FORECASTING MANILA SOUTH HARBOR MEAN SEA LEVEL USING SEASONAL ARIMA MODELS

Flerida Regine Q. Fernandez[1], Neil B. Montero[2], Rodolfo B. Po III[2],
Rizavel C. Addawe[2], Hazel Mae R. Diza[1]

[1] Mathematics Department, Pangasinan State University, Lingayen, Philippines
[2] Department of Mathematics and Computer Science, University of the Philippines,
Baguio City, Philippines

*Corresponding E-mail : fqfernandez1@up.edu.ph

Abstract

Global warming has adverse effects which include the rise of the mean sea level. This could be a problem especially those countries that are surrounded by bodies of water. The Manila South Harbor is a part of the South China Sea being one of the bodies of water that surround the Philippines. This paper aims to find Seasonal Autoregressive Integrated Moving Average (SARIMA) models that fits the given time series composed of the mean sea level of the Manila South Harbor from 2008 to 2014 measured in millimeters. Results show that there are three possible models that fit the time series but the chosen one is the model $SARIMA(1,0,0)(0,1,1)_{12}$. The forecasted values were then compared to the actual values of the mean sea level for the year 2015.

1.0   Introduction

The human population, even the planet itself, have already suffered a lot because of the unwanted effects of global warming or the so-called heating of the earth. Several adversities have already been observed such as worldwide melting of ice, death of some animal species, loss of habitat, increased flooding and submergence of low-lying areas, and many more. Although some countries have already taken steps to stop and fight global warming, it still one of the major problems today.

2.0   Literature Review

According the article entitles Effects of Global Warming (Bradford, Pappas,2017), one of the obvious effects of global warming is the increase of the average global temperature. It has increased by about 1.4 degrees Fahrenheit for the last 100 years according to the records of the National Oceanic and Atmospheric Administration (NOAA). And since temperature dictates the melt or formation of ice, the change on the global temperature has greatly affected the snow covers on North America, Europe, and Asia. The said continents have noticed less snow cover on their territories from 1960 to 2015 as written on the study of Kunkel, et al. (2016). Also, there is a lag on the peak of the formation of ice in the Arctic sea last 2016 and 2016 (NASA, Goddard Space Flight Center, 2017).

Journal of Technology Management and Business (ISSN: 2289-7224)
Vol 5, No 1, Special Edition, 2018
Managing Epidemic and Natural Disaster using Mathematical Modelling

As expected, due to the melting of ice on the different parts of the world, sea levels started to rise. This rise affected both the relative sea level and absolute sea level. Relative sea level pertains to the sea level relative to the nearest ocean floor. On the other hand, absolute sea level pertains to the sea level relative to the center of the earth. These two types of sea level can be measured using tide gauges or satellite radar altimeters. Actually, it has been reported that the rise of sea levels accelerated by an averaged 0.12 inches per year worldwide (World Meteorological Organization, 2014).

South China Sea is one of the water bodies that surround the Philippines. Several provinces are near the said sea. Major changes on its sea level will obviously affect those who live near that area. Several studies have been made regarding the South China sea level. For instance, as compared to the global sea level rise of the South China sea from 2003 to 2009, the variations in sea level from the measurements of satellite altimeters are definitely higher (Feng, et al., 2012). Hence, it is very important to be able to forecast the possible sea levels of the South China sea for economic and livelihood preparation. And so, in 2017, Fernandez, Montero, Po, and Addawe fitted a Seasonal Autoregressive Integrated Moving Average (SARIMA) model given by SARIMA $(1,0,0,)(0,1,1)_{12}$ to predict the next values of the South China sea level.

Manila South Harbor is a part of the South China Sea. It is considered to be the Philippine's currently most important gateway to foreign commerce and one of the important maritime hub in Asia-Pacific Region (Migu, 2009). Hence, this paper seeks to find SARIMA models for the Manila South Harbor sea level measurements obtained from the Permanent Service Mean Sea Level (PSMSL) of the Philippines obtained by using tide gauges.

3.0 Methodology

There are several ways on how to measure the sea level. One of which is the use of tide gauges. Technically it measures the height of the sea with respect to the nearest ocean floor. This study used the mean sea level of the manila south harbor taken by the Permanent Service for Mean Sea Level (PSMSL). They have used tide gauges in order to take the sea level measurements in millimeters. The data from the year 2008-2014 were taken to compose the time series used in this study. The remaining data for the year 2015 were then compared to the forecasted values.

The R statistical packages were used to generate appropriate models for the said time series. It was also used to compute errors and generate the necessary plots for the time series and forecasted values.

UBJ-ARIMA models are used to forecast future values of a given time series. The said statistical model would only work if the time series used is stationary. That is, the variance and average of the time series is constant over time. Otherwise, stationarity should be induced by means of several methods such as logarithmic transformation, differencing, etc. The model is usually of the form ARIMA(p, d, q) where the p denotes the AR order, the d denotes the number of times we applied differencing on the data set, while the q denotes the MA order. On the other hand, for seasonal time series, it is of the form SARIMA $(p, d, q)(P, D, Q)_s$ where P, D, Q, s are the seasonal AR order, seasonal differencing induced, seasonal MA order, and seasonality order, respectively. (Pankratz, 1983).

The Augmented Dickey Fuller (ADF) test was used to check if the given series is stationary or not. A p-value f less than 0.05 shows that the series is stationary.

Journal of Technology Management and Business (ISSN: 2289-7224)
Vol 5, No 1, Special Edition, 2018
Managing Epidemic and Natural Disaster using Mathematical Modelling

To identify possible models, the autocorrelation (ACF) and autocorrelation (PACF) plots were observed. ACF shows the correlation of ordered pairs which are separated by time spans. On the other hand, PACF shows the correlation of ordered pairs while taking into account the observations between them. Since we have 120 observations, only the first 30 observations are significant and should be examined. (Pankratz, 1983).

For the diagnostic checking, the error measures Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were used to test the accuracy of the forecasts of the models. Also, to measure the relative quality of the models, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used. The AIC shows the amount of information lost when a model is used to forecast future values while BIC shows the effectiveness of the model when used to forecast future values. For all the said criteria, we take the model with the lowest values.

Lastly, Ljung-Box test was used to see if the chosen model is fit for forecasting. A model with p-value of greater than 0.05 shows that the model is adequate for forecasting. The forecasted values were then compared to the actual values of the year 2015.

4.0 Results and Discussions

Ensuring the stationarity of the time series is a priority before using the UBJ-ARIMA modelling.
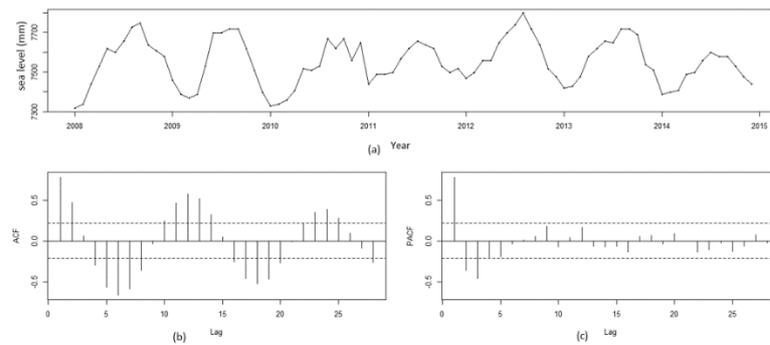
Figure 1: (a) Data plot, (b) ACF plot, and (c) PACF plot of the monthly values of the sea level of the Manila South Harbor.

By observation in Fig.1 (a), the mean which is 7550 seems to be constant over time since the graph fluctuates around it. Also, the distances between the data plots does not seem to vary which shows that the variance does not change over time. From those observations, we may conclude that the time series is stationary. However, to be sure, we used the Augmented Dickey Fuller (ADF) test. The p-value given by the said test is equal to 0.01 is less than 0.05, hence, the time series is stationary.

Now, we see that as shown in Fig.1(b), there is a seasonality of the time series. Also, the seasonal lags seem to slowly approach zero which means that there is a need for seasonal differencing. Another observation is that there is a spike on lag1 as shown on Fig.1(c). Considering all those observations, we have the model $SARIMA(1,0,0)(0,1,0)_{12}$. Fig.2 shows its corresponding ACF and PACF plots.

Journal of Technology Management and Business (ISSN: 2289-7224)
Vol 5, No 1, Special Edition, 2018
Managing Epidemic and Natural Disaster using Mathematical Modelling
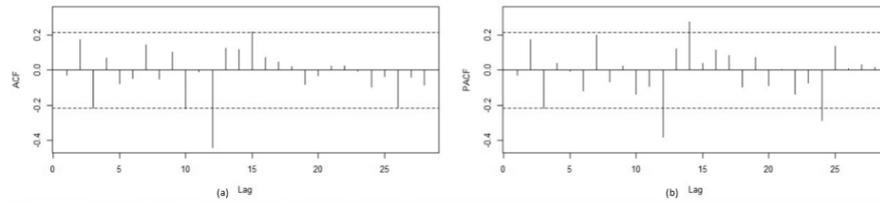
Figure 2: (a) ACF plot, and (b) PACF plot of SARIMA$(1,0,0)(0,1,0)_{12}$

As observed in Fig.2, there is a spike on the twelfth lag of the ACF and PACF of the model SARIMA$(1,0,0)(0,1,0)_{12}$. Note that it is positioned on a seasonal lag. This means that adding either a seasonal AR or MA coefficient is necessary. However, adding only a seasonal MA coefficient was proved to be necessary to remove the spikes. And so SARIMA$(1,0,0)(0,1,1)_{12}$ was taken together with the following plots for ACF and PACF.
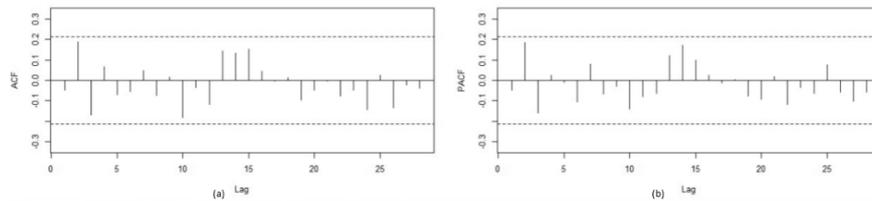
Figure 3: (a) ACF plot, and (b) PACF plot of SARIMA$(1,0,0)(0,1,1)_{12}$

Clearly, Fig.3 shows that all the lags, fall behind the 95% confidence interval shown by the dotted lines. This means that, SARIMA$(1,0,0)(0,1,1)_{12}$ may already be a model that can be used to represent the time series and can be used to forecast future values of the Manila South Harbor. However, the second lag of the ACF plot seems to be too close to the dotted lines. Adding a nonseasonal AR coefficient gives the following plots.
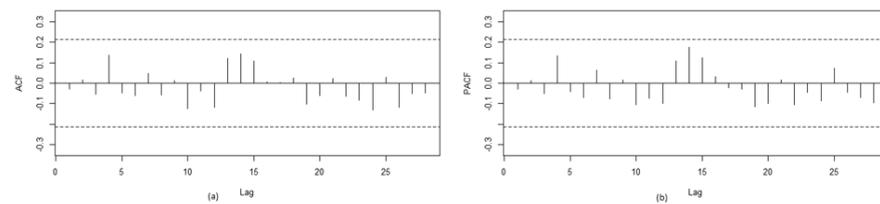
Figure 4: (a) ACF plot, and (b) PACF plot of SARIMA$(1,0,2)(0,1,1)_{12}$

The lags became shorter and farther form the 95% confidence limits which is a good sign. Although, this model already has 4 coefficients: 1 nonseasonal AR coefficient, 2 nonseasonal MA coefficient, and 1 seasonal MA coefficient. Fixing the first nonseasonal MA coefficient to zero reduces the number of coefficients to three and gives the following plots.

Journal of Technology Management and Business (ISSN: 2289-7224)
Vol 5, No 1, Special Edition, 2018
Managing Epidemic and Natural Disaster using Mathematical Modelling
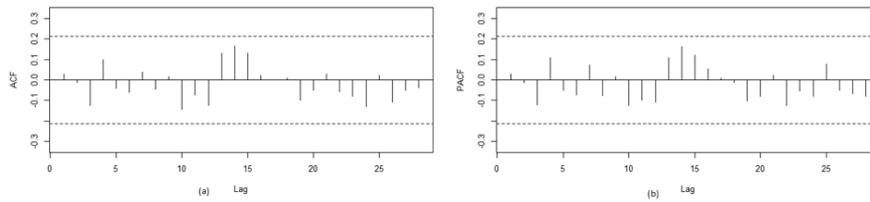
Figure 5: (a) ACF plot, and (b) PACF plot of $SARIMA(1,0,2)(0,1,1)_{12}$ where the first nonseasonal MA coefficient is fixed to zero.

Again, as shown in Fig.5. all the lags fall behind the dotted lines which means that this may be one of the possible models that can be used to forecast future values.

To summarize, there are three possible models for the given time series. Ljung-Box test was used to check if the models are adequate for forecasting. $SARIMA(1,0,0)(0,1,1)_{12}$ has p-value of 0.662, $SARIMA(1,0,2)(0,1,1)_{12}$ has p-value of 0.797, and $SARIMA(1,0,2)(0,1,1)_{12}$ where the first MA coefficient is fixed to zero has p-value of 0.7796. All of the models has given a p-value of greater than 0.05, hence, all are fit for forecasting.

Table 1: AIC, BIC, RMSE, MAPE, Average % error, and number of coefficients of the possible models.

Model	AIC	BIC	RMSE	MAPE	Average % error	No. of Coefficients
$(1,0,0)(0,1,1)_{12}$	783.23	790.06	0.4106919	41.95291	1.2125	2
$(1,0,2)(0,1,1)_{12}$	782.83	794.21	0.4104662	40.63701	1.255833	4
$(1,0,2)(0,1,1)_{12}$ with fixed MA coefficient	781.62	790.73	0.4141278	40.86847	1.224167	3

Table 1 shows the summary of the possible models. The values for the AIC, BIC, RMSE, and MAPE of the models are close to each other so we take note the average percent error and the idea of the parsimony. The average percent error was computed by averaging the percent error of the forecasted values as compared to the actual values. On the other hand, the idea of parsimony pertains to the thrift on the number of coefficients of the models. In that case, the model $SARIMA(1,0,0)(0,1,1)_{12}$ has the least average percent error and least number of coefficients. This implies that $SARIMA(1,0,0)(0,1,1)_{12}$ is the best fit model for the time series.

Journal of Technology Management and Business (ISSN: 2289-7224)
Vol 5, No 1, Special Edition, 2018
Managing Epidemic and Natural Disaster using Mathematical Modelling

(a)

(b)

(c)
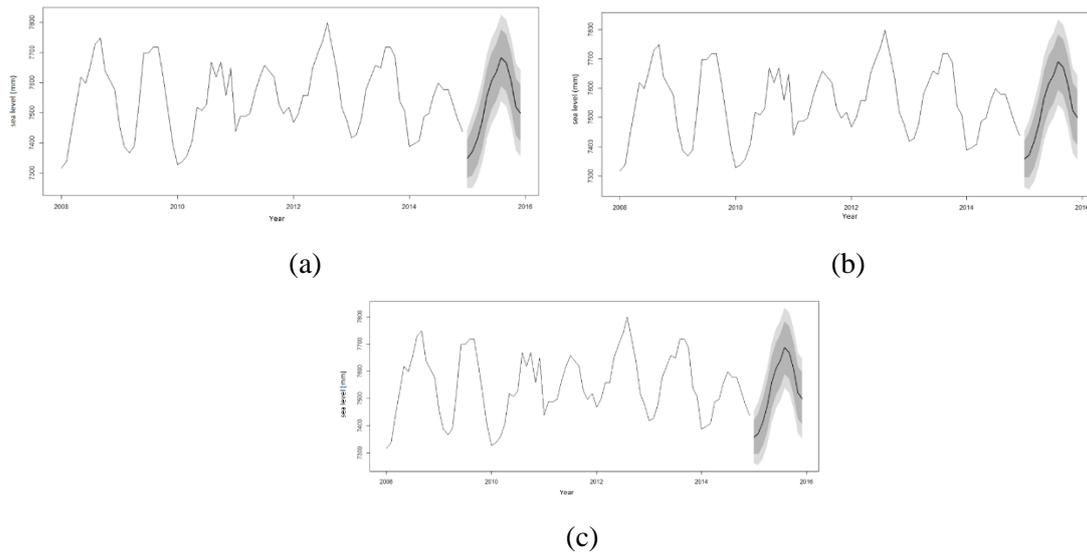
Figure 6: A 1 year forecast using (a) $SARIMA(1,0,0)(0,1,1)_{12}$, (b)$SARIMA(1,0,2)(0,1,1)_{12}$, and (c) $SARIMA(1,0,2)(0,1,1)_{12}$ where the first nonseasonal MA coefficient is fixed to zero.

As shown in Fig.6, the forecasted values of the three models are quite close to each other. Hence, any of the three models may be used to forecast the future values for the Manila South Harbor.

5.0   Conclusions

Fig.6 (a), (b), and (c) show the possible values for the Manila South Harbor for the year 2015. Actual sea level values are already available for the year 2015 and majority of the actual values are within the 80% confidence level shown by the dark grey area of the three figures. This implies that any of the three chosen models may be used to forecast possible values for the sea level. However, because of the idea of parsimony and average percentage error, the model $SARIMA(1,0,2)(0,1,1)_{12}$ was chosen to be the best-fit model for the time series. Only the actual values for the months of January and February did not fall on the 80% confidence interval. The actual values are higher than the grey areas. Which means to say that the mean sea level truly is rising quite unpredictably.

This model may be improved as soon as the data set for the sea level of Manila South Harbor is updated. For future works, using nonlinear modeling along with UBJ-ARIMA modeling may be used for increased forecasting accuracy. The method used in this paper may also be applied to forecast sea levels of other water bodies.

Journal of Technology Management and Business (ISSN: 2289-7224)
Vol 5, No 1, Special Edition, 2018
Managing Epidemic and Natural Disaster using Mathematical Modelling

References

Permanent Service for Mean Sea Level. (2017). *Tide Gauge Data*. Retrieved from http://www.psmsl.org/data/obtaining/rlr.monthly.data/145.rlrdata

National Geographic. *Effects of Global Warming*. Retrieved from https://www.nationalgeographic.com/environment/global-warming/global-warming-effects/

Bradford, A., & Pappas, S. (2017, August 12). *Effects of Global Warming*. Retrieved from https://www.livescience.com/37057-global-warming-effects.html

Feng, W., Zhong, M., & Xu, H. (2012) Sci. Cn. Earth Sci. 55, p.16961701

Pankratz, A. (1983). *Forecasting with Univariate Box-Jenkins Models Concepts and Cases.* John Wiley and Sons, New York

R. C. Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Viena, Austria (2013), ISBN3-90051-070

National Oceanic and Atmospheric Administration (2017, October 5). *Is Sea Level Rising?* Retrieved from http://oceanservice.noaa.gov/facts/sealevel.html.

Kunkel, K., Robinson, D., Champion, S., Yin, X., Estilow, T., & Frankson, R. (2016, April 8). *Trends and Extremes in Northern Hemisphere Snow Characteristics.* Springer International Publishing

NASA/Goddard Space Flight Center. (2017, July 24). *Summer ice melt in the Arctic*. Retrieved from https://www.sciencedaily.com/releases/2017/07/170724133153.htm

Fernandez, FR., Po III, R., Montero, N., & Addawe, R. (2017, December). *Prediction of South China Sea Level Using Seasonal ARIMA Models.* Proceedings of the 13th IMT-GT International Conference on Mathematics, Statistics, and their Applications (IMSA 2017). AIP Conf. Proc 1905, 050018-1-050018-6; https://doi.org/10.1063/1.5012237. AIP Publishing.

Migu. (2009, June 27). *Featured Philippine Ports: The Manila South Harbor*. Retrieved from https://mcgutib.wordpress.com/2009/06/27/featured-ports-pmo-south-harbor/