish, 2015). The most popular third generation wave model that are used for for wave simulation and prediction are WAVE WATCH III (Tolman et al. 1999), WAM model (WAMDI group, 1988) and SWAN model (Booij et al. 1999). These models require wind as an input for the simulation.

A. SWAN Model

The SWAN model was introduced by (Booij et al. 1999). The model is written in a term of spectral wave action density N, i.e. the spectral wave variance density E that is normalized by frequency σ. The action balance equation of SWAN is as follows

$$\frac{\partial N}{\partial t} + \nabla \cdot \left(C_g \cdot N\right) = \frac{S}{\sigma}$$

where $S = S_{in} + S_{ds} + S_{nl}$. Here the C_g is the wave group velocity, S is souce terms (as well as sink terms), S_{in} is the source term by wind, S_{ds} is a sink term due to dissipation of energy and S_{nl} is a shift of energy due to nonlinear wave-wave interactions, see (Booij et al. 1999). SWAN model has been used for many applications for wave simulation, in offshore as well as in coastal region, as operational model as well as forecasting model, (Akpınar, van Vledder, Kömürcü, & Özger, 2012) (Moeini & Etemad-Shahidi, 2007) (Adytia et al. 2012).

B. ARIMA model

In this subsection, the concept of Autoregressive Integrated Moving Average or ARIMA will be decribed. Time series is a series of events that occur based on certain time frames (Wei, 2006). The ARIMA model is a method for obtaining a prediction of a quantity based on its historical time series data, i.e. past and present values of the quantity. The method is also referred as the Box-Jenkins method (Huang, 2013). The ARIMA model is obtained from a combination between the Autoregressive (AR(p)) method with p-order with the method of Moving Average (MA(q)) with q-orde that includes differencing. If a nonstationer data is combined with ARMA model, then ARIMA(p,d,q) is satisfied (Box, Jenkins, Reinsel, & Ljung, 2015). The formula for simple ARIMA(p,d,q), with d as the order for differencing value, is as follows

$$y_t = \sum_{i=1}^{p} \theta_i y_{t-1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1} + \varepsilon_t \tag{1}$$

where y_t is observation value at time *t*, *p* is autoregressive parameter, *q* is a parameter for moving average, ε is an error at time *t*, and ε_{t-1} is an error at time *t-j* (Cadenas & Rivera, 2010). The forecasting for one step ahead at *t+1* as an example y_{t+1} is obtained from conditional expectation from eq. (1), is given by

$$\hat{y}_{t+1} = \sum_{i=1}^{p} y_t + \theta_j \varepsilon_t$$

where y_t is the observation data at *t*, and $\hat{y}_{t+1}$ is the prediction of y_t, where $\varepsilon_t = y_t - \hat{y}_t$(Radziukynas & Klementavicius, 2014a). The paramaters *p*, *d*, and *q* in the ARIMA model is the basic problem in the model. These values can be obtained from analysis of Autocorrelations (ACF) and Partial Autocorrelations (PACF) (Box et al., 2015). The ARIMA model has many applications, such as for prediction of wind speed (Radziukynas & Klementavicius, 2014b), prediction of increase of electricity cost (Conejo, Plazas, Espinola, & Molina, 2005), prediction of commodity prices (Kohzadi, Boyd, Kermanshahi, & Kaastra, 1996), and the most popular one is for prediction of stock price (Pai & Lin, 2005) (French, Schwert, & Stambaugh, 1987) (Tseng, Tzeng, Yu, Yuan, et al., 2001). In this paper, we will use the ARIMA model for prediction of wind wave.

III. WAVE PREDICTION

A. Historical Wave Data by SWAN Model

To obtain historical data of wind wave, in this paper, we use SWAN model for simulating historical wind wave in a location in Jakarta Bay. As an input for the wave model, we use global wind data from ERA-INTERIM provided by ECMWF (European Centre for Medium-Range Weather Forecasts). The wave simulation is performed in 3 nested domains. The first one is the global domain, the second one is the intermediate domain, and the last one is the regional model in the Jakarta Bay. We simulate only wave

conditions during 2017. The wind field at 12 February 2017 is shown in top plot of Fig.1, whereas the resulting wave height at the same time is shown in lower part of Fig. 1.
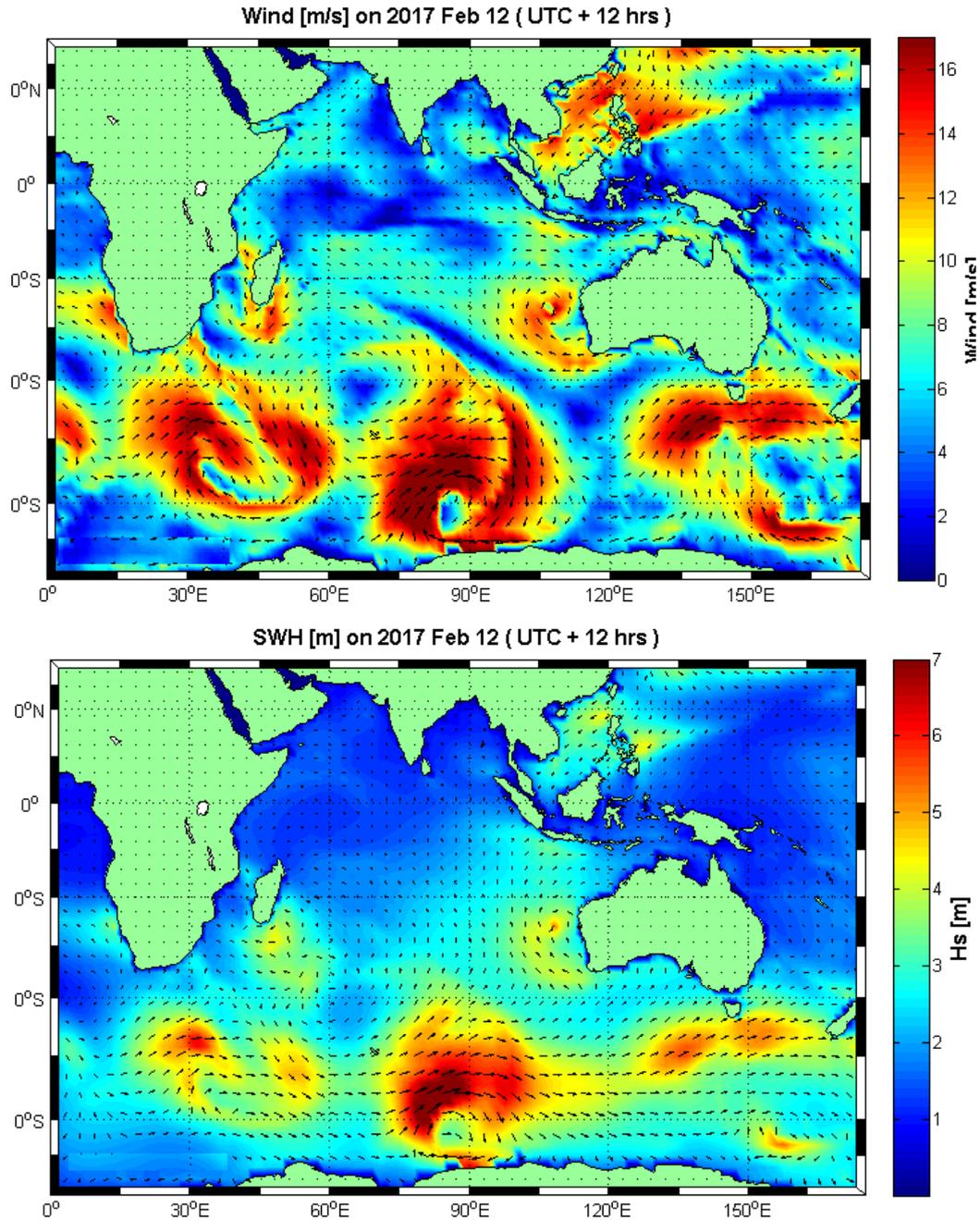

Fig. 1. The wind fields (top plot) and the resulting wave height (lower plot) at the global domain.
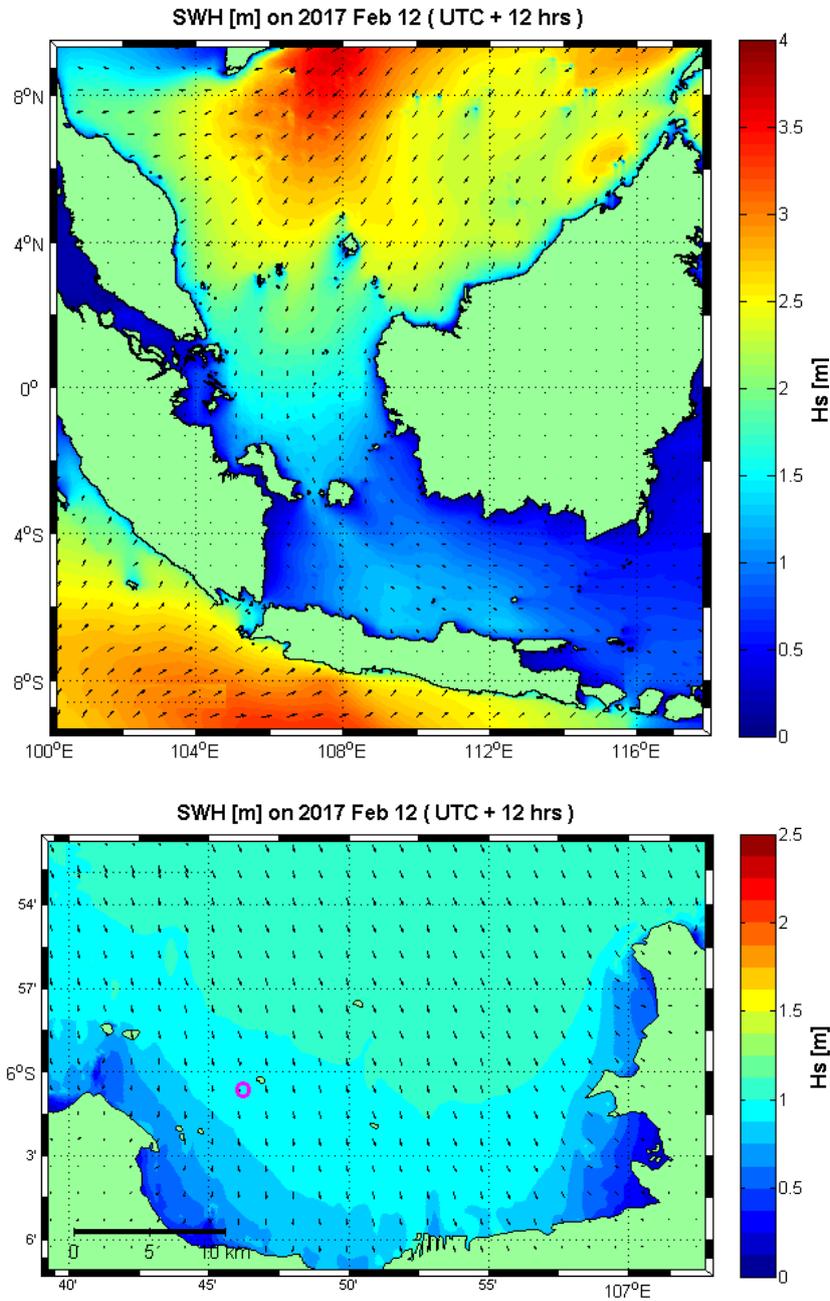
Fig. 2. The resulting wave height at the second domain (upper plot) and the regional domain (lower plot).

The wave height at 12 February 2017 in the second or intermediate domain is shown in top plot of Fig.2, whereas in the lower plot of Fig. 2 is the wave height at the Jakarta Bay. For this study, we extract a time series signal of significant wave height at 6.0108720˚ S and 106.7653970˚ E during 2017.

M. Alif R. Yonanta et.al.
Wind Wave Prediction by...

38

B. Wave Prediction by using ARIMA Model

To perform wave prediction, we use the wave data that is obtained from the SWAN simulation in previous subsection. The time series has hourly values. To be able to apply the data in the ARIMA model, the data should be converted such that it is stationer, by performing so-called differencing. The differencing is to calculate changes in the observation data. There two tests for checking stationary of a data. The first one is by simply observing the data by its graphics, and the other one is to test the data by using corelogram. After the data is stationer, the next step it is to choose maximal orde from Autoregressive (AR) and Moving Average (MA). The orde p in the AR is determined by observing the period of the time lag in Parcial Autocorrelation (PACF). The orde q in MA can be determined by observing the period of the time lag in the Autocorrelation (ACF). The flowchart that describes the prediction system for wind wave by using ARIMA model is shown in Fig. 3. In this paper, there are 1060 values of wave height for the observation data, model fitting, actual data, and model prediction. The prediction is done for the next 24 hours ahead.
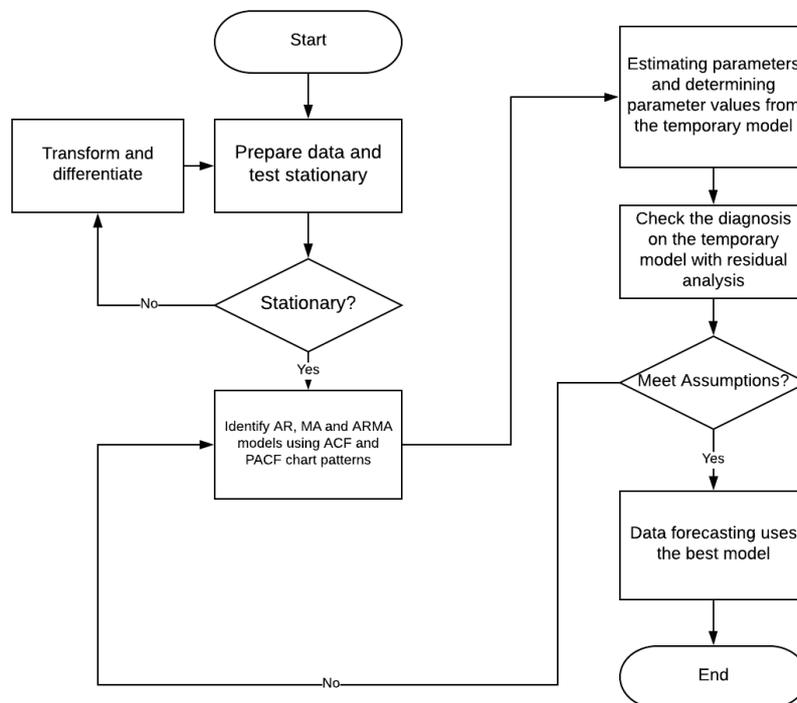

Fig. 3. Flowchart of the prediction system by using ARIMA model.

IV. Results And Discussion

The wind wave prediction by done for 24 hours ahead by using the available data that is obtained from the previous step. Here we use the ARIMA model with various combination values of p, d, and q, i.e. ARIMA(1,2,1), ARIMA(2,2,2), ARIMA(2,2,1) and ARIMA(1,2,2). These ARIMA models is calculated by using historical wave height data with temporal discretization of 1 hour. In Fig. 4, we show the trainning and prediction results by using various combination of parameters of ARIMA model, i.e. ARIMA(1,2,1), ARIMA(2,2,2), ARIMA(2,2,1) and ARIMA(1,2,2). Comparisons of the results of the ARIMA model in scatter plot are shown in Fig. 5.
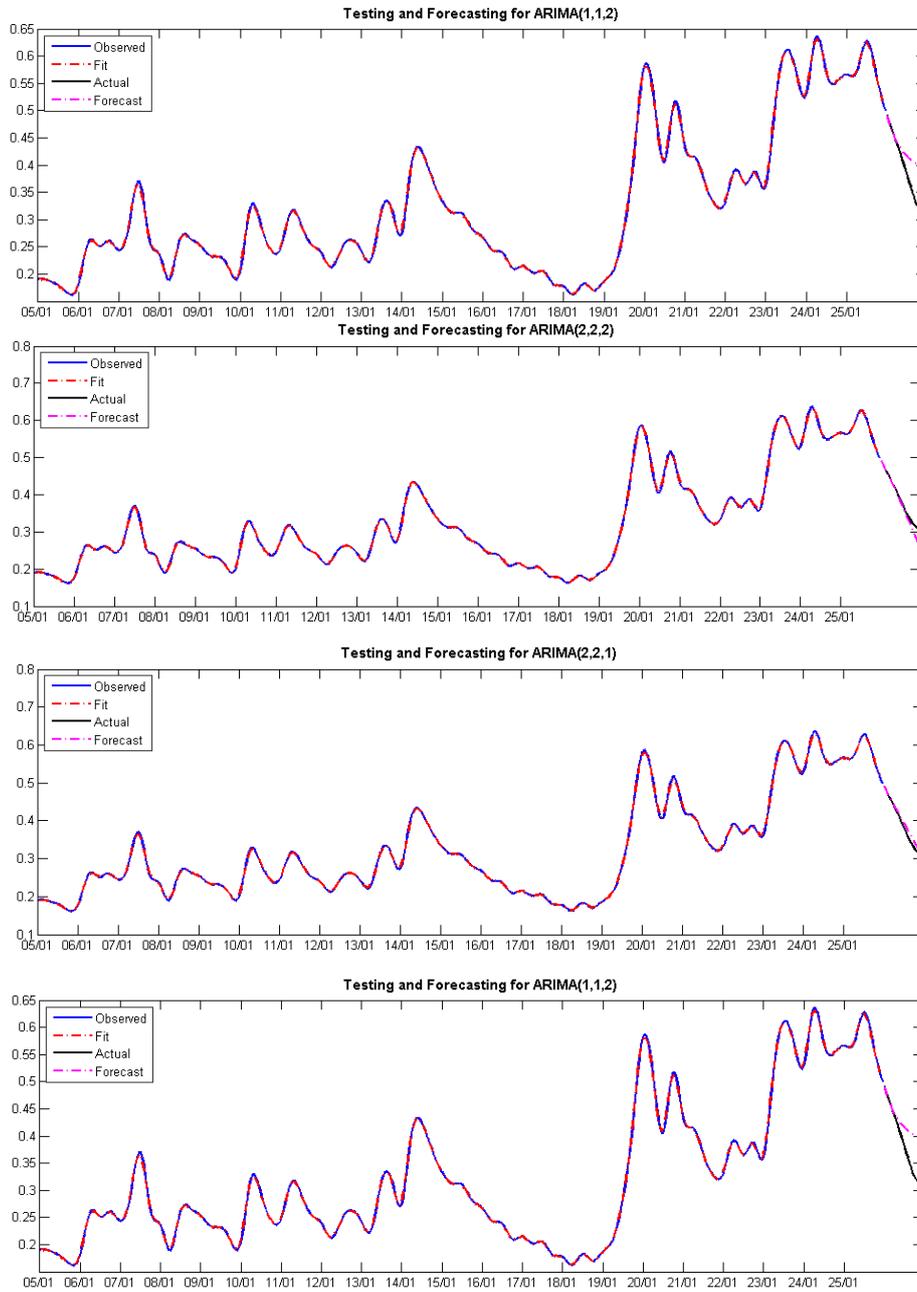
Fig. 4. The results of wave prediction by using ARIMA(1,2,1), ARIMA(2,2,2), ARIMA(2,2,1) and ARIMA(1,2,2).

M. Alif R. Yonanta et.al.
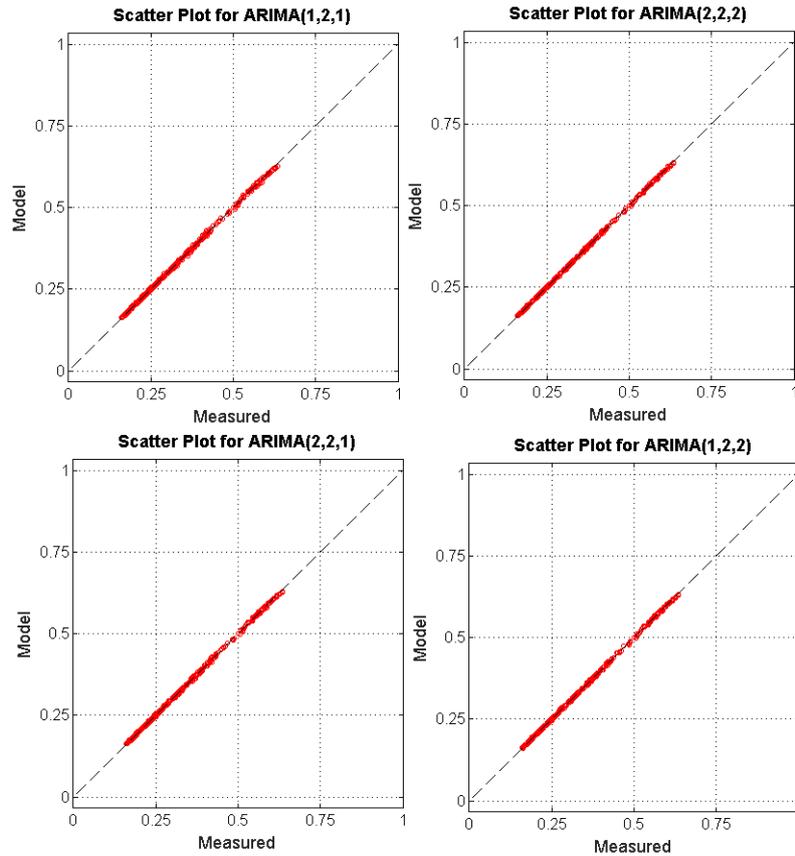Wind Wave Prediction by...

40

Fig. 5. Scatter plots of the results of wave prediction by using ARIMA(1,2,1) (top left), ARIMA(2,2,2)(top right), ARIMA(2,2,1)(lower left) and ARIMA(1,2,2)(lower right).

In Fig. 4, it can be seen the ARIMA(2,2,2) give best prediction values compare to the other setting of ARIMA model. From all settings of ARIMA, it can also be concluded that the ARIMA model fit better for short term prediction. It can be seen from Fig. 4 that qualitatively the deviation between prediction model tends to become large as the prediction time increases. To show quantitative comparison between the prediction models with the data, we use the Root Mean Square Error or RMSE that is defined as follows

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y_i} - y_i)^2}{n}}$$

where $\hat{y_i}$ and y_i are the prediction value and data at time $t = t_i$, respectively, n is the number of the data (Willmott & Matsuura, 2005). The smaller the value of the RMSE means that the prediction has smaller error compared to the data. Besides the RMSE for calculting the error, we also calculate R^2 to show the correlation between the prediction model with the data. The value of R^2 is between 0 and 1, where value 1 means that the prediction has high correlation with the data.

Summary of RMSE and R^2 is shown in Table I for each ARIMA model. From Table I, the RMSE value for ARIMA(2,2,2) has the smallest value, i.e. 0.0106 compared to the others ARIMA model, whereas the R^2 value is 0.999 which is very high. Besides that, the ARIMA(1,2,2) has also RMSE value that is close with the ARIMA(2,2,2), i.e. 0.014, with also high R^2 value, i.e. 0.999.

TABLE I
RMSE AND R^2 VALUES OF EACH ARIMA MODELS

ARIMA model	RMSE	R^2
ARIMA(0,1,0)	0,1116	0,95
ARIMA(1,1,1)	0,08	0,948
ARIMA(1,1,2)	0,0802	0,99
ARIMA(1,2,2)	0,014	0,999
ARIMA(1,2,1)	0,0128	0,997
ARIMA(2,1,1)	0,1540	0,942
ARIMA(2,2,1)	0,084	0,998
ARIMA(2,2,2)	0,0106	0,999

V. CONCLUSION

In this paper, a prediction of wind wave is approached by using stochastic process, i.e. ARIMA model. To obtain historical wind wave data, we simulate the phase-averaged wave model SWAN during 2017. The SWAN model uses wind input from ERA-INTERIM from ECMWF. We choose a location in offshore of Jakarta Bay as study area. Results of prediction show that the ARIMA(2,2,2) has smallest RMSE and highest R^2 values compared to the other ARIMA model. From the prediction, it can also be observed that the ARIMA model can calculate good prediction value for short term prediction. Here, the wind wave prediction is up to 24 hours ahead. When the prediction time is increased, the ARIMA model tend to give prediction with larger error.

REFERENCES

Akpınar, A., van Vledder, G. P., Kömürcü, M. İI., & Özger, M. (2012). Evaluation of the numerical wave model (swan) for wave simulation in the black sea. Continental Shelf Research, 50, 80–99.

Adytia, D., Ramdhani, M. & van Groesen (2012). Phase resolved and averaged wave simulations in Jakarta harbour. Proceedings 6th Asia-Pacific Workshop on Marine Hydrodynamics-APHydro, pp. 218-223.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Cadenas, E., & Rivera, W. (2010). Wind speed forecasting in three different regions of mexico, using a hybrid arima–ann model. Renewable Energy, 35(12), 2732–2738.

Conejo, A. J., Plazas, M. A., Espinola, R., & Molina, A. B. (2005). Day-ahead electricity price forecasting using the wavelet transform and arima models. IEEE transactions on power systems, 20(2), 1035– 1042.

Deka, P. C., & Prahlada, R. (2012). Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. Ocean Engineering, 43, 32–42.

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. Journal of financial Economics, 19(1), 3–29.

Huang, T. (2013). The box-jenkins methodology for time series models. America: Addison Wesley Publishing Company Inc.

Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. Neurocomputing, 10(2), 169– 181.

Mandal, S., & Prabaharan, N. (2010). Ocean wave prediction using numerical and neural network models.

Moeini, M., & Etemad-Shahidi, A. (2007). Application of two numerical models for wave hindcasting in lake erie. Applied Ocean Research, 29(3), 137–145.

Pai, P.-F., & Lin, C.-S. (2005). A hybrid arima and support vector machines model in stock price forecasting. Omega, 33(6), 497–505.

Radziukynas, V., & Klementavicius, A. (2014a). Short-term wind speed forecasting with arima model. In Power and electrical engineering of riga technical university (rtucon), 2014 55th international scientific conference on (pp. 145–149).

Radziukynas, V., & Klementavicius, A. (2014b). Short-term wind speed forecasting with arima model. In Power and electrical engineering of riga technical university (rtucon), 2014 55th international scientific conference on (pp. 145–149).

Salcedo-Sanz, S., Borge, J. N., Carro-Calvo, L., Cuadra, L., Hessner, K., & Alexandre, E. (2015). Significant wave height estimation using svr algorithms and shadowing information from simulated and real measured x-band radar images of the sea surface. Ocean Engineering, 101, 244–253.

Sorensen, R. M. (1993). Basic wave mechanics: for coastal and ocean engineers. John Wiley & Sons.

Sverdrup, H. U. (1947). Wind, sea and swell. theory of relations for forecasting. US Navy Hydrog. Office, Pub., 601, 44002E

Thomas, T. J., & Dwarakish, G. (2015). Numerical wave modelling–a review. Aquatic Procedia, 4, 443–448.

Tseng, F.-M., Tzeng, G.-H., Yu, H.-C., Yuan, B. J., et al. (2001). Fuzzy arima model for forecasting the foreign exchange market. Fuzzy sets and systems, 118(1), 9–19.

Weggel, J. R., & Sorensen, R. M. (1986). Ship wave prediction for port and channel design. In Ports' 86 (pp. 797–814).

Wei, W. (2006). Time analysis univariate and multivariate methods. America: Addison Wesley Publishing Company Inc.

Weissman, D. (1973). Two frequency radar interferometry applied to the measurement of ocean wave height. IEEE Transactions on Antennas and Propagation, 21(5), 649–656.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. Climate research, 30(1), 79–82.