Environmental Modelling through Chaotic Approach for Malaysian West Coast Sea Level.

Nazliah Mohd Ali* and Nor Zila Abd Hamid

Maritime Engineering Technology Section, Malaysian Institute of Engineering Technology, Universiti Kuala Lumpur, 32200 Lumut, Perak, Malaysia.
Department of Mathematics, Faculty of Sciences and Mathematics, Sultan Idris Education University, 35900 Tanjong Malim, Perak, Malaysia.

*nazliah@unikl.edu.my

Abstract. Sea level forecasting is an essential task for coastal engineering, geodetic application, navigation, and recreational activities. Predicting the behavior of future sea level is important for monitoring and forecasting of changes in fishery and marine ecosystems as well as for protection of coastal. This research is focused on the analysis and prediction of hourly sea level time series data at the benchmark station located in Penang by using chaotic approach. The purpose of this research was to identify the presence of chaotic behavior by the phase space reconstruction and Cao methods, and also a local linear approximation method is applied for prediction purposes. The results notified that the value of correlation coefficient between the observed and predicted time series is 0.8838 which is near to one. This reveals that the local linear approximation method can be used to predict the sea level time series in Malaysia. Certainly, the result of this research is expected to help stakeholder such as Department of Survey and Mapping Malaysia (JUPEM) in having a better sea level management.

1. Introduction

Sea water occupied almost 70% of the earth and its level rising every year. The rising of sea water level is not a good sign as it will give big impacts especially in low-elevation coastal zones. Most countries in South, Southeast, and East Asia including Malaysia expected to be at high risk due to large amount of population occupied the coastal area [1]. Besides, the low level islands such as Maldives or Tuvalu also expected to encounter the sea level rise risk and its consequences in the future [1]. Numerous researchers such as [2,3] have outlined that, the main causes of global mean sea level change are the additional freshwater input to due to land ice loss as well as from terrestrial reservoirs and thermal expansion of the sea waters in response to ocean warming. Due to this issue, predicting sea level is important as an increasing of sea level may lead to inundation, flood and coastal erosion.

Over the past few years, chaotic approach has attracted increasing attention and has been considered for application in environmental modeling. Chaotic approach has been applied successfully, for example, in modeling of sea level [4, 5], modeling of ozone [6], modeling of particulate matter [7] and modeling of river flow [8]. In Malaysia, there are a number of studies that have been successfully carried out by applying chaotic approach, for example, time series of ozone by [9], particulate matter by [10], and river flow by [11]. As far as the authors aware, there is no work carried out on sea level
time series in Malaysia by using the chaotic approach. Therefore, this study will contribute to the enhancement on the application of chaotic approach for the environmental modeling in Malaysia.

In this research, Cao method and Phase space plot are used to identify the presence of chaotic behavior in time series of sea level. If the presence of chaotic behavior has been identified, then the prediction model will be developed by using chaotic approach. In this study, the basic method from chaotic approach namely local linear approximation method is used to predict the observed time series. This method successfully applied by previous researchers [11, 12] and revealed very satisfactory results.

Overall, in this study modeling by using chaotic approach is divided into two parts: (i) the reconstruction of phase space which contributes to determination of chaotic behavior of the sea level time series and (ii) the development of prediction model which contributes to predict the future sea level time series.

2. Study Area and Data

2.1 Sea Level Data

The observed time series of sea level data was recorded hourly at the tidal gauge station of Penang, Malaysia. The period of the data used is 7 months from 1st June 2016 until 31st December 2016. The first 6 months of data which is 4392 data, were used as a training data and the last month of 2016 data which is 744 data were retained for testing the prediction performance purposes. The total of data for both training and testing is 5136.

Table 1. Descriptive Statistics for Penang

| Descriptive Statistics for Sea Level Time Series | Penang                        |
|-----------------------------------------------|-------------------------------|
| Mean                                          | 288.6701713                   |
| Standard Error                                | 0.745233297                   |
| Median                                        | 291                           |
| Mode                                          | 312                           |
| Standard Deviation                            | 53.40780856                   |
| Sample Variance                               | 2852.394016                   |
| Kurtosis                                      | -0.661916358                  |
| Skewness                                      | -0.110217053                  |
| Range                                         | 267                           |
| Minimum                                       | 148                           |
| Maximum                                       | 415                           |
| Sum                                           | 1482610                       |
| Count                                         | 5136                          |

Table 1 above shows the descriptive statistics of the data. The sea level data observed and recorded in centimetre (cm).

3. Methodology

3.1 Chaotic Approach

In this research, the chaotic approach will be used to discover the presence of chaotic dynamic. The
chaotic approach involves two phases, i.e. the reconstruction of phase space and prediction. The average mutual information and the Cao method are used in this study to detect the presence of chaotic behaviors. Meanwhile, the local linear approximation method will be used for prediction purposes.

3.2 Phase Space Reconstruction
According to Takens (1980), the phase space reconstruction is defined by:

\[ Y_i^m = \{ x_i, x_{i+\tau}, x_{i+2\tau}, x_{i+3\tau}, \ldots, x_{i+(d-\tau)} \} \]  

where \( \tau \) is the delay time and \( d \) is the embedding dimension. In this research, the average mutual information method was applied to calculate the proper delay time of the dynamic system while Cao method will be implemented to estimate the embedding dimension.

3.3 Determination of Time Delay, \( \tau \)
3.3.1 Average Mutual Information

\[ I(T) = \frac{1}{N} \sum_{n=1}^{N} p(u_a, u_{a+\tau}) \log_2 \left( \frac{p(u_a, u_{a+\tau})}{p(u_a) p(u_{a+\tau})} \right) \]  

Where \( p(u_a) \) and \( p(u_{a+\tau}) \) are the probability to get \( u_a \) and \( u_{a+\tau} \) respectively in \( X_{\text{training}} \), where \( p(u_a, u_{a+\tau}) \) is the joint probability density for \( p(u_a) \) and \( p(u_{a+\tau}) \). The graph \( T \) against \( I(T) \) is plotted and \( \tau \) is the first minimum value of \( T \) which gives minimum value of \( I(T) \).

3.4 Determination of Embedding Dimension, \( d \)
The value of \( d \) is required after the determination of \( \tau \) value have been done. \( d \) is the minimum number of variables needed to explain the dynamics of the time series data [13]. The pattern of an attractor can be described when \( d \) is at the most optimal value. To determine the most effective set of the dimensions, the value of \( d \) may varied which will give appropriate estimation outcomes. In this research, \( d \) is calculated by using the Cao method because other than searching parameter \( d \), this method can also distinguish between chaotic and random data [14].

This method was introduced by Cao [14] where the minimum embedding dimension is determined from a scalar time series. It has the following advantages where it contains the time-delay parameter for the embedding only. This mean, there is no other subjective parameters exist in this method. Besides, this method is independent on the availability of data which mean that this method is not relying to the quantity of data and can determine stochastic signals directly from the deterministic signals. The formula is defined by:

\[ E1(d) = \frac{E(d+1)}{E(d)} \]  

where

\[ E(d) = \frac{1}{N-d\tau} \sum_{n=1}^{N-d\tau} \left\| Y_n^{d+1} - Y_{jj}^{d+1} \right\| \left\| Y_n^d - Y_{jj}^d \right\| \]  

(4)
is the maximum norm. $Y_i^d$ is the nearest neighbor to $Y_n^d$. Graph $d$ against $E1(d)$ is plotted. If $E1(d)$ stop changing when the value of $d$ is greater than $d_0$, thus $d_0 + 1$ is the minimum embedding dimension which is the value of $d$ [10].

Cao[15] also introduced calculations in order to identify the presence of chaotic behavior. This method can differentiate between chaotic and random data by observing the value of $E2(d)$. $E2(d)$ is calculated by using:

$$E2(d) = \frac{E^*(d + 1)}{E^*(d)}$$

(5)

Where the value of $E^*(d)$ is given by:

$$E^*(d) = \frac{1}{N - d} \sum_{n=1}^{N - d} \left| x_{n+dr}^d - x_{ij+dr}^d \right|$$

(6)

If the chaotic dynamic is present in the observed time series, there is at least one $d$ where $E2(d) \neq 1$. Contrarily, if all the values of $E2(d)$ are equal to one, thus the observed time series is random.

3.5 Data Prediction Model

3.5.1 Local Linear Approximation

Local Linear Approximation Method (LLAM) is used to predict the data in sea level time series. LLAM has been widely used by previous researchers in order to predict the future value of chaos data. This method has been used by [15] in the study on forecasting monthly stream flow dynamics in the western United States. Besides, [16] along with [17] also using local linear approximation method to predict the data of river and sea level respectively.

In order to predict $Y_{t+1}$, the nearest neighbor(s) to $Y_t$ are searched. In this study the Euclidean distance between $Y_t$ and the vectors before $Y_i (i=1,2, \ldots, t-1)$ is calculated. Assume that the minimum distance to the nearest neighbor is $Y_m$, thus the value $Y_m$ and $Y_{m+1}$ are used to satisfy a linear equation $Y_{m+1} = AY_m + B$. Least square method is used to calculate the constant value of $A$ and $B$. Thus, the predictive value $Y_{t+1}$ can be calculated by using $Y_{t+1} = AY_t + B$.

3.6 Data Analysis Tools

After all data have been obtained, it must be analyze by using comprehensive tools to ensure the data is precise. The tools used to analyze the data are MATLAB 2009 and TSTools. These tools are used since the data is in a large volume and it is not possible to be done by using the manual method. Thus, these tools improved the process of analyzing the data and make the study efficient.

4. Results and Discussions

4.1 Phase space plot

Phase space plot is under phase space reconstruction method which can detect the presence of chaos in data. It can be detected by the existing of attractor in the plot. The graph of $\{x(t), x(t+\tau)\}$ were constructed using $\tau = 4$ to describe the chaotic behavior of data.
The Figure 1 presents the reconstruction of the Penang hourly sea level data series. The plot was constructed by using \( \{x(t), x(t+3)\} \) as the 4. It shows the same characteristics of plots where there is an existence of the attractor in the plot. To determine whether the data is chaotic or not is depends on the presents of the attractor in the systems. This confirms that the studied data of sea level is in chaotic behavior.

It shows that the nature of data is chaos and it is different with the random and linear data plot as there is an existence of attractor in the phase space plot as suggested by [17].

\[
\begin{align*}
Penang \text{ Phase Space Plot (} \tau = 4 \text{)}
\end{align*}
\]

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{penang_phase_space_plot}
\caption{Phase Space Plot}
\end{figure}

\subsection*{4.2 Cao Method}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{cao_method}
\caption{\( E1(d) \) and \( E2(d) \) from Cao Method}
\end{figure}

The results of \( E1(d) \) and \( E2(d) \) from Cao Method are depicted in Figure 2. It can be seen that \( E1(d) \) starts to saturated after \( d=7 \). Therefore, the minimum embedding dimension value is \( d=8 \). This indicates that the observed sea level time series at Penang influenced by at least 8 factors.

Moreover, according to Cao method, if \( E1(d) \) saturates with increasing \( d \), the chaotic behavior is presence in the data series. Clearly observed that at \( d=8 \), the value of \( E1(d) \) start to saturate. It can be concluded that \( E1(d) \) saturates continuously as \( d \) increases. Thus, the result shows the existence of the chaotic behavior. Furthermore, for \( E2(d) \), it can be seen that there exist \( E2(d) \neq 1 \). The existence of \( E2(d) \neq 1 \) can confirm that the presence of chaotic behavior in the observed sea level time series.

\subsection*{4.3 Prediction results}

The LLAM is used to predict the observed sea level data. In this research, the prediction is done for one month periods from 1\(^{st}\) to 31\(^{st}\) December 2016 (744 hours). The prediction models are developed by reconstructed the phase space of equation (1) using 4 and \( d = 8 \). The comparison between
predicted and observed value can be seen as shown in Figure 3. Clearly, it reveals that the trend of the data can be predicted well.

Figure 3. A comparison of predicted and observed sea level time series.

Based on the Table 2 below, the value of correlation coefficient (CC) between the observed and predicted data is \( r = 0.8839 \). This result indicates that there is a strong correlation between the observed and predicted data since the value of \( r \) is near to one. Overall, the presented results demonstrate that the local linear approximation method is good and reliable in predicting sea level time series data.

| Method | Local Linear Approximation |
|--------|---------------------------|
| Location | Penang |
| Time Delay | \( \tau = 4 \) |
| AAE | 20.5269 |
| RMSE | 24.7848 |
| CC | 0.8839 |

5. Conclusion
In this research, it is certify that the chaotic approach is a good approach which can be implemented to predict the sea level time series data at Malaysian west coast area Penang. This is because, it has satisfied the two objectives of this research which is to identify the presence of chaotic dynamics in time series of sea level at the selected location and to predict the sea level in selected areas using the development of the chaotic model.

Those proofs can be seen when the data of sea level in the location is in the chaotic state and this matter fulfill the objectives number 1 in this research. The phase space plot showed how the nature of data in which the existence of attractor in the plot can be seen. In addition, the value of \( E2(d) \) also plays an important role in determining the presence of chaotic behavior in data when the value of \( E2(d) \neq 1 \) through Cao Method.

Moreover, the Local Linear Approximation Method shows the accurate and precise result based on its correlation coefficient value which is in range of 0.7 to 1. Since its value is 0.8839 which is very good value, it shows that this approach have a strong correlation between the predicted and the observed data. Therefore, this method is significant to be used in other fields for future prediction.
From this research also, it is hope that this it will benefit to all peoples who live near coastal area for their awareness and preparation measure in order to counter the catastrophic disaster such inundation, coastal erosion and death of peoples. Besides, from this research, it can gives an opportunity for the govern bodies such as Department of Mapping & Survey Malaysia, National Hydraulic Research Institute Malaysia, Department of Town and Country Planning (JPBD), the Drainage and Irrigation Department (DID), the Public Works Department (PWD) and Local Authorities to take an early preparation in measuring and predicting the sea level.

Acknowledgment
Greatest appreciation to the Department of Mapping and Survey Malaysia for providing sea level data and assisted direct or indirectly upon completing this research.

References
[1] Nicholls, R. J., & Cazenave, A. (2010). Sea-level rise and its impact on coastal zones. Science, 328(5985), 1517-1520.
[2] Ahmad Radzi, A., & Ismail, H. (2013). Trend analysis of sea level rise for Kukup (Johor), west coast of peninsular Malaysia. In International Conference of Emerging Trends in Engineering and Technology.
[3] Cazenave, A., & Llovel, W. (2010). Contemporary sea level rise. Annual review of marine science, 2,145-173.
[4] Domenico, M. De, Ghorbani, M. A., Makarynskyy, O., Makarynska,D., & Asadi, H. (2013). Chaos and Reproduction in Sea Level. Applied Mathematical Modelling, 37(6), 3687–3697.
[5] Khatibi, R., Ghorbani, M.A., Aalami,M.T., Kocak, K., Makarynskyy, O., Makarynska,D., & Aalinezhad, M. (2011). Dynamics of hourly sea level at Hillarys Boat Harbour, Western Australia: a chaos theory perspective. Ocean Dynamics, 61(11), 1797-1807.
[6] Kocak, K., Saylan, L., & Sen, O. (2000). Nonlinear Time Series Prediction of O3 Concentration in Istanbul. Atmospheric Environment, 34(8), 1267–1271.
[7] Chelani, A. B., & Devotta, S. (2006). Nonlinear Analysis and Prediction of Coarse Particulate Matter Concentration in ambient Air. Journal of the Air & Waste Management Association, 56(1), 78–84.
[8] Khatibi, R., Sivakumar, B., Ghorbani, M. A., Kisi, O., Koçak, K., & Zadeh,D. F. (2012). Investigating chaos in river stage and discharge time series. Journal of Hydrology, 414, 108-117.
[9] Hamid, N. Z. A., Noorani, M. S. M., Juneng, L., & Latif, M. T. (2013). Prediction of Ozone Concentrations Using Nonlinear Prediction Method. AIP Conference Proceedings, 1522, 125–131.
[10] Hamid, N. Z. A., & Noorani, M. S. M. (2014). A Pilot Study using Chaotic Approach to Determine Characteristic and Forecatng of PM10 Concentration Time Series. Sains Malaysia, 43(3), 2014.
[11] Adenan, N. H., & Noorani, M. S. M. (2013). River Flow Prediction Using Nonlinear Prediction Method. International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering, 7(11), 62 – 66.
[12] Adenan, N. H., & Noorani, M. S. M. (2013). River flow prediction using nonlinear prediction method. World Academy of Science, Engineering and Technology, International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering, 7(11), 1589-1592.
[13] Regonda, S., Rajagopalan, B., Lall, U., Clark, M., & Moon, Y. I. (2005). Local polynomial method for ensemble forecast of time series. Nonlinear Processes in Geophysics, 12(3), 397-406.
[14] Cao, L. (1997). Practical method for determining the minimum embedding dimension of a scalar time series. Physica D: Nonlinear Phenomena, 110(1-2), 43-50.

[15] Sivakumar, B. (2003). Forecasting monthly streamflow dynamics in the western United States: a nonlinear dynamical approach. Environmental Modelling & Software, 18(8), 721-728.

[16] Khatibi, R., Sivakumar, B., Ghorbani, M. A., Kisi, O., Koçak, K., & Zadeh, D. F. (2012). Investigating chaos in river stage and discharge time series. Journal of Hydrology, 414, 108-117.

[17] Sivakumar, B. (2002). A phase-space reconstruction approach to prediction of suspended sediment concentration in rivers. Journal of Hydrology, 258(1), 149-162.