Chi, Yeong Nain. (2020), Analyzing Long-Term Records of Global Average Sea Level Change Using ARIMA Model. In: Journal of Economics and Business, Vol.3, No.2, 672-681.

ISSN 2615-3726

DOI: 10.31014/aior.1992.03.02.230

The online version of this article can be found at: https://www.asianinstituteofresearch.org/
Analyzing Long-Term Records of Global Average Sea Level Change Using ARIMA Model

Yeong Nain Chi

1 Department of Agriculture, Food and Resource Sciences, University of Maryland Eastern, Maryland, U.S.A.

Correspondence: Yeong Nain Chi, Department of Agriculture, Food and Resource Sciences, University of Maryland Eastern, Princess Anne, Maryland 21853, U.S.A. E-mail: ychi@umes.edu

Abstract
The purpose of this study was to demonstrate the role of time series model in predicting process and to pursue analysis of time series data using long-term records of global average sea level change from 1880 to 2013 extracted from the U.S. Environmental Protection Agency using data from Commonwealth Scientific and Industrial Research Organization, 2015. Following the Box–Jenkins method, ARIMA(0,1,1,) model was the best fitted model in prediction for the data, Global Average Absolute Sea Level Change, 1880-2013, in this study. Forecasting process with ARIMA(0,1,1) model for prediction indicated global average sea level change at a constant increasing rate in the short-term. Understanding past sea level is important for the analysis of current and future sea level changes. In order to sustain these observations, research programs utilizing the resulting data should be able to significantly improve our understanding and narrow projections of future sea level rise and variability.

Keywords: Sea Level Rise, Long-Term Records, Time Series Analysis, ARIMA

1. Introduction

Climate change is the long term change that occurs in the average weather patterns of the earth. It is primarily triggered by human activities like burning of fossil fuels, deforestation, etc. or natural events like volcanic eruptions. National Oceanic and Atmospheric Administration (NOAA) 2019 Global Climate Annual Report summarized that the global annual temperature has increased at an average rate of 0.07°C (0.13°F) per decade since 1880. To make it worse, this rate (+0.18°C/ +0.32°F) of increase has doubled since 1981 (https://www.ncdc.noaa.gov/sotc/global/201913).

Sea level rise is caused primarily by two factors related to climate change: the added water from melting of land ice (ice sheets and glaciers) to the world’s oceans and the thermal expansion of seawater temperature rise. The potential impacts of sea level rise include, but not limited to, increasing coastal flooding and erosion, damages on
agricultural land cover and crops, damages on coastal/urban settlements and infrastructures, damages on coastal flora and fauna ecosystems, increasing environmental sanitation problem, and increasing public health problem. According to Climate.gov, global average sea level has risen about 8–9 inches (21–24 cm) since 1880 (https://www.climate.gov/news-features/understanding-climate/climate-change/global-sea-level). Church et al. (2004) used TOPEX/Poseidon satellite altimeter data and combined with historical tide gauge data to estimate monthly distributions of large-scale sea level variability and change over the period 1950–2000. The computed rate of global average sea level rise from the reconstructed monthly time series was $1.8 \pm 0.3$ mm year$^{-1}$.

After that, Church and White (2006) indicated that a larger rate of sea level rise after 1993 and other periods of rapid sea level rise but no significant acceleration over this period using tide-gauge data from 1950 to 2000. They extended the reconstruction of sea level back to 1870 and found a sea level rise from January 1870 to December 2004 of 195 mm, a 20th century rate of sea level rise of $1.7 \pm 0.3$ mm yr$^{-1}$ and a significant acceleration of sea level rise of $0.013 \pm 0.006$ mm yr$^{-2}$.

In 2011, Church and White (2011) found that the estimated rate of sea level rise was $3.2 \pm 0.4$ mm year$^{-1}$ from the satellite data and $2.8 \pm 0.8$ mm year$^{-1}$ from the in situ data. The global average sea level rise from 1880 to 2009 was about 210 mm. The linear trend from 1900 to 2009 was $1.7 \pm 0.2$ mm year$^{-1}$ and since 1961 was $1.9 \pm 0.4$ mm year$^{-1}$. They also documented that there was considerable variability in the rate of sea level rise during the twentieth century but there has been a statistically significant acceleration since 1880 and 1900 of $0.009 \pm 0.003$ mm year$^{-2}$ and $0.009 \pm 0.004$ mm year$^{-2}$, respectively.

The Intergovernmental Panel on Climate Change (IPCC) (2014) estimated that the sea level has risen by 26–55 cm (10–22 inches) with a 67% confidence interval. If emissions remain very high, the IPCC projected sea level could rise by 52–98 cm (20–39 inches). In its Fourth National Climate Assessment Report (2017) the U.S. Global Change Research Program (USGCRP) estimated that sea level has risen by about 7–8 inches (about 16–21 cm) since 1900, with about 3 of those inches (about 7 cm) occurring since 1993 (very high confidence). Relative to the year 2000, sea level was very likely to rise by 1.0–4.3 feet (30–130 cm) in 2100, and 0.3–0.6 feet (9–18 cm) by 2050.

There had many studies pointed out that sea level is rising at an increasing rate (Church et al., 2008; Cazenave and Llovel, 2010; Cazenave and Cozannet, 2013; Horton et al., 2018; Kulp and Strauss, 2019; Haasnoot, 2020). Thus, understanding past sea level is important for the analysis of current and future changes. Modeling sea level change and understanding its causes has considerably improved in the recent years, essentially because new in situ and remote sensing observations have become available (Foster and Brown, 2014; Visser et al., 2015; Bolin, 2015; Srivastava et al., 2016). Despite the importance of sea level rise and its consequences, there is a lack of studies in the technical literature available on forecasting schemes.

The purpose of this study is to demonstrate the role of time series model in predicting process and to pursue analysis of time series data using long-term records of global average sea level change from 1880 to 2013. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data, while time series forecasting is the use of a model to predict future values based on previously observed values. Time series forecasting is one of the most applied data science techniques in business, used extensively in finance, supply chain management, production and inventory planning. It has a well-established theoretical grounding in statistics and dynamic systems theory. It is important because there are so many prediction problems that involve a time component. These problems are neglected because it is this time component that makes time series problems more difficult to handle.

2. Materials and Methods

The data, Global Average Absolute Sea Level Change, 1880-2013 (Figure 1), used for this study is available to the general public from the US Environmental Protection Agency (http://www3.epa.gov/climatechange/images/indicator_downloads/sea-level_fig-1.csv) using data from CSIRO
The data contains “cumulative changes in sea level for the world’s oceans since 1880, based on a combination of long-term tide gauge measurements and recent satellite measurements. It shows average absolute sea level change, which refers to the height of the ocean surface, regardless of whether nearby land is rising or falling. Satellite data are based solely on measured sea level, while the long-term tide gauge data include a small correction factor because the size and shape of the oceans are changing slowly over time. (On average, the ocean floor has been gradually sinking since the last Ice Age peak, 20,000 years ago.)” (Quoted from https://datahub.io/core/sea-level-rise#readme).

Figure 1. Time Series Plot of Global Average Absolute Sea Level Change, 1880-2013

A common approach in the analysis of time series data is to consider the observed time series as part of a realization of a stochastic process. In time series analysis and its various applications, a common assumption is that the data are stationary. Intuitively, stationarity means that the statistical properties (i.e., mean and variance) of the process do not change over time. In mathematics and statistics, stationarity is a property of a stochastic process.

ARIMA models are regression models that use lagged values of the dependent variable and/or random disturbance term as explanatory variables. In ARIMA model, the Autoregressive Process (AR) part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values; the Moving Average Process (MA) part indicates that the regression error is actually a linear combination of error terms whose values occurred at various times in the past; and the I (for “integrated”) indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model better fit to predict future points in a time series data as well as possible (Montgomery et al., 2008).

The general form of ARIMA model of order (p,d,q) is

\[ Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \]  

where \( p \) = order of time lags, \( d \) = order of differencing, \( q \) = order of moving average, \( Y_t \) is value of the series at time \( t \), \( Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p} \) are dependent on the past values of the variable at specific time points, \( \phi_0, \phi_1, \phi_2, \ldots, \phi_p \) are the regression coefficients, \( \theta_1, \theta_2, \ldots, \theta_q \) are the (weights) coefficients applied to \( \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q} \) previous forecast errors and \( \varepsilon_t \) is error part (not explained by model).

In time series analysis, the Box–Jenkins method, named after the statisticians George Box and Gwilym Jenkins (Box and Jenkins, 1970), applies ARIMA models to find the best fit of a time series model to past values of a time series. The Box–Jenkins method refers to a systematic method of identifying, fitting, checking, and using ARIMA
models (Box et al., 2016), which is the process for estimating ARIMA models based on the autocorrelation function (ACF) and partial autocorrelation function (PACF) as a means of determining the stationarity of the variable in question and the lag lengths of the ARIMA model. Thus, the Box–Jenkins method starts with the assumption that the process that generated the time series can be approximated using an ARMA model if it is stationary or an ARIMA model if it is non-stationary.

For ARIMA model, there are four important steps, including identification, estimation, diagnostic checking and forecasting, should be put in consideration to apply it. ARIMA models rely heavily on the autocorrelation pattern in the time series data. Identification step applied to achieve stationarity and to build a suitable model using ACFs, PACFs, and transformations (differencing). If the time series is not stationary, it needs to be stationarized through differencing. Stationarizing a time series through differencing is an important part of the process of fitting an ARIMA model. In practice, it is almost not necessary to go beyond second difference, because real data generally involve only first or second level non-stationarity.

In the estimation step, plots and summary statistics can be used to identify trends and autoregression elements to get an idea of the amount of differencing and the size of the lag that will be required for model identification. The following estimation step is to use a fitting procedure to find the coefficients of the model. In order to discover good parameters for the model, Akaike’s Information Criterion (AIC) or Bayesian Information Criterion (BIC) can be used to determine the order p and q of an ARIMA model. Good models are obtained by minimizing the AIC or BIC.

Diagnostic checking step is primarily to use plots and statistical tests of the residual errors to determine the model fitting, and to evaluate the fitted model in the context of the available data and check for areas where the model may be improved. The process is repeated until a desirable level of fit is achieved, and non-significant parameters can be removed from the model. There are many accuracy metrics applied after model identification and estimation helping in choosing the best fitted model. The commonly used accuracy metrics to judge forecasts are: Mean Square Error (MSE) = \( \frac{1}{n} \sum (Y_t - \hat{Y}_t)^2 \), Root Mean Square Error (RMSE) = \( \sqrt{\text{MSE}} \), Mean Absolute Percentage Error (MAPE) = \( \frac{1}{n} \sum \left| \frac{Y_t - \hat{Y}_t}{|Y_t|} \right| \times 100 \), and Mean Absolute Error (MAE) = \( \frac{1}{n} \sum |Y_t - \hat{Y}_t| \) (Montgomery et al., 2008).

The last step is forecasting step that can be applied in prediction process after checking the model in the previous steps. The Forecasting Module of IBM SPSS Statistics 26 was used as the tool to estimate the model parameters to fit the ARIMA models in predicting global average sea level change to achieve the purpose of this study.

3. Empirical Results

3.1 Identification of ARIMA Model

By looking at the time plot of the data, ACF plot is useful for identifying non-stationary time series. For a stationary time series, the ACF will drop to zero relatively quickly, while the ACF of non-stationary data decreases slowly. The ACF plot of the data, Global Average Absolute Sea Level Change, 1880-2013, used for this study (Figure 2) showed strong positive statistically significant correlations at up to 16 lags that never decay to zero. This suggested that the time series is non-stationary and should be differenced. As a result, there is no need yet to examine the PACF for the time series.
In terms of non-stationary time series, differencing can be used to transform a non-stationary time series into a stationary one. The first difference of a time series is the time series of changes from one period to the next. Notice that the graph of the first difference of the data, Global Average Absolute Sea Level Change, 1880-2013, looked approximately stationary (Figure 3).

Figures 4 and 5 presented the ACF and PACF, respectively, of first difference of the data, Global Average Absolute Sea Level Change, 1880-2013. The ACF in Figure 4 and the corresponding PACF in Figure 5 showed a similar pattern that looked mostly like a steady decay toward zero after the first few lags. Together, this could suggest that the time series of first difference followed a first-order moving average process. Further analysis of the data, Global Average Absolute Sea Level Change, 1880-2013, should be conducted by estimating an ARIMA model.
There were many time series models suggested to represent the data with the normalized BIC values as in Table 1. The most fitted model was with the smallest normalized BIC value (-3.002). The most suitable model was Holt’s exponential smoothing method. Also, ARIMA model with ordering (0,1,1) is good in prediction process with BIC value (-2.994), followed by (1,1,0), (0,1,2) and (1,1,1) of BIC values (-2.961), (-2.957), and (-2.957), respectively. The other model statistics which indicated fitting of the model to the data were higher MAPE value with the lowest normalized BIC and lower RMSE value. All these measures were a good indicator of fitting of the model to the data well.

Table 1. Comparison of Different ARIMA Models

| Model                  | Normalized BIC | RMSE   | MAPE   | MAE   |
|------------------------|----------------|--------|--------|-------|
| Holt’s Exponential Smoothing | -3.002         | 0.215  | 11.887 | 0.168 |
| ARIMA(0,1,1)            | -2.994         | 0.216  | 12.014 | 0.169 |
| ARIMA(1,1,0)            | -2.961         | 0.219  | 12.042 | 0.171 |
| ARIMA(0,1,2)            | -2.957         | 0.216  | 11.708 | 0.168 |
| ARIMA(1,1,1)            | -2.957         | 0.216  | 11.724 | 0.169 |
3.2 Estimation of the Model Parameters

The parameters of Holt’s Exponential Smoothing Method were presented in Table 2. Even Holt’s exponential smoothing method was the most suitable model for the time series data, but the smoothing parameter for the trend ($\alpha_2$) had a very small value (= 0.00001951), which means that the slope hardly changes over time, and also was not significant statistically. Thus, in this case, Holt’s exponential smoothing method is identical to the simple Exponential Smoothing method.

Table 2. Parameters of Holt’s Exponential Smoothing Method

| Parameter | Estimate | SE  | t     | Sig. |
|-----------|----------|-----|-------|------|
| $\alpha_1$ (Level) | 0.601 | 0.083 | 7.232 | 0.000 |
| $\alpha_2$ (Trend) | 1.951E-5 | 0.032 | 0.001 | 1.000 |

Given this option, ARIMA(0,1,1) was chosen for further forecasting process, and the parameters of ARIMA(0,1,1) model were presented in Table 3. The ARIMA(0,1,1) model for the data, Global Average Absolute Sea Level Change, 1880-2013, can be expressed as follows:

$$\hat{Y}_t = 0.069 + Y_{t-1} + 0.395\varepsilon_{t-1} + \varepsilon_t$$

Table 3. Parameters of ARIMA(0,1,1) Model

| Parameter | Estimate | SE  | t     | Sig. |
|-----------|----------|-----|-------|------|
| Constant  | 0.069 | 0.011 | 6.029 | 0.000 |
| Difference | 1      |     |       |      |
| MA Lag 1  | 0.395  | 0.081 | 4.891 | 0.000 |

3.3 Diagnostic Checking of the Model

The Ljung–Box Q-test (Ljung and Box, 1978) is a diagnostic tool used to test the lack of fit of a time series model. The test statistic of the Ljung–Box Q-test was $Q = 9.216$ with 17 degrees of freedom and the p-value of the test was 0.933, which is much larger than 0.05. Thus, the null hypothesis that autocorrelations up to lag k equal zero (that is, the data values are random and independent up to a certain number of lags) was accepted. It concluded that the model was correctly specified.
Figure 6 illustrated the ACF and the PACF, respectively, for the residuals resulting from the estimated ARIMA (0,1,1) model. Reading from the bottom up, both figures showed no pattern in the correlations reported among the residuals nor did any of the correlations extend beyond the vertical 95% confidence intervals included in the plots. This, combined with the Ljung–Box Q-test statistic, suggested that the ARIMA (0,1,1) model appropriately modeled the dynamics for this time series.

![Figure 6. Residuals of ACF and PACF](image)

### 3.4 Using the Model in Forecasting Process

The actual and predicted values of Global Average Absolute Sea Level Change were shown in Table 4 and Figure 6. Forecasting process with the ARIMA(0,1,1) model for prediction, i.e., 10 years of future Global Average Absolute Sea Level Change indicated a good fitting of the model for prediction at a constant increasing rate in the short term.

| Year | Forecast | 95% Lower Control Limit | 95% Upper Control Limit |
|------|----------|-------------------------|-------------------------|
| 2014 | 9.159    | 8.732                   | 9.585                   |
| 2015 | 9.227    | 8.729                   | 9.726                   |
| 2016 | 9.296    | 8.734                   | 9.857                   |
| 2017 | 9.364    | 8.746                   | 9.982                   |
| 2018 | 9.433    | 8.763                   | 10.102                  |
| 2019 | 9.501    | 8.784                   | 10.219                  |
| 2020 | 9.570    | 8.807                   | 10.333                  |
| 2021 | 9.638    | 8.833                   | 10.444                  |
| 2022 | 9.707    | 8.861                   | 10.553                  |
| 2023 | 9.776    | 8.891                   | 10.660                  |
4. Conclusion

In this study the best choice time series model was ARIMA(0,1,1) model as its high R-squared and the lowest BIC values among other models. It was noticed that global average sea level would increase as this ARIMA(0,1,1) model gave evidence about future global average sea level rise. This model also provided information which are important in decision making process related to the future sea level rise impacts. Also, ARIMA model is a good time series model in forecasting the future performance not only for regional but also for national sea level rise outcomes.

There were many studies revealed that sea level is rising at an increasing rate. Thus, understanding past sea level is important for the analysis of current and future changes. Tides, for example, are predictable because they are produced by a combination of forces that are predictable. These forces are determined by the movements and positions of objects in our solar system, particularly the earth and the moon in relation to the sun. To understand tides, therefore, it is important to understand the movements of the earth and the moon relative to the sun. However, most of the time, and in most places, tides are much more complex.

Sea level rise is a relatively slow process, and the majority of impacts, with the exception of seasonally flooded low-lying coastal areas, are predicted and modeled for the future. In order to sustain these observations, research programs utilizing the resulting data should be able to significantly improve our understanding and narrow projections of future sea level rise and variability.

References

Bolin, D., Guttorp, P., Januzzi, A., Jones I, D., Novak, M., Podschwit, H., Richardson, L., S"arkk"a, A., Sowder, C., & Zimmerman, A. (2015). Statistical prediction of global sea level from global temperature. Statistica Sinica, 25, 351-367. doi: 10.5705/ss.2013.222w

Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: forecasting and control. Holden-Day, San Francisco.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). Time series analysis: forecasting and control (5th ed.). Hoboken, N.J.: John Wiley and Sons Inc.

Cazenave, A., & Llovel, W. (2010). Contemporary sea level rise. Annual Review of Marine Science, 2, 145-173. doi: 10.1146/annurev-marine-120308-081105.

Cazenave, A., & Cozannet, G. L. (2013). Sea level rise and its coastal impacts. Earth’s Future, 2, 15–34, doi:10.1002/2013EF000188.

Church, J. A., White, N. J., Coleman, R., Lambeck, K., & Mitrovica, J.X. (2004). Estimates of the regional distribution of sea level rise over the 1950–2000 period. Journal of Climate, 17(13), 2609-2625. http://hdl.handle.net/102.100.100/189020?index=1
Church, J. A., & White, N. J. (2006). A 20th century acceleration in global sea-level rise. *Geophysical Research Letters, 33*, L01602, 1-4. doi:10.1029/2005GL024826

Church, J. A., White, N. J., Aarup, T., Wilson, W. S., Woodworth, P. L., Domingues, C. M., Hunter, J. R., & Lambeck, K. (2008). Understanding global sea levels: past, present and future. *Sustainability Science, 3*, 9-22. doi: 10.1007/s11625-008-0042-4

Church, J. A., & White, N. J. (2011). Sea-level rise from the late 19th to the early 21st century. *Surveys in Geophysics, 32*, 585-602. doi:10.1007/s10712-011-9119-1

Foster, G., & Brown, P. T. (2014). Time and tide: analysis of sea level time series. *Climate Dynamics, 45*(1-2), 291-308. doi:10.1007/s00382-014-2224-3

Haasnoot, M., Kwadijk, J., Alphen, J. van, Bars, D. Le, Hurk, B. van den, Diermanse, F., Spek, A. van der, Essink, G. O., Delsman, J., & Mens, M. (2020). Adaptation to uncertain sea-level rise; how uncertainty in Antarctic mass-loss impacts the coastal adaptation strategy of the Netherlands. *Environmental Research Letters, 15*, 034007, 1-15. doi.org/10.1088/1748-9326/ab666c

Horton, B. P., Kopp, R. E., Garner, A. J., Hay, C. C., Khan, N. S., Roy, K., & Shaw, T. A. (2018). Mapping sea-level change in time, space, and probability. *Annual Review of Environment and Resources, 43*, 481-521. doi.org/10.1146/annurev-environ-102017-025826

IPCC, (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp. Retrieved from: https://www.ipcc.ch/report/ar5/syr/

Kulp, S. A., & Strauss, B. H. (2019). New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding. *Nature Communications, 10*, 4844, 1-12. doi.org/10.1038/s41467-019-12808-z

Ljung, G. M., & Box, G. E. O. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*(2), 297-303. doi: 10.2307/2335207

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to time series analysis and forecasting*. Hoboken, N.J.: John Wiley & Sons. Inc.

Srivastava, P. K., Islam, T., Singh, S. K., Petropoulos, G. P., Gupta, M., & Dai, Q. (2016). Forecasting Arabian sea level rise using exponential smoothing state space models and ARIMA from TOPEX and Jason satellite radar altimeter data. *Meteorological Applications, 23*, 633-639. doi:10.1002/met.1585

U.S. Global Change Research Program (USGCRP), (2017). *Climate science special report: Fourth National Climate Assessment*, Volume I. [Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, 470 pp. Retrieved from: https://science2017.globalchange.gov/

Visser, H., Dangendorf, S., & Petersen, A. C. (2015). A review of trend models applied to sea level data with reference to the “acceleration-deceleration debate”. *Journal of Geophysical Research: Oceans, 120*(6), 3873-3895. doi:10.1002/2015JC010716