ASM Sc. J., 17, 2022

https://doi.org/10.32802/asmscj.2022.1159

Prediction Modelling Through Chaotic Approach on Ozone (O_3) Pollutant Time Series in Shah Alam City

Ahmad Basri Ruslan[1*], Nor Zila Abd Hamid[2], Khairunnisa Che Jusoh[2], Shahidatul Idayu Shahizam[2] and Franklin Chai Jun[1]

[1]*Sekolah Menengah Sains Pendang @ Akademi Sains Pendang (ASP), Jalan Jenun, 06700, Pendang, Kedah, Malaysia*

[2]*Universiti Pendidikan Sultan Idris, Jabatan Matematik, 35900, Tanjung Malim, Perak, Malaysia*

Studies on Ozone (O_3) particles and their effect have become a significant concern. Exposure to high concentrations of O_3 can cause an adverse reaction in the human respiratory system. Therefore, this study aims to predict the monthly O_3 time series in the highly populated area of Shah Alam through a chaotic approach. The phase space plot and Cao methods successfully detected the chaotic behaviour O_3 time series data. The three parameters are determined before the prediction process: time delay t, embedded dimension m, and nearest neighbour k. This study used $t = 1$ and the value of m is calculated through the Cao method. The last parameter that needs to be determined is by graph plotting k versus correlation coefficient(cc). The combination of parameters t and m will be used for prediction, and the performance measure will be recorded. The prediction process is done using three non-linear methods, namely LMAM, LLAM and ILLAM. LMAM gives the best prediction value by using the combination of parameter t and the value of the performance measure is 0.8486. Trial-and-error method, m = 6, LMAM gives cc = 0.8977 with k = 17, LLAM gives cc = 0.8033 with k =1, and ILLAM gives cc = 0.8265 with k =24. Mean-top-bottom method, m = 9, LMAM gives cc = 0.8863 with k =10, LLAM gives cc = 0.8841 with k = 6 and ILLAM gives cc = 0.9370 with k = 22. This finding indicates that the O_3 time series can be predicted using a chaotic approach and the improved method in determining the value can compute better prediction performance.

Keywords: ozone; chaotic approach; non-linear; phase space plot

I. INTRODUCTION

Ozone (O_3) is naturally formed in between stratosphere and troposphere levels. National Aeronautics Space Administration (NASA) defines O_3 as a natural gas formed from three oxygen atoms (NASA, 2018). O_3 layer is used to protect the earth from ultraviolet rays. The troposphere level is the lowest. The gas composition in the troposphere is adequate for humans to breathe. However, breathing with O_3 is toxic to human health at the troposphere level. It may cause an adverse reaction in the respiratory system (Cisneros *et al.*, 2010). United States Environmental Protection Agency (EPA) founds that high concentrations of O_3 at the troposphere level are dangerous for health and categorised as toxic (EPA, 2017). At the troposphere level, O_3 is formed through a chemical reaction of volatile organic compounds (VOC, i.e. NO_2, CO, etc.) and a free oxygen atom in the air with sunlight as the catalyst. VOC is released by the combustion of vehicle engines and factories from industries. When the emission of waste from vehicles and industries increases, then the formation of O_3 will also increase. The nature of O_3 is not soluble in water. It is freely moving at the troposphere level (Nuvolone *et al.,* 2017). It makes O_3 is highly dangerous at ground level.

There are five types of pollutants observed by the Department of Environment Malaysia (DOE) which are particulate matter below 10 millimetres (PM_{10}), carbon monoxide (CO), nitrogen monoxide (NO), ozone (O_3) and sulfur dioxide (SO_2). At the end of 2019, 8.3 million deaths are reported due to air pollution (McCarthy, 2019). Factors that affected pollution are the economy and the number of

*Corresponding author's e-mail: ahmadbasriruslan@gmail.com

populations (Myllyvirta, 2020). India and China are the top two countries that are highly populated in the world. These two countries also give the highest number of death due to pollution which is 1.243 million deaths (China) and 1.241 million deaths (India) at the end of 2019. It is proven that a highly populated country may lead to high pollution. Therefore, this study aims to predict the monthly O_3 time series in the highly populated area of Shah Alam through a chaotic approach.

Prediction modelling through a chaotic approach in Malaysia is new. Besides, there are several methods in prediction such as artificial neural network (ANN), multilinear regression, support vector machine, fuzzy logic and many more. These methods require many variables such as meteorology, gas and many more before predictions can be made (Hamid & Noorani, 2014). While through a chaotic approach, this method only requires previous time series data to make the prediction. For example, to predict the CO time series, only previous time series data of CO is required to make the prediction. Several predictions had been successfully done by using chaotic approach onto river flow time series (Adenan & Noorani, 2013), O_3 time series (Hamid & Noorani, 2013), PM_{10} time series (Hamid & Noorani, 2014), temperature time series (Bahari & Hamid, 2019), sea level time series (Ali & Hamid, 2019), CO time series (Ruslan & Hamid, 2019), CO time series (Ruslan *et al.,* 2020a) and SO_2 time series (Ruslan *et. al.,*. 2020b; Ruslan *et al.,* 2020c) These studies show excellent prediction performance by using correlation coefficient. That is why this study also uses the chaotic approach for prediction.

Study through chaotic approach involves two steps which are i) reconstruction of phase space and ii) prediction process. First, the dynamical of time series needs to be analysed to verify chaotic behaviour in the chosen time series. Time series are divided into deterministic and random (Abarbanel, 1996; Sprott, 2003). Chaos is between deterministic and random. To proceed this study for prediction process, the presence of chaotic behaviour needs to be confirmed. If the chaotic behaviour does not exist in the observed time series, the study cannot proceed to the prediction process. There are several methods to detect the presence of chaotic behaviour in the chosen time series. This study will use phase space and the Cao method to detect the presence of chaotic behaviour. If the chaotic behaviour exists, the prediction process can be done. Phase space plot method had been used to detect the presence of chaotic behaviour in onto river flow time series (Adenan & Noorani, 2013), O_3 time series (Hamid & Noorani, 2013), PM_{10} time series (Hamid &N Noorani, 2014), temperature time series (Bahari & Hamid, 2019), sea level time series (Ali & Hamid, 2019), CO time series (Ruslan & Hamid, 2019) and rainfall time series (Mashuri *et al.*, 2019). The Cao method was chosen because: i) does not involve another parameter except τ and ii) is not dependable on the number of time series data used (Cao, 1997). Therefore, the phase space plot and the Cao method are chosen to detect chaotic behaviour in this study.

Before starts with the prediction process, three parameters need to be determined, which are delay time (τ), embedded dimension (m) and the number of nearest neighbours (k). The choice of parameter τ is important to fully capture the structure of the attractor (Velickov, 2004). If τ is too big, different coordinates will not correlate and may cause the loss of information from the original system (Regonda *et al.,* 2005). The previous study uses $\tau = 1$, and the final result shows an excellent correlation coefficient (cc) between the original data and prediction onto hydrology and air pollution data (Sivakumar, 2002; Hamid, 2018). For m, it will be calculated through Cao method for getting the optimum value of m (Velickov, 2004). In the previous study, k is chosen with $k = 2m$, where m is the embedded dimension (Adenan & Noorani, 2015). Besides, several another few studies used the trial-and-error method in determining the value of parameter k, such as $k = 50$ (Hamid & Noorani, 2014; Jayawerdana, 1997; Zaim & Hamid, 2017; Hamid *et al.,* 2017), $k = 100$ (Hamid *et al.,* 2017) and $k = 200$ (Hamid & Noorani, 2013). In this study, an optimal of k will be analysed through graph plotting cc against k testing all numbers k from 1 until 200. The value k that gives the maximum value cc will be chosen as the best model and will be recorded. Three methods for predictions will be used, which are Local Mean Approximation Method (LMAM), Local Linear Approximation Method (LLAM) and Improved Local Linear Approximation Method (ILLAM). These three methods are known as chaotic models because they involve reconstructing phase space at m - dimension.

As mentioned above, a different time series will give a different prediction result. This study is focusing on O_3 time series in an urban area in Shah Alam. While for the previous study, the focus of the study was in a different geographical area which is in Shah Alam (Hamid *et al.,* 2013). Hence, it will give different prediction results. In addition, the previous study does not find an optimal parameter k before continuing with the prediction. In this study, the optimal number will be used to make the prediction model predict a better value by using cc it as the performance measure.

A. Time Series

The monthly reading of O_3 real-time series data is prepared by the Department of Environment Malaysia between 1st of June until 30th of June 2014 at Shah Alam Station. The O_3 time series are recorded in the unit of part per billion (ppb). Shah Alam is a city and state capital of Selangor, Malaysia, situated within Petaling District and a small part of the neighbouring Klang District. Shah Alam city is filled with residential, commercial centres, an education hub and recreational attractions. Summary from the Department of Statistics Malaysia (2010) shows that Shah Alam is ranked tenth among 20 cities with a large population in Malaysia (DOSM, 2010). Population density affects the number of pollutants released (Idris & Mahmud, 2017). The time-series data are recorded hourly for 30 days. The total number of data used is 720 hours. The O_3 time series is recorded in the one-dimensional form vector X :

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

N is the total number of data used in hours. In this study, $N = 720$. The overall time series for all 720 consecutive hours is illustrated in Figure 1. Statistical details for O_3 concentration time series is recorded in Table 1. Next, the data is divided into two different parts. The first part is the training time series for analysis purposes, and the other is the testing time series for measuring the prediction model's performance. The training time series is for three weeks long, 552 hours, while the test is one week long, which are 168 hours. These two parts are written as follows:

$$X_{train} = \{x_1, x_2, x_3, \ldots x_{552}\} \tag{2}$$

$$X_{test} = \{x_{552}, x_{553}, x_{554}, \ldots x_{720}\} \tag{3}$$

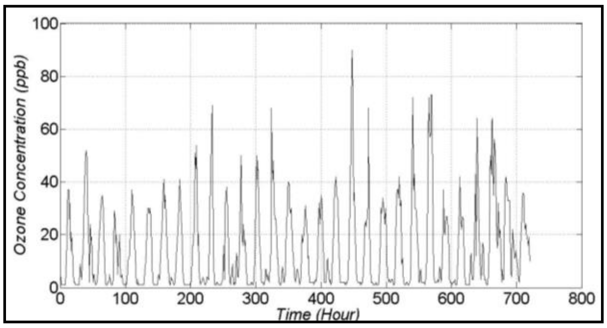

Figure 1. Hourly O_3 concentration time series

II. CHAOTIC APPROACH

A. Phase Space Plot

The training time series is recorded accordingly, like Equation (2). With x_t is the concentration of O_3 time series, the two-dimensional graph is plotted on an axis $\{x_t, x_{t+1}\}$. Before that, the value for τ the need to be determined. A few methods previously are used in determining the value of τ such as average mutual information (Fraser & Swiney, 1986), autocorrelation method (Schuster, 1988) and many more. However, few past studies used $\tau = 1$ real-time series such as flood time series (Lakshmi & Tiwari, 2009), temperature time series (Bahari & Hamid, 2019) and suspended sediment concentration (Sivakumar, 2002). By setting $\tau = 1$, the prediction performance shows an excellent result. Hence, $\tau = 1$ is set to be used in this study.

After set parameter $\tau = 1$, phase space plot $\{x_t, x_{t+1}\}$ is built. The existence of an accumulator or strange attractor on the graph shows the presence of chaotic behaviour on the observed time series (Sivakumar, 2002; Lakhsmi & Tiwari, 2009). Figure 2 shows the strange attractor on the phase space plot for O_3 time series in Jun 2014 in Shah Alam. Therefore, O_3 time series is chaos.

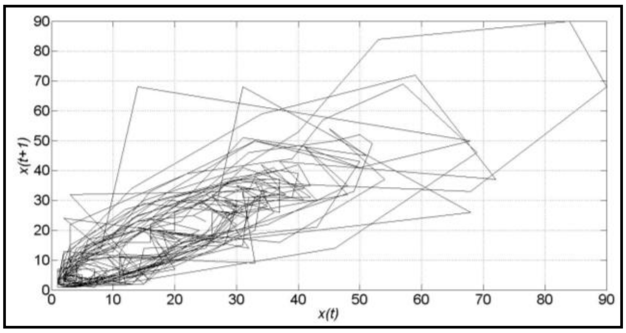

Figure 2. Phase space plot

B. Cao Method

Training time series data are reconstructed into m -embedded dimension:

$$Y_j^m = (x_j, x_{j+\tau}, x_{j+2\tau}, \dots, x_{j+(m-1)\tau}) \quad (4)$$

The value for τ has already set to 1, while parameter m is set through the Cao method. According to (Regonda *et al.,* 2005), m is the minimum embedded value needed for explaining the nature of the data. The nature of the data can be characterised as low dimensionally chaotic if the value m is less than 10 (Hamid & Noorani, 2017). Chaos at a low dimensional means that few variables can only explain the observed time series. Cao method is not limited to determining the value of m. This method also can be used to detect the presence of chaotic behaviour (Cao, 1997). There are two parameters included in the Cao method, which are $E1(m)$ and $E2(m)$.

$E1(m)$ is used to compute the value of m. If the value of m starts to saturate as m an increase from m_0, then m_{0+1} is the minimum embedded dimension m. For a chaotic time series, the graph m against $E1(m)$ will shows a saturation state. While for random time series, the graph m against $E1(m)$ will not show any saturation state.

As explained earlier, $E2(m)$ also introduced by (Cao, 1997). For a random time series, the value of $E2(m) = 1$ for all m value. Besides, there are few possibilities where $E2(m) \neq 1$. If there is at least one value $E2(m) \neq 1$, then the observed time series is chaos.

The overall analysis from the Cao method onto the observed time series are displayed in Figure 3. The saturation value is not specifically stated by (Cao, 1997) since saturation may vary with different time series. For O_3 time series, the value of m_0 observed is 2 where m starts to saturate from 0.95 to 1.00. Since there is saturation state in the observed graph, the observed time series is chaos. Hence, the value of $m_{0+1} = m = 3$. For $E2(m)$ there are few values where $E2(m) \neq 1$ which are at m = 1,2,3,4,5,8 and 10. According to Abarbanel (1996) [16], the observed time series is chaos.

It is confirmed that the observed time series is chaos through the phase space plot and Cao method. The presence of chaotic behaviour is recorded in Table 1.

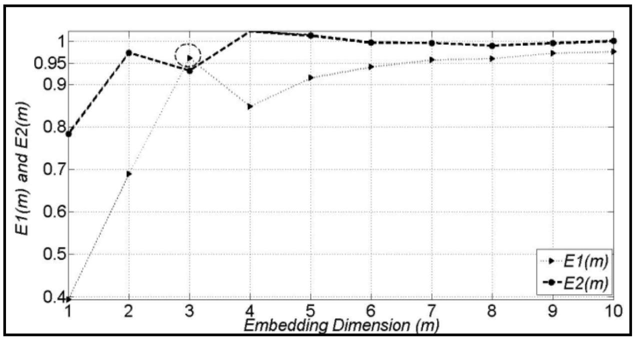

Figure 3. The Cao method

Table 1. Chaos determination

Method	**Chaos**
Phase space plot	Yes
$E1(m)$	Yes
$E2(m)$	Yes

III. PREDICTION

Three prediction models will be used in this study onto O_3 time series in Shah Alam. The prediction can be made after the chaotic behaviour is detected in the observed time series. Two parameters $\tau = 1$ and $m = 3$ has been chosen respectively to be used in prediction. The last parameter that needs to be determined is k. Previous studies on O_3 time series and another time series use trial-and-error methods to determine the value k. The improvement in this study is by varying the number of k use $1 \le k \le 200$. The combination of parameters τ , m and k , will be used for prediction and the performance measure will be recorded.

A. Local Mean Approximation Method (LMAM)

Prediction through chaotic approach by using LMAM is based on equation:

$$Y_{j+1}^{m} = f Y_{j}^{m} \tag{5}$$

Prediction for value Y_{j+1}^{m} is done based on the value of k for Y_{j}^{m} . The value of k nearest neighbor for $Y_{j'}^{m}$ are chosen based on the minimum value of Euclidean distance $\left\| Y_{j'}^{m} - Y_{j}^{m} \right\|$ with $j' < j$. Let us say $k = 1$ is being used, then the approximation for Y_{j+1}^{m} is $Y_{j'+1}^{m}$. Prediction is taken as the mean value of $Y_{j'}^{m}$:

$$Y_{j+1}^{m} = \frac{1}{k} \sum_{q=1}^{k} Y_{j_q'+1}^{m} \tag{6}$$

Figure 4 shows the prediction by using k that is ranged between $1 \le k \le 200$. The value of k that gives the maximum cc is recorded as well. The best k value is 17 that gives the maximum value of $cc = 0.8486$. The combination of the parameter used are $\tau = 1$, $m = 3$ and $k = 17$ give the maximum prediction performance. Table 3 shows the prediction result by using k the value through $k = 2m$ and trial-and-error method for LMAM.

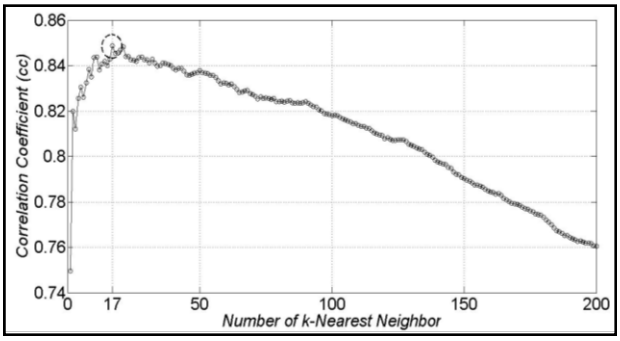

Figure 4. Prediction by using LMAM with the best k

Table 2. Prediction results by using k value from previous studies for LMAM

k	cc
50	0.8379
100	0.8181
200	0.7604
$k = 2m$	0.8261

B. Local Linear Approximation Method (LLAM)

For prediction by using LLAM, the linear equation that being use is:

$$x_{n+1} = Ax_n + B \tag{7}$$

Equation (7) is used to be standardised with the training set at (7). Prediction for x_{n+1} is calculated by referring to the value of x_n . Constant value A and B are calculated through the smallest square root method. Next, the phase space is constructed by using $\tau = 1$ and $m = 3$. The reconstruction of phase space is:

$$Y_{j}^{3} = (x_j, x_{j+1}, x_{j+2}, x_{j+3}, x_{j+4} x_{j+5}) \tag{8}$$

with $j = 1, 2, 3, 4, \ldots, N-5$. Since $N = 552$, the last phase space is:

$$Y_{547}^{3} = (x_{547}, x_{548}, x_{549}, x_{550}, x_{551}, x_{552}) \tag{9}$$

Linear equation $x_{n+1} = Ax_n + B$ is calculated. The value for A and B are changed accordingly depends on the value of k used. Let say k is 20 . It produces equation

$$x_{n+1} = 0.3421 x_n + 3.3923 \tag{10}$$

Figure 5 shows that at $k = 1$ gives the maximum $cc = 0.8453$. Even though there are several values of k that gives the same maximum value, the smallest value of k is chosen. A small value of k is enough to make an excellent prediction (Casdagli, 1992). Table 4 shows the prediction result by using k the value through $k = 2m$ and trial-and-error method for LLAM.

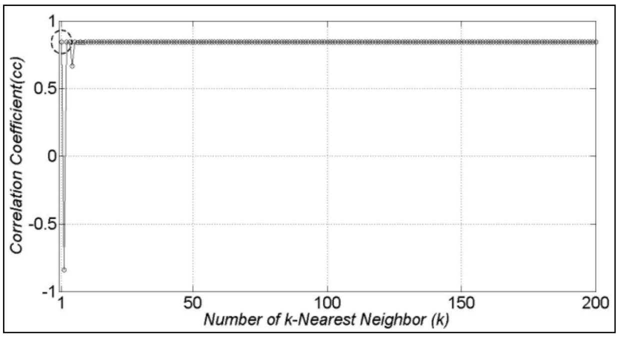

Figure 5. Prediction by using LLAM with the best k

Table 3. Prediction results by using k value from previous studies for LLAM

k	cc
50	0.8453
100	0.8453
200	0.8453
$k = 2m$	0.8453

C. Improved Local Linear Approximation Method (ILLAM)

This method is first used in predicting O_3 concentration time series, and the results are excellent compared to LLAM (Hamid & Noorani, 2013). Previously, this method used trial-and-error methods to determine the value for k, $k = 200$ (Hamid & Noorani, 2013). In this study, the number of k will be tested from 1 up to 200. ILLAM will use Equation (7). For LLAM, constant values A and B are calculated by using a training set. The improvement is made on the training set. The training set is updated for every prediction made, and a new equation is formed as follows:

$$Y^m_{jk^k} = C_n Y^m_{jk} + D_n \tag{11}$$

Prediction through LLAM only produced one equation for the whole 168 predictions. While for ILLAM, for every 168 predictions, 168 equations will be made. Figure 6 shows the best value of k where $k = 146$ that gives maximum $pk = 0.8606$. The value of k used from 1 to 75 cannot be calculated. The calculation process cannot be solved by using a simultaneous equation and finally produces *NaN* (Not a Number). Let's say $k = 2$, $x_{jk} = (5,5)$ and time series $x_{jk^1} = (5,9)$. By referring to Equation (7), equations as below are formed:

$$\begin{bmatrix} 5 \\ 9 \end{bmatrix} = A + B \begin{bmatrix} 5 \\ 5 \end{bmatrix} \tag{12}$$

Simultaneously, the equation cannot be solved. That is why the calculation cannot be solved and produces *NaN* for the first 75 k used.

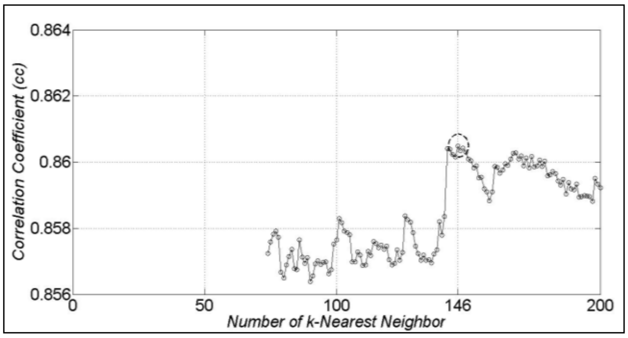

Figure 6. Prediction by using ILLAM with the best k

Table 4. Prediction results by using k value from previous studies for ILLAM

k	cc
50	*NaN*
100	0.8575
200	0.8594
$k = 2m$	*NaN*

IV. RESULTS

The time series observed is chaotic. Prediction is made after determining three parameters. The prediction performance gives an excellent result. This study is focused onto the selection of k value. The improved method in determining the value k is successfully done. The method of trial-and-error in determining the value of k are less correlated. Figure 4 until Figure 6 shows the maximum cc generated by using the best value of k . Table 3 until Table 4 shows the prediction using the trial-and-error method by LMAM, LLAM and ILLAM, respectively. These tables prove that the improved method in determining the value k gives a better prediction result. Figure 7 until Figure 9 shows the comparison graphs of the real and predicted data by using LMAM, LLAM and ILLAM.

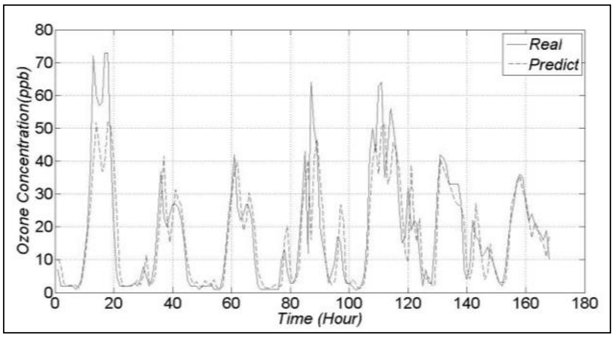

Figure 7. Prediction by using LMAM

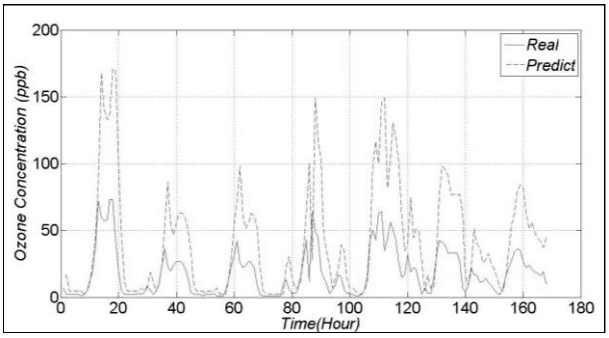

Figure 8. Prediction by using LLAM

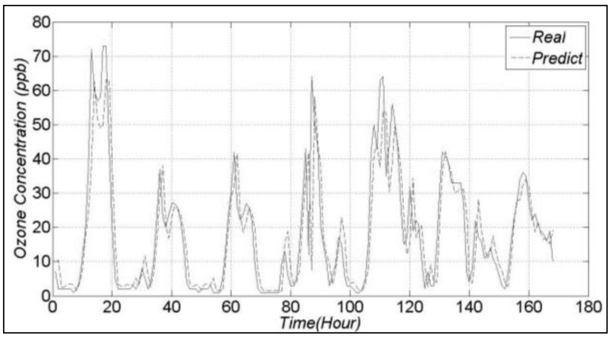

Figure 9. Prediction by using ILLAM

V. CONCLUSIONS

Phase space plot and Cao method had been used to detect the presence of chaotic behaviour. This study shows that the used time series is chaotic. All the past studies determine the value k by using the trial-and-error method. For this study, the value k varies from 1 to 200. Prediction by using the value of k by using trial-and-error method also had been recorded for comparison purposes. It is proven that the improved method in determining the value k is better than the trial-and-error method. Table 5 shows that for all three methods, a value of k that gives maximum cc for each method. Therefore, it is suggestable to use an improved method by varying the number of k to compute a better prediction performance. This study may help the authority such as Department of Environment Malaysia improve and control Malaysia's air quality.

Table 5. Prediction by using parameter k from the improved method

Method	k	cc
LMAM	17	0.8486
LLAM	1	0.8453
ILLAM	146	0.8606

VI. ACKNOWLEDGMENT

I want to express my utmost gratitude to the Department of Environment Malaysia for contributing the necessary data. The Ministry of Education Malaysia supports this research under the Fundamental Research Grant Scheme (FRGS) with the research code 2019-0005-102-02(FRGS/1/STG06/UPSI /02/1). Not to forget, I would like to thank the Ministry of Education for Fee Waiver Programme for encouraging participation among teachers for ICSTEM 2021.

VII. REFERENCES

Abarbanel, HDI 1996, 'Analysis of Observed Chaotic Data', Institute for Nonlinear Science, Springer-Verlag New York.

Adenan NH & Noorani, MSM 2013, 'River flow prediction using non-linear prediction method', International Journal of Mathematical and Computational Sciences, vol. 7, no. 11.

Adenan, NH & Noorani, MSM 2015, 'Peramalan Data Siri Masa Aliran Sungai di Dataran Banjir dengan Menggunakan Pendekatan Kalut', Sains Malaysiana, vol. 44, no. 3, pp. 463-471.

Ali, NM & Hamid, NZA 2019, 'Chaotic analysis for Malaysia west coast sea level: a case study of Kukup, Johor', IOP Conferences Series: Earth and Environmental Science, 289.

Bahari, M & Hamid NZA 2019, 'Analysis and prediction of temperature time series using chaotic approach', IOP Conferences Series: Earth and Environmental Science. 289.

Cao L 1997, 'Practical Method for Determining the Minimum Embedding Dimension of a Scalar Time Series', Physica D: Non-linear Phenomena, vol. 110, pp. 1-2.

Casdagli M 1991, 'Chaos and Deterministics Versus Stochastic Non-Linear Modelling', Santa Fe Institute, vol. 54, no. 2, pp. 303-328.

Cisneros, R, Bytnerowicz, A, Schweizer, D, Zhong, S, Traina, S & Bennet, DH 2010, 'Ozone, Nitric Acid and Ammonia Air Pollution is Unhealthy for People and Ecosystem in Southern Sierra Nevada, California', Enviromental Pollution, vol. 158, no, 10, pp. 3261-3271.

Department of Statistics, Malaysia 2010, Largest Cities and Municipalities in Malaysia, viewed 20 December 2019, <https://web.archive.org /web/20150205090002/http://www.statistics.gov.my/portal/download_Population/files/population/03ringkasan_kawasan_PBT_Jadual1.pdf>.

Environmental Protection Agency 2017, Ground-level Ozone Pollution, United States of America.

Fraser AM & Swiney HL 1986, 'Independent Coordinates for Strange Attractors from Mutual Information', Phys Rev A., vol. 33, no. 2, pp. 1134-1140.

Hamid, NZA & Noorani, MSM 2017, 'Aplikasi Model Baharu Penambahbaikan Pendekatan Kalut ke atas Peramalan Siri Masa Kepekatan Ozon', Sains Malaysiana, vol. 46, no. 8, pp. 1333-1339.

Hamid NZA & Noorani, MSM 2013, 'An improved prediction model of ozone concentration time series based on chaotic approach', International Journal of Mathematical and Computational Sciences, vol. 7, no. 11.

Hamid, NZA 2018, 'Application of Chaotic Approach in Forecasting Highland's Temperature Time Series', IOP Conference Series: Earth and Environmental Science, vol. 169, pp. 1-9.

Hamid, NZA & Noorani, MSM 2014, Suatu kajian perintis menggunakan pendekatan kalut bagi pengesanan sifat dan peramalan siri masa kepekatan PM10, Sains Malaysiana.

Hamid, NZA, Noorani, MSM, Juneng, L & Latif, MT 2013, 'Prediction of Ozone Concentration Using Nonlinear Prediction Method', Proceedings of the 20th National Symposium on Mathematical Sciences, vol. 1522, pp. 125-131.

Hamid, NZA, Noorani, MSM & Adenan, NH 2017, 'Chaotic Analysis and Short-Term Prediction of Ozone Pollution in Malaysian Urban Area', Journal of Physics, 890.

Idris NA & Mahmud M 2017, 'Kajian Jejak Karbon di Kuala Lumpur', Journal of Social Sciences and Humanities, vol. 12, no. 2, pp. 165-182.

Jayawardena, AW 1997, 'Runoff Forecasting Using a Local Approximation Method', IAHS, vol. 239, pp. 167-171.

Lakshmi, S & Tiwari, RK 2009, 'Model Dissection from Earthquake Time Series: A Comparative Analysis Using Modern Non-Linear Forecasting and Artificial Neural Network Approaches', Computers & Geosciences, vol. 35, no. 2, pp. 191–204.

Mashuri, A, Adenan, NH & Hamid, NZA 2019, 'Determining the Chaotic Dynamics of Hydrological Data in Flood-Prone Area', Civil Engineering and Architecture, vol. 7(6A), pp. 71-76.

McCarthy, N 2019, Study: Pollution Kills 8.3 Million People, Anually. URL: https://www.statista.com/chart/20360/premature-pollution-related-deaths/.

Myllyvirta, L 2020, 'Quantifying the Economic Cost of Air Pollution from Fossil Fuels', Centre of Research on Energy and Clean Air.

NASA 2018, Ozone Facts, URL https://ozonewatch.gsfc.nasa.gov/facts/SH.html.

Nuvolone, D, Petri, D & Voller, F 2017, 'The effects of ozone on human health', Environmental Science and Pollution Research, vol. 25, no. 9, pp. 8074–8088.

Regonda, S, Rajagopalan, B, Lall, U, Clark, M & Moon, Y-I 2005, 'Local Polynomial Method for Ensemble Forecast of Time Series', Non-linear Processes in Geophysics, vol. 12, no. 3, pp. 397–406.

Ruslan, AB & Hamid, NZA 2019, 'Application of improved chaotic method in determining number of k-nearest neighbor for CO data series', International Journal of Engineering and Advanced Technology, vol. 8(6S3).

Ruslan, AB, Hamid, NZA & Jusoh, KC 2020a, 'Peramalan aplikasi pendekatan kalut bahan pencemar siri masa CO menggunakan kaedah penambahbaikan dalam penentuan parameter bilangan *k*-jiran terdekat', Borneo International Journal, vol. 2, no. 4, pp. 11-16.

Ruslan, AB, Hamid, NZA & Jusoh, KC 2020b, 'Kajian rintis peramalan siri masa SO2 melalui pendekatan kalut di kawasan bandar', Journal of Quality Measurement and Analysis, vol. 16, no. 1, pp. 17-27.

Ruslan, AB, Hamid, NZA & Jusoh, KC 2020c, 'Peramalan tak linear bagi siri masa karbon monoksida di kawasan dengan taburan penduduk tinggi di Sabah', Journal of Quality Measurement and Analysis, vol. 16, no. 2, pp. 157-170.

Schuster, HG 1988, Deterministic Chaos, VCH, Weinheim.

Sivakumar, B 2002, 'A phase-space reconstruction reconstruction approach to prediction of suspended sediment concentration in rivers', Journal of Hydrology, vol. 258, pp. 149-162.

Sprott, C 2003, Chaos and Time Series Analysis, Oxford University Press.

Velickov, S 2004, Nonlinear Dynamics and Chaos, London; Taylor and Francis Group plc.

Zaim, WNAWM & Hamid, NZA 2017, 'Peramalan Bahan Pencemar Ozon (O3) di Universiti Pendidikan Sultan Idris, Tanjung Malim, Perak, Malaysia Mengikut Monsoon dengan Menggunakan Pendekatan Kalut', Sains Malaysiana, vol. 46, no. 12, pp. 2523-2528.