Journal of Quality Measurement and Analysis *JQMA* 17(2) 2021, 1-8
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*

A MATHEMATICAL MODELLING THROUGH CHAOTIC APPROACH TO FORECAST THE SEA LEVEL IN PENANG

(Pemodelan Matematik Melalui Pendekatan Kalut bagi Meramal Paras Laut di Pulau Pinang)

NAZLIAH MOHD ALI*, NORZILA ABD HAMID & FATIN ZAWANI ZAINAL AZAIM

ABSTRACT

Forecasting of sea level is significant since a rising sea level can cause erosion, flood, and inundation. Prediction of the behaviour of a future sea level is crucial for coastal engineering, geodetic applications, navigation, coastal ecosystems, and recreational activities along with observation and prediction of the changes in fisheries and marine environments also coastal fortifications. The data examined is a time series of sea levels recorded hourly in a high-risk area at Penang tidal gauge station. The chaotic approach technique was used to analyze the observed time series data which was divided into two steps: phase space reconstruction and prediction. The purpose of the study comprises to discover the existence of chaotic nature through the phase space reconstruction process and the Cao method. A local linear approach has been used for prediction purposes. The findings indicated that the coefficient of correlation value among the data observed and data predicted was .9356. The result implies the Malaysian sea level of time series can be predicted using the local linear approximation method. These findings are anticipated to support agencies in particular the Department of Survey and Mapping Malaysia (JUPEM) to organize improved management of the sea level.

Keywords: Cao method; chaotic approach; local linear approximation method; phase space approach; sea-level prediction

ABSTRAK

Peramalan paras laut adalah signifikan kerana kenaikan paras laut boleh menyebabkan hakisan, banjir, dan penggenangan air. Peramalan paras laut pada masa hadapan juga sangat penting untuk kejuruteraan pantai, aplikasi geodesi, navigasi , ekosistem pantai dan kegiatan rekreasi juga bagi tujuan pemerhatian dan peramalan perubahan perikanan di persekitaran laut dan pantai. Kajian ini menggunakan data siri masa paras laut yang dicerap mengikut jam di kawasan berisiko tinggi yang terdapat di stesen pasang surut, Pulau Pinang. Data siri masa yang dicerap dianalisis menggunakan kaedah pendekatan kalut yang membabitkan dua proses iaitu pembinaan semula ruang fasa dan proses peramalan. Tujuan kajian adalah untuk mengesan kehadiran dinamik kalut menggunakan proses pembinaan semula ruang fasa dan kaedah Cao. Kaedah penghampiran linear setempat pula digunakan untuk tujuan peramalan. Hasil kajian menunjukkan nilai pekali korelasi antara data yang diperhatikan dan data yang diramalkan ialah .9356. Keputusan membuktikan bahawa data siri masa paras laut Malaysia dapat diramalkan menggunakan kaedah penghampiran linear setempat. Dapatan ini diharapkan dapat membantu agensi-agensi khususnya Jabatan Ukur dan Pemetaan Malaysia (JUPEM) dalam pengurusan paras laut.

Kata kunci: kaedah Cao; pendekatan kalut; penghampiran linear setempat; pembinaan semula ruang fasa; peramalan paras laut

1. Introduction

There is no doubt that the future of the sea level certainly will give a physical, economic and social impact especially on Malaysia's west coastal area (Sarkar *et al*. 2014). The immediate consequences include submergence and increased flooding and inundation, and also coastal erosion and saltwater invasion (Nicholls *et al*. 2010; Cazenave *et al*. 2010). Penang Island is an important economic hub in Malaysia, with the majority of its population living along the coast. Over the last three decades, Penang Island has experienced an average rate of sea-level rise of 3.2 mm/year (Gao *et al*. 2021). The worst impact occurred in 2004, where Tsunami splashed the Penang coastal area and killed over 52 people. Many facilities were damaged as a result of the tragedy, and the Penang government had to spend a huge amount of money to recover. These impacts will affect the continuous development of Malaysia especially in economic and social aspects which require a huge amount of cost accumulated just for the recovery process (Davies 2017). Moreover, Ghazali *et al*. (2018) examined the effects of both sea-level rise and tsunami on possibly inundated locations of Penang Island. The findings revealed that 30 percent of coastal land would be drowned in the worst-case tsunami inundation model, despite the fact that the majority of people reside in these areas. As a consequence of this issue, the development of the sea level prediction models is essential for improving the management of the sea level.

Chaos theory has drawn more attention and was proposed for use in modelling various time series. This technique has been widely used in various fields, particularly in hydrology, which it has been used to predict sea level (Domenico *et al*. 2013; Khatibi *et al*. 2011), ozone (Kocak *et al*. 2000), particulate matter (Chelani & Devotta 2006), and river flow (Khatibi *et al*. 2012). There is a variety of research that has been well implemented by using a chaotic approach in Malaysia, such as, modelling of ozone by Zaim dan Hamid (2017) and Abd Hamid *et al*. (2017), modelling of carbon monoxide by Jusoh *et al*. (2021) and temperature by Abd Hamid (2021). On the other hand, Mashuri (2021) has focused on the modelling of river flow. As regarded by researchers, there are still no studies yet been performed on the use of chaos theory on the prediction of sea level time series in Malaysia. Therefore, this study will encourage the development of mathematical modelling specifically the implementation of chaos theory in Malaysia.

Generally, the chaotic analysis is distributed into two sections: (i) identifying the chaotic nature of sea level data and (ii) forecasting the time series of sea level data. To determine the presence of chaotic dynamics in sea level data, the Cao method and the phase space diagram have been used. Once the existence of a chaotic nature was established, the prediction model would be constructed using the chaotic approach. The basic technique of the chaotic approach, i.e. the local linear approximation approach, has been used to forecast the observed data. Previous researchers (Hamid & Noorani 2014; Adenan & Noorani 2013) successfully applied this method and discovered adequate results.

2. Methodology

2.1. *Data description*

The time series of sea level data was observed and recorded hourly at an existing tide gauge station (Figure 1) in Penang, Malaysia. The duration of the data employed was seven (7) months, starting at 1 a.m. on 1st June 2016 and ending at 11 p.m. on 31st December 2016. Therefore, this study involved a total of 5136 data. The data is presented in the form of a scalar time series:

$$X = \{x_1, x_2, x_3, \ldots, x_w\}\ , \tag{1}$$

where w represents the total number of hours of observation. In this study, w=5136. Data from the first six months, i.e. 4392 data, have been used as training data where:

$$X_{training} = \{x_1, x_2, x_3, \ldots, x_{4392}\}. \tag{2}$$

Meanwhile, for verification of the prediction performance, data from the last month of 2016, that is 744 data, has been retained and recorded as:

$$X_{test} = \{x_{4393}, x_{4394}, x_{4395}, \ldots, x_{5136}\}. \tag{3}$$

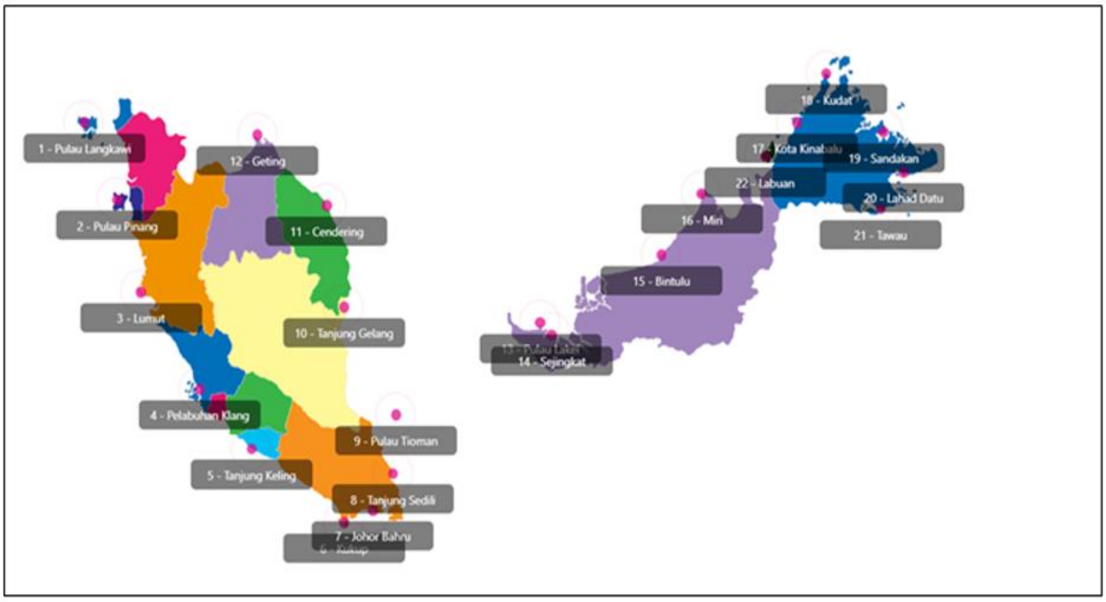

Figure 1: Existing tidal gauge in Malaysia (https://www.jupem.gov.my).

2.2. *Chaos theory*

The chaotic method involves two phases, specifically phase space reconstruction and prediction. In the study, The Cao method and the average mutual information function were used to identify the existence of a chaotic nature. Meantime, for prediction purposes, a local linear approximation method was applied.

2.3. *Phase space reconstruction*

The existence of chaos in data can be discovered by the construction of the phase space diagram. It can be defined by the presence of an attractor in the diagram.
The time series of training data in (2) is reconstructed to d-dimensional vector:

$$\mathbf{Y}_i^d = \{x_i, x_{i+\tau}, x_{i+3\tau}, \ldots, x_{i+(d-\tau)}\}. \tag{4}$$

Here d is a chosen embedding dimension and τ is a time delay. For the construction of (4), the embedding dimension, d and the time delay, τ must be determined. The average mutual information approach has been employed to compute the optimum time delay of a dynamic system, whereas the Cao method has been applied to determine the sufficient embedding dimension. According to Fraser & Swinney (1986), the average mutual information function can be described as:

$$I(T) = \frac{1}{N}\sum_{a=1}^{N} p(u_a, u_{a+T}) \log_2 \left[\frac{p(u_a, u_{a+T})}{p(u_a)\, p(u_{a+T})} \right], \tag{5}$$

where N is the total number of time series data meanwhile u is the observed time series data at a time a and $a+T$. $p(u_a)$ and $p(u_{a+T})$ are the probability to obtain u_a and u_{a+T} respectively on $x_{training}$, whereas $p(u_a, u_{a+T})$ is the joint probability of $p(u_a)$ and $p(u_{a+T})$. The graph T is plotted against $I(T)$ and τ would be the first minimum value T that represents the minimum value $I(T)$.

To discover the sufficient embedding dimension, d, an optimal τ has been used as an input. d is defined as the optimal number of factors necessary to depict the behaviour of the data (Regonda *et al*. 2005). In this study, the Cao method was preferred to compute d since other than discovering parameters d, this method also can be used to determine the existence of chaotic nature (Cao 1997). The pattern of the attractor could be portrayed once d is at the most optimum value. The formula is described by the following:

$$E1(d) = \frac{E(d+1)}{E(d)}, \tag{6}$$

where

$$E(d) = \frac{1}{N - d\tau} \sum_{n=1}^{N-d\tau} \frac{\left\| \mathbf{Y}_n^{d+1} - \mathbf{Y}_{jj}^{d+1} \right\|}{\left\| \mathbf{Y}_n^{d} - \mathbf{Y}_{jj}^{d} \right\|}. \tag{7}$$

$\|\bullet\|$ is the maximum norm. $\mathbf{Y}_{jj}^{d}$ is the nearest neighbour to $\mathbf{Y}_n^{d}$. Hence, the graph d is plotted against $E1(d)$. When $E1(d)$ is saturated if the value d is exceeding d_0, hence $d_0 + 1$ is considered to be a minimum embedding dimension (Hamid *et al*. 2013).

Cao (1997) also developed a calculation to determine the existence of chaotic behaviour. This approach may distinguish between random and chaotic data by perceiving the values of $E2(d)$. $E2(d)$ has been computed using the following:

$$E2(d) = \frac{E(d+1)}{E(d)}. \tag{8}$$

If for one or more d where $E2(d) \neq 1$ the chaotic dynamics have appeared in the observed time series. On the other hand, if the values $E2(d)$ are equivalent to 1 for any d, therefore, the observed time series behaviour is random.

2.4. *Local linear approximation model (LLAM)*

The Prediction in this study was performed through a Local Linear Approximation Model (LLAM). This technique is successfully used by Sivakumar (2003) in the research on monthly streamflow dynamics. Moreover, Zakaria *et al*. (2021) apply the local linear approximation model to forecast water level time series.

For LLAM, a linear equation $Y_{n+1} = AY_n + B$ must be fitted to the training set (2). The least square method is used to compute the constant values of A and B. Once A and B have been determined, the prediction is done through the equation $x_{n+1} = Ax_n + B$, with the real value of x_n. For example, to predict x_{4393}, x_{4392} is used ($x_{4393} = Ax_{4392} + B$).

3. Results and discussion

3.1. *Phase space plot*

To illustrate the chaotic behaviour of data, the graph $\{x(t), x(t+\tau)\}$ was fabricated by using $\tau = 4$. The reproduction of the Penang hourly sea level data has been fabricated using $\{x(t), x(t+4)\}$. The result is represented in Figure 2. As can be seen, the graph depicts a deterministic structure in the trajectory, indicating the presence of an attractor within the plot.

Identifying whether the time series data is in a chaotic nature or not depends on the existence of attractors in dynamic systems. The graph reveals that the behaviour of the data is contradictory compared to the random and linear data since there is a presence of the attractors in the phase space plot (Sivakumar 2002). This verifies that the experimental data of the sea level time series is in chaotic nature.

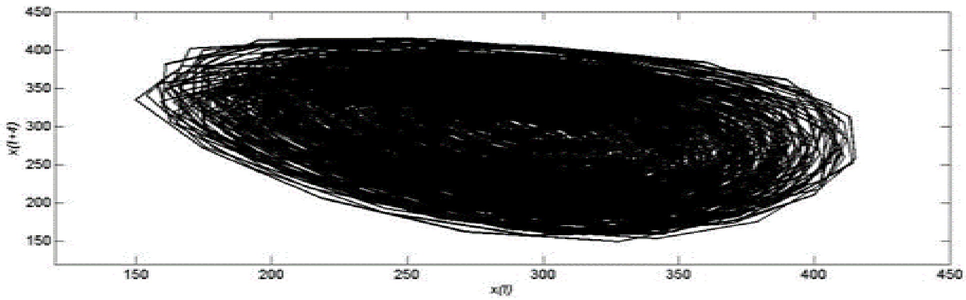

Figure 2: Phase space diagram ($\tau = 4$)

3.2. *Cao method*

Figure 3 describes the findings of $E1(d)$ and $E2(d)$ from the Cao Method. It can be perceived that $E1(d)$ begins to saturate after $d_0 = 7$. Consequently, the value of the minimum embedding dimension is $d = 8$. As stated by Regonda *et al*. (2005), d is the minimum number of factors that influence the dynamics of a time series. Hence, the result demonstrates at least 8 factors influence the observed time series of sea level in Penang. According to Cao (1997), the presence of chaotic dynamics in the data series can be determined if $E1(d)$ saturates as d increases. It has been observed that the values of $E1(d)$ beginning to saturate at $d = 8$. It could be inferred that $E1(d)$ is constantly saturated as d increases. Therefore, the outcome demonstrates the presence of a chaotic nature. Moreover, this finding can be validated by the determination of $E2(d)$ through Cao method. Figure 3 displays the result of $E2(d)$ which indicates the existence of $E2(d) \neq 1$. The presence $E2(d) \neq 1$ indicates the existence of a chaotic nature in the sea level data (Cao 1997).

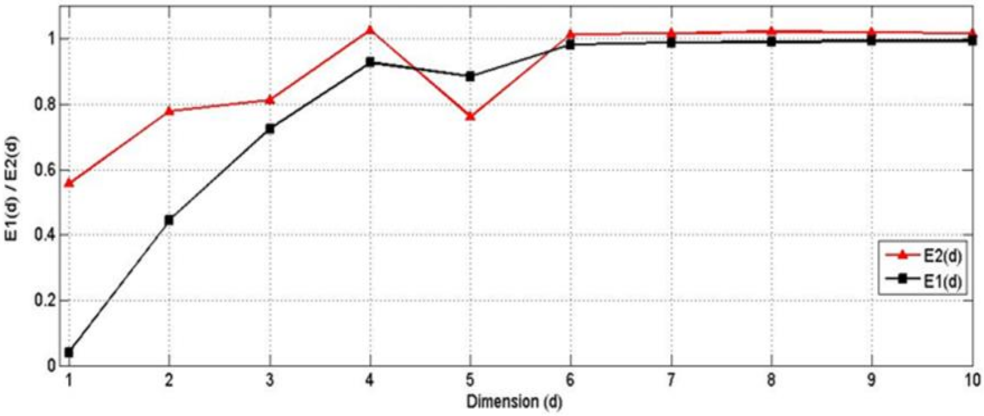

Figure 3: Cao method

Conclusively, the LLAM was adopted to forecast the observed sea level data. This study forecasts a period of one month of data (744 hours) from 1st to 31st December 2016. The forecasting model was fabricated by reconstructing the phase space of equation (1) with $\tau = 4$ and $d = 8$. The correlation between the values forecasted and the values observed is demonstrated in Figure 4. Obviously, it indicates that the data trend can be well predicted.

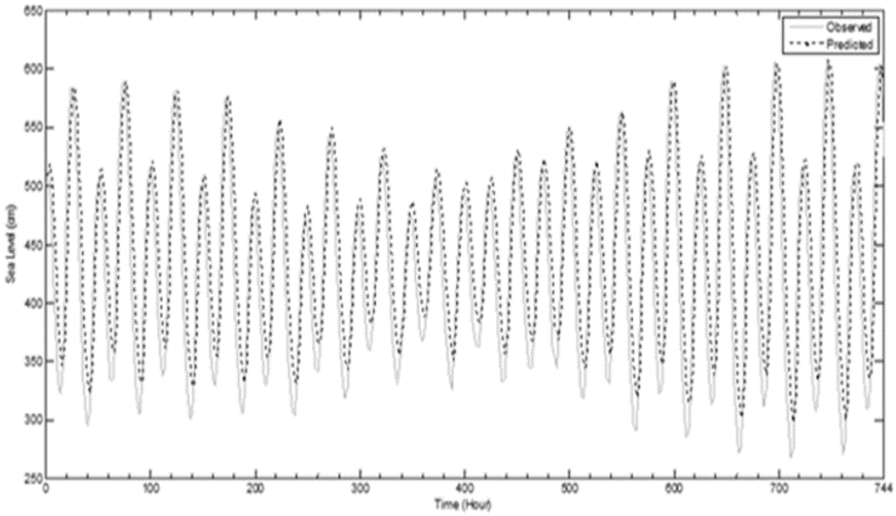

Figure 4: Prediction results

The performance of the prediction model can be determined by using the performance indicator, the coefficient of correlation (cc). The cc value among the experimental data and predicted data is r = 0.9356. Since the value of r is more than 0.8, this implies that there is a very strong positive relationship between the experimental and the predicted data (Schober *et al*. 2018). Thus, the results obtained prove that the LLAM was reliable and applicable to predict the time series of sea level data.

4. Conclusion

This study shows that the chaotic method is a proven method that could be employed to forecast the Penang sea level data. This analysis satisfies all the purposes of the study i.e., identifying the existence of chaotic behaviour in a time series of sea levels at the particular area as well as forecasting sea levels in a selected area using the chaotic models.

Such evidence can be observed when the sea level data at the chosen area is in chaotic nature. The nature of data is demonstrated by the phase space plot whereby the presence of the attractor is shown in the plot. Moreover, the value of $E2(d)$ has played a significant role to identify the existence of chaotic behaviour in the data when there exists $E2(d) \neq 1$ through the Cao method.

Additionally, it was found that the LLAM approach displays an excellent result according to the value of the coefficient of correlation between 0.7 and 1. As the result is 0.9356, it proved that this method implies a very good relationship between the data predicted and the data observed. Consequently, this present approach theoretically can be used for future prediction in many other areas.

In implication, the study is believed to support communities who live along the coastal region for their awareness to encounter devastating catastrophe including the flood, coastal erosion, inundation, and casualties. In the future, it is expected that this study would support agencies, particularly, the Department of Mapping & Survey Malaysia, Drainage and Irrigation Department (DID), Department of Town and Country Planning (JPBD), National Hydraulic Research Institute Malaysia (NAHRIM), and local authorities in measurement and management of the sea levels.

Acknowledgments

Special thanks to the Department of Mapping and Survey Malaysia (JUPEM) in favour of supplying the required sea level time series data.

References

Abd Hamid N.Z. 2021. A pilot study using chaos theory to predict temperature time series in Malaysian semi urban area. *Turkish Journal of Computer and Mathematics Education* **12**(3): 997-1003.

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.

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.

Cazenave A. & Llovel W. 2010. Contemporary sea level rise. *Annual Review of Marine Science* **2**: 145-173.

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.

Davies R. 2017. Malaysia-severe storm and floods leave 7 dead, 10,000 displaced. https://floodlist.com/asia/malaysia-penang-kedah-floods-november-2017 (31 December 2020).

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.

Fraser A.M. & Swinney H.L. 1986. Independent coordinates for strange attractors from mutual information. *Physical review A* **33**(2): 1134-1140.

Gao G., San L.H. & Zhu Y. 2021. Flood inundation analysis in Penang Island (Malaysia) based on InSAR maps of land subsidence and local sea level scenarios. *Water* **13**(11): 1518.

Ghazali N.H.M., Awang, N.A. Mahmud, M. & Mokhtar, A. 2018. Impact of sea level rise and tsunami on coastal areas of North-West Peninsular Malaysia. Irrigation and Drainage **67**: 119-129.

Hamid N.Z.A. & Noorani M.S.M. 2014. A pilot study using chaotic approach to determine characteristic and forecating of PM_{10} concentration time series. *Sains Malaysiana* **43**(3): 2014.

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**(1): 125-131.

Jusoh K.C., Abd Hamid, N.Z. & Side S. 2021. Forecasting through the chaotic approach on carbon monoxide time series in industrial area. *Journal of Science and Mathematics Letters* **9**: 55-62.

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.

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.

Kocak K., Saylan L. & Sen O. 2000. Nonlinear time series prediction of O3 concentration in Istanbul. *Atmospheric Environment* **34**(8): 1267-1271.

Mashuri A. 2021. Chaotic identification of hourly and daily water level time series data in different areas of elevation at Pahang River. *Turkish Journal of Computer and Mathematics Education* **12**(3): 2427-2435.

Nicholls R.J. & Cazenave A. 2010. Sea-level rise and its impact on coastal zones. *Science* **328**(5985): 1517-1520.

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.

Sarkar M., Kabir S., Begum R.A., Pereira J.J., Jaafar A.H. & Saari M.Y. 2014. Impacts of and adaptations to sea level rise in Malaysia. *Asian journal of Water, Environment and Pollution* **11**(2): 29-36.

Schober P., Boer C. & Schwarte L.A. 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia* **126**(5): 1763-1768.

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

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

Zaim W.N.A.W.M. & Hamid N.Z.A. 2017. Peramalan bahan pencemar ozon (O_3) di Universiti Pendidikan Sultan Idris, Tanjung Malim, Perak, Malaysia mengikut monsun dengan menggunakan pendekatan kalut. *Sains Malaysiana* **46**(12): 2523–2528.

Zakaria N.H., Adenan N.H., Abd Karim N.S. & Mashuri A. 2021. Prediction of water level time series data for dam at Selangor using chaotic approach and local linear approximation method. *Journal of Science and Mathematics Letters* **9**: 10-17.

Department of Mathematics,
Faculty of Sciences and Mathematics,
Sultan Idris Education University,
35900 Tanjong Malim, Perak, Malaysia.
email: nazliahbintimohdali@gmail.com, nor.zila@fsmt.upsi.edu.my*

Universiti Kuala Lumpur Malaysian Institute of Marine Technology (MIMET),
Dataran Industri Teknologi Kejuruteraan Marin
Bandar Teknologi Maritim, Jalan Pantai Remis,
32200 Lumut, Perak, Malaysia.
email: fatinzawani@unikl.edu.my

Received: 9 October 2021
Accepted: 14 December 2021

[*]Corresponding author