Prediction of Tide level by using Holtz-Winters Exponential Smoothing: Case study in Cilacap Bay

Dwinov Satrio Wibowo
School Of Computing
Telkom University
Bandung,Indonesia
dwinovsatriow@student.telkomuniversity.acid

Didit Adytia
School Of Computing
Telkom University
Bandung,Indonesia
adytia@telkomuniversity.ac.id

Deni Saepudin
School Of Computing
Telkom University
Bandung,Indonesia
denisaepudin@telkomuniversity.ac.id

***Abstract*—Sea level rise is a phenomenon that causes the sea level to rise to some extent, And the impact from The changes in the tide level can influence flooding in coastal area that can damage the structure of the building around that area and also disturb the health of functionally linked neighboring ecosystems. But now with the development of technology and science it is possible to make some projection of the future tidal level by using a time series data, which is very important for an island country like Indonesia, this forecasted data can be used to make a planning and implementing a projects in port and coastal area. Now, there are many methods to predict the future value of several things. In this paper, the Holt-Winters Exponential Smoothings applied to forecast the tidal level in Cilacap. Then the Holt-Winters forecasting performance compared with the Autoregressive Integrated Moving Average (ARIMA), and Seasonal-Autoregressive Integrated Moving Average (SARIMA), in order to see which one that can produce the best forecast. The method performance measured by using root mean square error (RMSE) and R-Square. The Holt-Winters Exponential smoothing produces RMSE and R-Square that are better than ARIMA and SARIMA. The choice of seasonal period significantly affects the forecasting result produced by the Holt-Winters method.**

***Index Terms*—tide level, forecasting, holt-winters exponential smoothing, ARIMA, SARIMA**

I. INTRODUCTION

Sea level rise is a phenomenon that causes the sea level to rise to some extent, which can cause various kinds of problems such as a flood in a residential area close to the beach. During the 1961 to 2003 period, the average rise estimation of the global mean sea-level is 1.8 ± 0.5 mm per year, and the rise accelerated along the mid 19th - the mid 20th centuries [1].

With the development of technology and science, now we can use many data to make some models to forecast the future tidal level, which is very important for an island country like Indonesia [2]. Nowadays, there are many studies and experiments about methods and models to predict the tidal level's future value based on time-series data. For example, Prashant K. Srivastava et al. used the exponential leveling spatial state model and the Autoregressive Integrated Moving Average (ARIMA) method to predict the sea-level rise in Arabi [3]. Sepideh Karimi, Ozgur Kisi, et al. used Neuro-fuzzy techniques and neural networks to predict the rise of sea level in Darwin port in Australia [4]. Ghorbani, Mohammad Ali, et al. compared the performance of genetic programming methods with ANN (artificial neural networks) to predict sea-level rise [5].

There are many forecasting methods, such as a statistical method like Holt-Winters Exponential Smoothing, ARIMA, Seasonal-Autoregressive Integrated Moving Average (SARIMA), and Deep Learning methods, by using the help of Artificial Intelligence (AI). The Holt-Winter Exponential Smoothing widely used to forecast business data that contain trends and seasonality. ARIMA and SARIMA is a model that comes from the combination of the autoregressive (AR) and moving average (MA) models approach. The predicted value from this method result highly depends on the historical values and moving average. This method also assumes that the data is stationary, which means the mean and variance of the data do not change in time, and future value results generated from SARIMA got another aspect that needs to be considered, and that is the seasonality of the data [6].

Like ARIMA and SARIMA, the Holt-Winters method is also sensitive to unusual events or outliers. The outliers can affect the smoothing values since they depend on the past and the current data. However, before selecting which method to choose, we must check whether there are particular behavioral components present within the time-series data, such as a trend seasonality [7]. The content of this paper will discuss the comparison performances of Holt-Winters Exponential Smoothing vs. ARIMA and SARIMA forecasting method to forecast Cilacap sea level rise seventh days ahead by comparing their error.

II. RELATED WORK

A. Sea Level

The changes in the tide level can influence flooding and impact the structure, growth, aquatic ecosystem, and the health of functionally linked neighboring ecosystems [4]. Based on that, the tidal level data is needed for planning and implementing projects in port and coastal area development [8]. That is why tidal level forecasting is vital for future planning and development of coastal and port areas and mitigating its serious consequences.

Some research study has conducted to predict tidal level by using the pass recorded tidal level data. There are many

978-1-7281-8235-3/20/$31.00 ©2020 IEEE
Authorized licensed use limited to: University of Exeter. Downloaded on November 14,2020 at 11:01:13 UTC from IEEE Xplore. Restrictions apply.

methods such as ARIMA, SARIMA, and Holt-Winters Exponential Smoothing that can make a future projection of a time series data. ARIMA is a combination of the autoregressive and moving average term. The autoregressive term represents the lags of the differenced data. In contrast, the lags of the forecast errors represented by the moving average and SARIMA is an extended technique of the ARIMA method that can fully consider the seasonal characteristic of time series data [3], [9]. In 2018 Fernandez et al. applied a seasonal ARIMA model to make a future projection of Manila South Harbor mean sea level [10]. Sun et al. combine a SARIMA and Long short-term memory(LSTM) model to make a short time tidal level variability in china sea, where the Sarima method used to forecast the seasonal and trend terms of the sea level while LSTM used to predict the random terms. The study indicates that the SARIMA+LSTM works well for forecasting short time sea level variation with centimeter-levels precision [9].

The Holt-Winters exponential smoothing is an extension from the holt method used to predict data with both trend and seasonality. It has two types, additive and multiplicative, depending on the characteristics of the data [11]. The Holt-winters method used to predict Caspian sea level anomaly by Moslem Imani et al. (2013). From their study Moslem Imani, et al. conclude that the method performs well in projecting the future Caspian sea level anomaly, and the main advantages of the method are its relative simplicity compared to other complicated projection methods such as ANN [12]. Besides that Holt-winters method has also been used by Howard Grubb et al. in 2001 to do a long term forecasting of UK air passanger [13].

III. System Model

In this study, Holt-winters, ARIMA, and SARIMA methods used to make a future projection of tide level data. First, the data set containing many missing values are preprocessed and also detrended. Then forecasting models of the Holt-Winters Exponential Smoothing, SARIMA, and ARIMA are constructed, and their results are compared to each other. The steps of the methodology described in Figure 1.

A. Sea Level Data

The data used for this study come from http://ioc-sealevelmonitoring.org in Cilacap, Indonesia. The location of Cilacap shown in Figure 2. The data contains the pressure quantity of sea level data per 5 minutes from January 2019 to June 2019, as shown in Figure 3. The sea level data of Cilacap contain some outliers and missing values and showed both seasonality and trend.

B. Preprocessing

The preprocessing step consists of handling the missing values in the original data by applying linear interpolation and converting it into hourly data, the graph of hourly data obtained from preprocessing depicted in Figure 4.

The next process is to detrend the data by removing a trend of the time series data. Figure 5 is the plot of the detrended data.

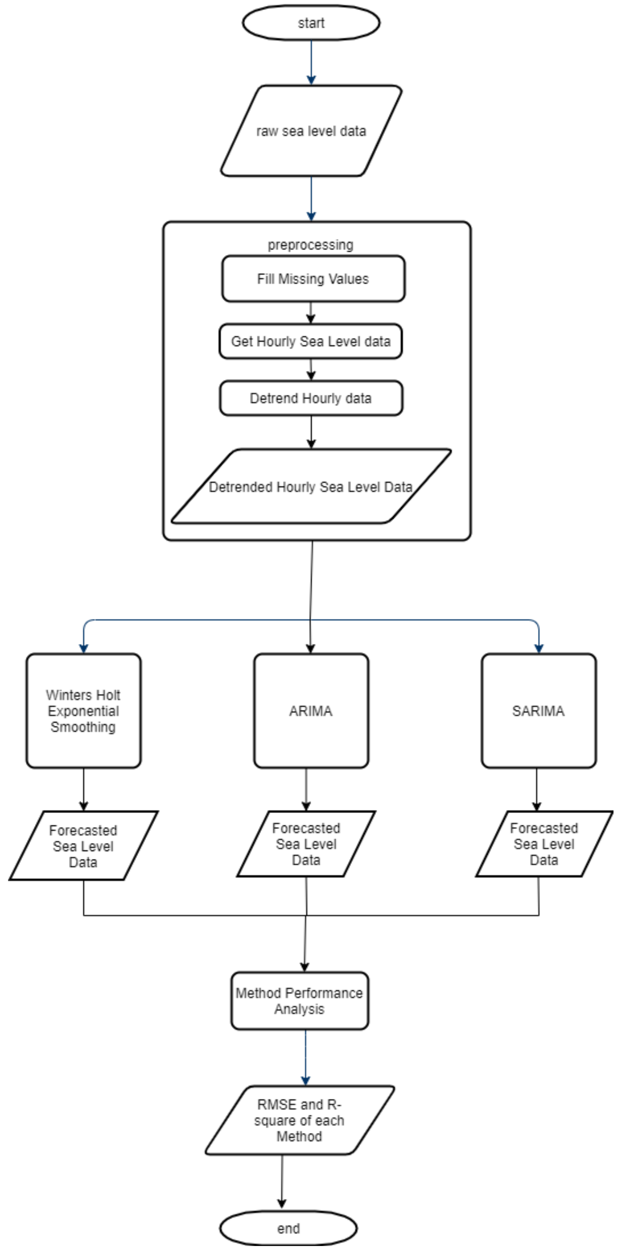

Fig. 1. Flow diagram of tide level forecasting.

C. Forecasting Method

a) Holt-Winters Exponential Smoothing: This method is designed for trended and seasonal time series [14]. It is an extension of the Holt method in exponential smoothing, usually applied to data with trend and seasonality behavior [15]. The method has three incorporate equations for calculating the level, the trend, and the seasonality [16]. Relatively to the seasonality type, it can be additive or multiplicative, depending on the oscillatory movement along the period. The equation of the winters is $Yt = Tt + St + \epsilon t$ when the seasonality is additive, and $Yt = TtxSt + \epsilon t$ in the multiplicative case, with Tt represent the trend, and St represents the seasonal component, andϵt represent the error with mean zero and constant variance

Authorized licensed use limited to: University of Exeter. Downloaded on November 14,2020 at 11:01:13 UTC from IEEE Xplore. Restrictions apply.

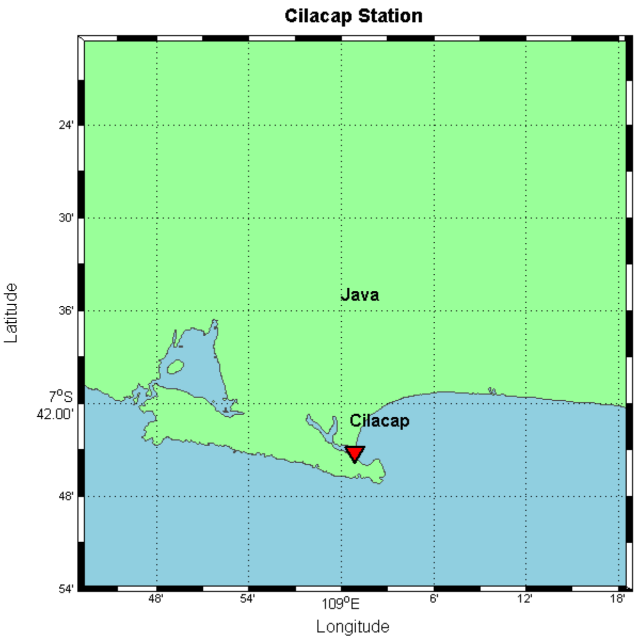

Fig. 2. Location of Cilacap, Indonesia.

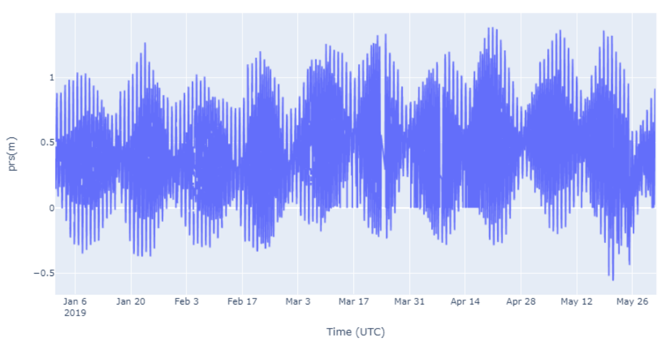

Fig. 3. The plot of original data.

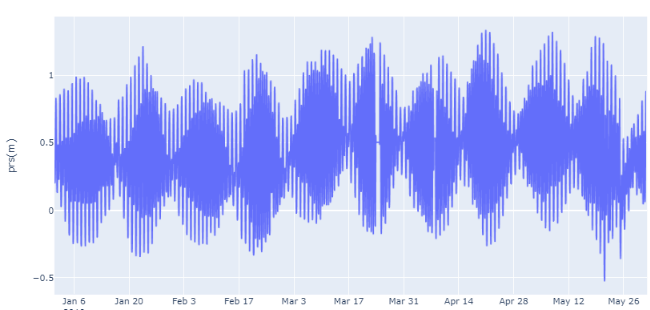

Fig. 4. The plot of hourly data.

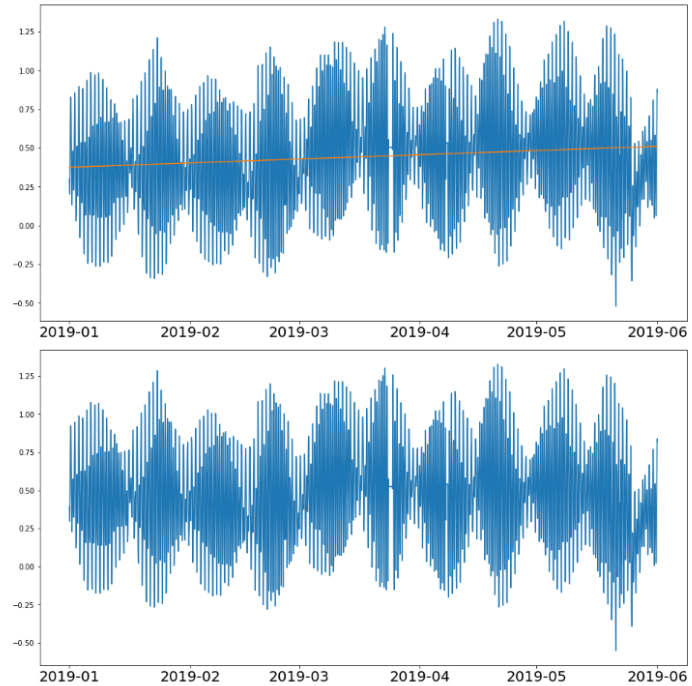

Fig. 5. Plot of detrended data.

[15]. The Additive seasonal type Holt-Winters Exponential Smoothing equations shown in the equations below:

$$level = l_t = \alpha(Y_t - s_t -_s) + (1+\alpha)(l_t -_1 + b_t -_1) \quad (1)$$

$$Trend : b_t = \beta(l_t - l_t -_1) + (1+\beta)(b_t -_1) \quad (2)$$

$$seasonal : s_t = \gamma(Y_t - l_t) + (1+\gamma)S_t -_s \quad (3)$$

The Multiplicative seasonal type Holt-Winters Exponential Smoothing equations shown in the equations below:

$$level : l_t = \alpha(\frac{Y_t}{s_t -_s}) + (1+\alpha)(l_t -_1 + b_t -_1) \quad (4)$$

$$Trend : b_t = \beta(l_t - l_t -_1) + (1+\beta)(b_t -_1) \quad (5)$$

$$seasonal : s_t = \gamma(\frac{Y_t}{l_t}) + (1+\gamma)S_t -_s \quad (6)$$

In this study, the method we use is the additive winters Holt method with detrended data as input, and parameter (α, β, γ) such as the level smoothing, trend smoothing, and seasonal smoothing and also one more parameter thas is the seasonal period which is the number of data that used to become one cycle in a season.

b) ARIMA and SARIMA: The ARIMA method is one of the most common models for time series forecasting analysis. It is an extension of the autoregressive moving average (ARMA) models that combine the AR and MA models. The ARIMA method requires a stationary time series data [17]. By adding the possibility to integrate a non-stationary process to stationary in ARMA, we can get the ARIMA models. ARIMA (p,d,q) models consist of an autoregressive parameter p, an order of integration d, and a moving average parameter q [17].

Authorized licensed use limited to: University of Exeter. Downloaded on November 14,2020 at 11:01:13 UTC from IEEE Xplore. Restrictions apply.

MATh

A non-seasonal ARIMA forecasting process represented by using an equation such as :

$$\alpha(B)(1-B^d)y_t = c + \theta(B)\epsilon_t \tag{7}$$

where ϵ_t is a noise with $\sigma^2 = 1$ and $\mu = 0$, B is the backshift operator,c is a function represents random walk with drift, and α and θ are polynomials function with order p and q, respectively [3].

In order to deal with the seasonal and nonseasonal behavior of the time series data, some modification has been made to the ARIMA models by adding a seasonal autoregressive notation (P), a seasonal moving average (Q), an order of seasonal differencing represented by (D), and the period of seasonality represented by (m). This modification leads to the SARIMA method, denoted by SARIMA(p,d,q)(P, D, Q)m or ARIMA(p,d,q)(P, D, Q)m [18]. The SARIMA forecasting method can be expressed by using the equation :

$$\phi(B)\Phi(B^s)(1-B^s)^D(1-B)^d z_t = \theta(B)\Theta(B^s)a_t, \tag{8}$$

where Φ and Θ are polynomial functions with orders P and Q, respectively, each having no root in a unit circle. P and Q represent the autoregressive and moving average for the ARIMA model, D refers to the order of differencing (D) whereas s represents the period of seasonality (m).

D. Performance Analysis

The performance analysis of the proposed methods conducted by measuring the error from each method using root mean square error (RMSE). The R^2 applied to measure the proportion of the variance for the dependent variable. The formula for the RMSE and R^2 value can be written equation (9) and (10) as follow

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - y)^2}{N}}, \tag{9}$$

$$R^2 = \frac{\sum_{i=1}^{N}(x_i - x)(y_i - y)}{\sqrt{\sum_{i=1}^{N}(x_i - x)^2}\sqrt{\sum_{i=1}^{N}(y_i - y)^2}}. \tag{10}$$

IV. Results and Discussion

In this section, a comparison analysis of prediction results performed from Holt-Winters Exponential Smoothing, ARIMA, and SARIMA by applying the methods provided by the statsmodels-package in python.

A. Result

A brute force method to search the best level smoothing, trend smoothing and seasonal smoothing parameter for the Holt-Winters Exponential Smoothing applied to produce the best result. The preprocessed data used as the input data with the seasonal period 696. The model used to forecast the tide level of Cilacap for 1st June- 7th June 2019, such as shown in Table 3. The comparison of the real data and the forecasting result is shown in Figure 6..

TABLE I
Holt-Winters Exponential Smoothing Forecasting Result

Winters Holt Exponential Smoothing Result		
date	***winters holt result***	***real value***
2019-06-01 00:00:00	0.822462	0.802250
2019-06-01 01:00:00	0.684480	0.717033
2019-06-01 02:00:00	0.475947	0.537683
2019-06-01 03:00:00	0.251616	0.313233
2019-06-01 04:00:00	0.044806	0.081850
...	...	...
2019-06-07 21:00:00	-0.013154	-0.016783
2019-06-07 22:00:00	0.114886	0.015667
2019-06-07 23:00:00	0.319314	0.139217
2019-06-08 00:00:00	0.566751	0.401533
2019-06-08 01:00:00	0.795249	0.546188

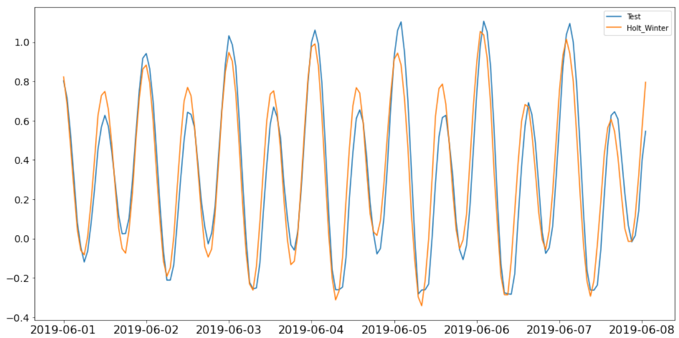

Fig. 6. The comparison plot between holt-winters forecasting and the actual data.

In order to determine the best (p,d,q) parameter suited for the ARIMA and SARIMA, the aic applied to produce the best forecasting result using the preprocessed data from 2019- 2019 as the input data. The sample of the forecasted data using ARIMA(12,0,4) and SARIMA(1,0,1)(10,2,7,7) compared with the original sea level data shown in Table 4. A comparison of the plotted original data and the forecasted results are shown in Figure 7.

TABLE II
Winters Holt Exponential Smoothing Forecasting Result Sample

ARIMA AND SARIMA Result			
date	***ARIMA result***	***SARIMA result***	***Actual value***
2019-06-01 00:00:00	0.865226	0.859817	0.802250
2019-06-01 01:00:00	0.742950	0.740300	0.717033
2019-06-01 02:00:00	0.567488	0.565117	0.537683
2019-06-01 03:00:00	0.362015	0.358909	0.313233
2019-06-01 04:00:00	0.171958	0.180101	0.081850
...	...	...	
2019-06-07 21:00:00	0.198746	0.011175	-0.016783
2019-06-07 22:00:00	0.193830	0.031246	0.015667
2019-06-07 23:00:00	0.255898	0.138356	0.139217
2019-06-08 00:00:00	0.372024	0.323076	0.401533
2019-06-08 01:00:00	0.514869	0.508719	0.546188

The comparison between the Holt-Winters Exponential Smoothing, ARIMA, and SARIMA is shown in Table 5. The table shows that the RMSE value from Holt-Winters is lower than the other method. The same as the RMSE, the R-square value for Holt-Winters, also show better value than the other

Authorized licensed use limited to: University of Exeter. Downloaded on November 14,2020 at 11:01:13 UTC from IEEE Xplore. Restrictions apply.

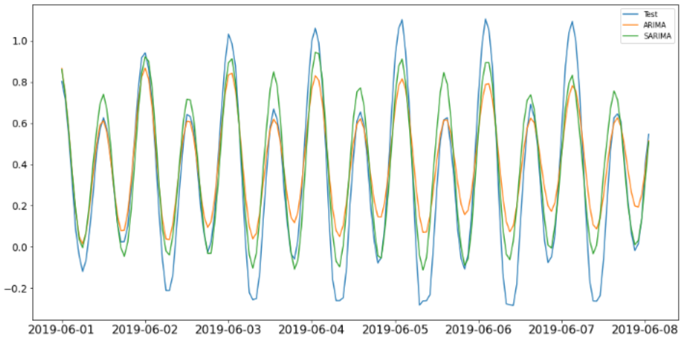

Fig. 7. The comparison plot between ARIMA, SARIMA forecasting and the actual data.

method that used in this study.

TABLE III
METHOD PERFORMENCE ANALYSIS

Method	***RMSE***	R^2
ARIMA	0.22593661	0.66571611
SARIMA	0.15454456	0.84295690
Winters holt	0.13384729	0.88220397

The effects of selecting the seasonal period parameter values to the forecasting results generated by Holt-Winters are investigated. The compared result of the method with the various seasonal period is shown in Tabel 6.

TABLE IV
SEASONAL PERIOD ANALYSIS

method	***winters holt RMSE***
seasonal period = 360 hour	0.18880094
seasonal period = 696 hour	0.13384729
seasonal period = 1031 hour	0.23168355
seasonal period = 1391 hour	0.28919736
seasonal period = 1702 hour	0.25156374
seasonal period= 2124 hour	0.21335184

V. CONCLUSION

A method for forecasting the tide level using Holt-Winters Exponential Smoothing is proposed in this study. The result produced by this method is compared with the result of ARIMA with (p, d, q) = (12,0,4), SARIMA (1,0,1)(10,2,7,7). The data used for this study is the recorded sea level data in Cilacap from January-June 2019. The Holt-Winters method produces a better result in forecasting the Cilacap tide level than the ARIMA and SARIMA. The Holt-Winter method produces R-square values of 0.88220397 and 0.13384729 for the RMSE values, and from this study, it shows that the Holt-Winters method greatly influenced by the seasonal period value that we set to get the best forecasting performance.

REFERENCES

[1] Change, IPCC Climate. "The physical science basis." Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change 996, 2007.

[2] E. Selitubun, Y. A. Lesnussa, and F. Kondolembang. "Pemodelan ARIMA untuk Prediksi Kenaikan Muka Air Laut dan Dampaknya Terhadap Luas Sebaran Rob di Kota Ambon." JMPM: Jurnal Matematika dan Pendidikan Matematika 2.1, 2017, pp.58-67.

[3] K. P. Srivastava, et al. "Forecasting Arabian Sea level rise using exponential smoothing state space models and ARIMA from TOPEX and Jason satellite radar altimeter data." " Meteorological Applications 23, no. 4, 2016,pp. 633-639.

[4] S. Karimi, et al. "Neuro-fuzzy and neural network techniques for forecasting sea level in Darwin Harbor Australia." ,Computers and Geosciences 52, 2013, pp. 50-59.

[5] M. A. Ghorbani, R. Khatibi, A. Aytek, O. Makarynskyy, and J. Shiri. "Sea water level forecasting using genetic programming and comparing the performance with artificial neural networks." Computers and Geosciences 36, no. 5, 2010,pp. 620-627.

[6] F. R. Fernandez, et al. "Prediction of South China sea level using seasonal ARIMA models." AIP Conference Proceedings. Vol. 1905. No. 1. AIP Publishing LLC, 2017.

[7] P. S. Kalekar. "Time series forecasting using holt-winters exponential smoothing." Kanwal Rekhi School of Information Technology 4329008.13, 2004.

[8] B. Triatmodjo. Teknik Pantai. Beta Offset: Yogyakarta, 1999.

[9] Q. Sun, J. Wan, and S. Liu. "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, 2020.

[10] F. R. Fernandez, et al. "Forecasting Manila South Harbor Mean Sea Level Using Seasonal ARIMA Models." Journal of Technology Management and Business 5.1, 2018.

[11] M. Imani, R. You, and C.Y. Kuo. "Accurate Forecasting of the satellite-derived seasonal Caspian sea level anomaly using polynomial interpolation and holt-winters exponential smoothing." TAO: Terrestrial, Atmospheric and Oceanic Sciences 24.4, 2013, p. 521.

[12] M. Imani, R. You, and C. Y. Kuo. "Accurate Forecasting of the satellite-derived seasonal Caspian sea level anomaly using polynomial interpolation and holt-winters exponential smoothing." TAO: Terrestrial, Atmospheric and Oceanic Sciences 24.4, 2013, p. 521.

[13] H. Grubb and A. Mason. "Long lead-time forecasting of UK air passengers by Holt–Winters methods with damped trend." International Journal of Forecasting 17.1, 2001, pp. 71-82.

[14] S. Gelper, R. Fried, and C. Croux. "Robust forecasting with exponential and Holt–Winters smoothing." Journal of forecasting 29.3, 2010, pp. 285-300.

[15] S. Lima, A. M. Gonçalves, and M. Costa. "Time series forecasting using Holt-Winters exponential smoothing: An application to economic data." AIP Conference Proceedings. Vol. 2186. No. 1. AIP Publishing LLC, 2019.

[16] K. Suppalakpanya, et al. "An Evaluation of Holt-Winters Methods with Different Initial Trend Values for Forecasting Crude Palm Oil Production and Prices in Thailand." Suranaree Journal of Science and Technology 26.1, 2019.

[17] V. S. Ediger and S. Akar. "ARIMA forecasting of primary energy demand by fuel in Turkey." Energy policy 35.3, 2007, pp. 1701-1708.

[18] J. Y. Kajuru, K. Abdulkarim, and M. M. Muhammed. "Forecasting Performance of Arima and Sarima Models on Monthly Average Temperature of Zaria, Nigeria." ATBU Journal of Science, Technology and Education 7.3, 2019, pp. 205-212.

Authorized licensed use limited to: University of Exeter. Downloaded on November 14,2020 at 11:01:13 UTC from IEEE Xplore. Restrictions apply.