International Journal of Engineering & Technology, 7 (4.38) (2018) 1635-1638

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Prediction of Ozone Pollution through Chaotic Approach

Nor Zila Abd Hamid*

Faculty of Science and Mathematics, Sultan Idris Education University,
35900, Tanjong Malim, Perak, Malaysia

Abstract

This study is about applying chaotic approach in analyzing and predicting ozone (O3) pollution. There are two stages in chaotic approach. The first is to detect the presence of chaotic nature by means of phase space plot and Cao method while the second stage is the prediction through the local mean approximation method. Result found that O3 pollution recorded at one of Malaysian leading educational university namely Sultan Idris Education University is detected as chaotic in nature. Thus, the prediction is continued using chaotic approach. The one-hour ahead prediction is carried out for one month period. The correlation coefficient value between the observed and predicted O3 pollution is 0.9165. This excellent result indicates that the local mean approximation method is suitable for predicting O3 pollution and chaotic approach is an appropriate approach which can be applied in analyzing and predicting O3 pollution. These results are expected to assist stakeholders such as the Department of Environment Malaysia, Malaysian Meteorological Department as well as educational institutions for more systematic pollution management.

Keywords*: Chaotic approach, local mean approximation method, prediction, ozone pollution.*

1. Introduction

Ozone (O_3) is an air pollutant that endangers health and has caused a variety of respiratory and cardiovascular diseases[1,2]. Thus, more research should be done towards the pollution. The main contribution of this study is the development of O_3 prediction model through chaotic approach.

In mathematics, the nature of a data can be classified into deterministic or random. Deterministic data is predictable while random data cannot be predicted. Chaotic is a nature, which is in between deterministic and random[3]. Chaotic data is predictable, however, by being sensitive dependence to initial conditions, then, for the chaotic data, only short-term prediction is allowed[4]. Research by [5–8] found that O_3 data is chaotic through various method such as correlation dimension, Lyapunov exponent and correlation integral method. From literature, it is found that the phase space plot as well as Cao method[9] are also can be used to distinguish the nature of the data. Both methods are easy to apply; however, both methods are rarely been used on O_3 pollution data. Therefore, in this study, both methods will be applied.

Apart from the detection of data nature, the chaotic approach is also used to predict the O_3. In this study, the pollution will be predicted through one of the basic method from chaotic approach namely the local mean approximation method.

In the current research in Malaysia, O_3 pollution is predicted using neural networks and multiple linear regression method (see [10–12]). Both are known as multivariate method since prediction of O_3 pollution through both methods are depending on factors affecting O_3 pollution such as methane, carbon monoxide (CO), nitrogen oxide (NOx), humidity, temperature, solar radiation and wind speed. In the event of information of the factors are inadequate, other methods need to be considered. Therefore, in this study, the chaotic approach is used. Prediction of O_3 pollution through chaotic approach is involving O_3 pollution data solely.

Respectively, [8] and [13] applied chaotic approach to predict hourly and daily average of O_3 pollution. Both researches gained excellent results. Undoubtedly, in this paper, chaotic approach also will be applied. There are a few methods under chaotic approach. The simplest yet always use is the local mean approximation method. Therefore, the method is chosen.

The first contribution of this study is to determine the nature of the O_3 pollution using two selected methods namely the phase space plot and Cao method. The second contribution is the prediction of the O_3 pollution using local mean approximation method. The performance of this method is determined by calculating the correlation coefficient (r) value.

2. Ozone Pollution Data

The O_3 pollution data used in this study was observed at one of the semi-urban area located in the state of Perak, Malaysia namely Tanjong Malim. The total area of Tanjong Malim is about 950 km^2. Tanjong Malim is known as educational area since one of the leading educational university namely Sultan Idris Education University is located there. Tanjong Malim is frequently visited by peoples. Hence, the prediction of air pollution such as O_3 is very necessary to maintain public health.

The O_3 data used in this study are the secondary data obtained from the Department of Environment Malaysia. The data are recorded hourly in ppb (part per billion) unit and observed for three months, from 1st June until 31st August 2014. Thus, the total number of the data is 2208. The overall hourly O_3 data observed in Tanjong Malim is as shown in Figure 1 and the statistical description of the data is as listed in Table 1.

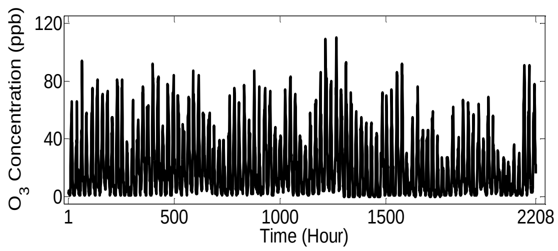

Fig 1: O_3 data observed in Tanjong Malim

Table 1: Statistical description of observed O_3 data

Statistics	Value (ppb)
Mean	21.7
Standard Deviation	22.2
Sample Variance	494.9
Kurtosis	0.4
Skewness	1.1
Minimum	0.0
Maximum	110.0
Sum	47855.0

3. Chaotic approach

3.1. Reconstruction of Phase Space

This section elaborates the method on analyzing data namely the reconstruction of phase space. With x_t is the O_3 data at t-th hour and N is the total hours of observation, O_3 data is recorded in the form of:

$$X = (x_1, x_2, \dots, x_N) \quad (1).$$

X is divided into two parts; X_{train} and X_{test}. The unknown parameters are calculated using X_{train} while the performance of the prediction model are determined using X_{test} .
O_3 pollution in Tanjong Malim is recorded for three months. Data on June and July were used as X_{train} while data on August were used as X_{test}. Therefore, there are 1464 X_{train} and 744 X_{test}. Data which recorded as Equation 1 were reconstructed into vector of Equation 2 namely the m-dimensional phase space:

$$\boldsymbol{Y}_j^m = \{x_j, x_{j+\tau}, x_{j+2\tau}, \dots, x_{j+(m-1)\tau}\} \quad (2).$$

Before any action to be taken, the delay time, τ and embedding dimension, m must be calculated first. In order to calculate τ, [14] have tested various method and detected that the average mutual information method by [15] is good in calculating τ. Therefore, average mutual information method is applied in this study. Through Equation 3, τ is calculated.

$$I(T) = \frac{1}{N}\sum_{a=1}^{N} p(u_a, u_{a+T})\, log_2\left[\frac{p(u_a,u_{a+T})}{p(u_a)p(u_{a+T})}\right] \quad (3)$$

$p(u_a)$ and $p(u_{a+T})$ are the probability of obtaining u_a and u_{a+T} in data X. $p(u_a, u_{a+T})$is a joint probability of $p(u_a)$ and $p(u_{a+T})$.
From Equation 3, graph $\{T, I(T)\}$ is plotted. The first minimum value of T is considered as τ. Once the value of τ is determined, the phase space is plotted. Phase space is the graph of $\{x_t, x_{t+\tau}\}$. The nature of the data is chaotic if there exist a well-defined attractor on the phase space [16].
In order to calculate m , Cao method [9] was applied. It has been scientifically proven by [9] that Cao method neither contain any subjective parameters nor depend on the number of recorded data. Undoubtedly, Cao method is applied to calculate m.

m is defined through $E1(m) = \frac{E(m+1)}{E(m)}$. Symbol $\|\blacksquare\|$ is the maximum norm and $\boldsymbol{Y}_n^m$ is the nearest neighbor to $\boldsymbol{Y}_j^m$;

$$E(m) = \frac{1}{N-m\tau}\sum_{\boldsymbol{j=1}}^{\boldsymbol{m}} \frac{\left\|\boldsymbol{Y}_j^{m+1} - \boldsymbol{Y}_n^{m+1}\right\|}{\left\|\boldsymbol{Y}_j^m - \boldsymbol{Y}_n^m\right\|} \quad (4).$$

If $E1(m)$ saturates at m_0, then $m_0 + 1 = m$. If $E1(m)$ does not saturates, then the data is random. Thus, besides calculating $m, E1(m)$ also can be used to differentiate between random and chaotic data.
In addition, Cao[9] also introduced the parameter of $E2(m)$. $E2(m)$ is calculated by:

$$E2(m) = \frac{E^*(m+1)}{E^*(m)} \quad (5)$$

where

$$E^*(m) = \frac{1}{U-m\tau}\sum_{j=1}^{U-m\tau}\left|x_{j+m\tau}^m - x_{w+m\tau}^m\right| \quad (6).$$

According to Cao method, if there exist $2(m) \neq 1$, its indicates that chaotic nature is presence in the recorded data. Otherwise, the data is random.

3.2. Prediction Model

Prediction of $\boldsymbol{Y}_{j+1}^m$is done based on

$$\boldsymbol{Y}_{j+1}^m = f(\boldsymbol{Y}_j^m) \quad (7).$$

k nearest neighbors $\boldsymbol{Y}_{j'}^m$ are selected based on the minimum value of $\|\boldsymbol{Y}_{j'}^m - \boldsymbol{Y}_j^m\|$, where $j' < j$ and $\|*\|$ is the Euclidean distance. In this study, $k = 50$ determined through trial and error method is used.The predicted $\boldsymbol{Y}_{j+1}^m$ is taken as the average of the $\boldsymbol{Y}_{j'+1}^m$values;

$$\boldsymbol{Y}_{j+1}^m = \frac{1}{k}\sum_{q=1}^{k}\boldsymbol{Y}_{j_q'+1}^m \quad (8).$$

3.3. Performance Measure

The correlation coefficient (r) was used to measure the method's performance. The r value is between 0 and 1. The closer the value of r to 1, the better the predition is. The formula of r is:

$$r = \left|\frac{\sum_{i=1}^{l}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{l}(P_i - \bar{P})^2}\sqrt{\sum_{i=1}^{l}(O_i - \bar{O})^2}}\right| \quad (9)$$

with P_i and O_i are the predicted and observed value at i −th hour and $\bar{P}$ and $\bar{O}$ are the average of predicted and observed values. Parameter l is the total value of X_{test} and in this study $l = 744$. All of the computational parts are carried out using the Matlab software.

4. Results and discussion

4.1. Chaotic Nature

Finding from average mutual information method is as shown in Figure 2. The best time interval to reflect the attractor structure of the O_3 pollution data is 8 since form Figure 2, $\tau = 8$. Thus, the phase space is plotted as shown in Figure 3 on the plane of $\{x_t, x_{t+8}\}$. The phase space displays that all points are located in one triangle trajectory. The existence of the trajectory shows that the observed O_3 pollution is chaotic in nature.
Meanwhile, referring to the Cao plot of Figure 4, after the value of five, $(m_0 = 5)$, $E1(m)$ starts to saturate. Thus, the embedding dimension is $m = 6$. Furthermore, according to Cao[9], the saturation of $E1(m)$ shows that observed O_3 data is chaotic. In addition, the existence of $E2(m) \neq 1$ shows the presence of chaotic nature in

the observed pollution. As in Figure 4, there exist $E2(m) \neq 1$ at $m = 1$, $m = 2$ and $m = 5$. Therefore, from $E2(m)$, it also can be concluded that the nature of the O_3 pollution at Tanjong Malim is chaotic.

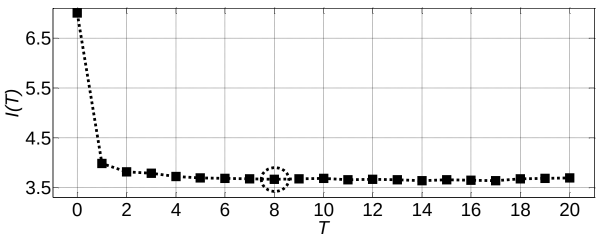

Fig. 2 : Average Mutual Information method

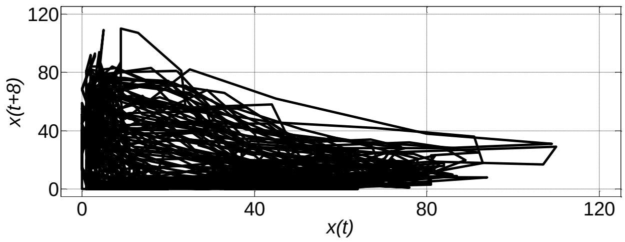

Fig. 3 : Phase space plot

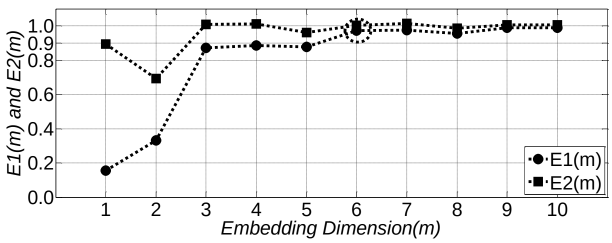

Fig. 4 : Cao method

4.2. Factors Influence Ozone Pollution

Researches by [17–19] found that the O_3 pollution has been influenced by the direction and speed of the wind, sea breeze, temperature, relative humidity, suspended dust, radiation, sunlight and solar energy. Research on sensitivity analysis of O_3 pollution driven by [20] found that O_3 was influenced by pollution of NO and PM_{10}, maximum and average temperature, maximum and average O_3 pollution, solar radiation, relative humidity, wind direction as well as sunlight.

$m = 6$ calculated from Cao method shows that there are at least six factors affecting O_3 pollution in the Tanjong Malim educational area. The factors listed in above paragraphs are over six. Therefore, $m = 6$ are consistent and truth.

4.3. Prediction Result

From both phase space and Cao method, it can be concluded that the observed hourly O_3 pollution data is chaotic. Hence, the prediction will be done using chaotic approach.

Prediction is done for accounting periods from 1st to 31st August 2014. Figure 5 shows the prediction result. From observation, the trend of up and down of the data can be predicted well.

The correlation coefficient (r) value is 0.9165. The r value is closer to 1. Therefore, the prediction through the local mean approximation method is success.

It can be seen that the too high data are unpredictable at best. In future, the local mean approximation method needs to be improved to overcome this problem.

Previous studies by [21–25] have shown that the chaotic approach is able to predict high O_3 pollution observed at the various areas. Thus, the power of chaotic approach in predicting high values is undeniable. In addition, studies have also been carried out on the PM_{10} pollutants by [26] and temperature time series by [27]. Hence, it is not beyond doubt that the local approximation method which is based on chaotic approach is capable of predicting the data of any air pollutants.

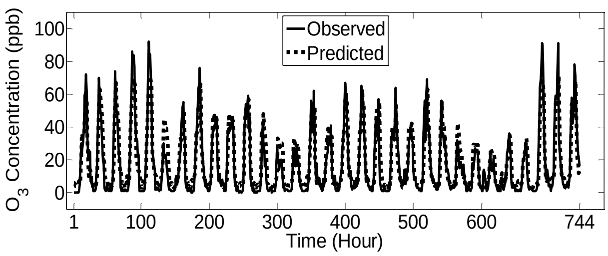

Fig. 5 : Prediction result

5. Conclusion and future research

The results from phase space plot and Cao method show the presence of chaotic nature in O_3 pollution at educational area of Tanjong Malim. The excellent prediction results with r approaching 1 indicates that chaotic approach is very suitable to be applied on O_3 pollution data.

In the future, the method such as autocorrelation function is proposed to calculate the τ value. Furthermore, chaotic approach can be tested on wind speed, relative humidity as well as pollution of carbon monoxide and sulphur dioxide. As urban areas are having uncertainties climate due to heavy pollution, then, the research area can be extended to Malaysian urban area such as Klang Valley, Johor Baharu, Johore and Georgetown, Penang.

Acknowledgment

Million appreciation goes to Department of Environment, Malaysia for providing the O_3 pollution data. This study is under the sponsorship of the research grant with code 2019-0005-102-02 (FRGS/1/2018/STG06/UPSI/02/1).

References

[1] S. Cakmak, C. Hebbern, J. Vanos, D. L. Crouse, and R. Burnett, Environ. Pollut. 214, 589–599 (2016).
[2] L. Madaniyazi, T. Nagashima, Y. Guo, X. Pan, and S. Tong, Environ. Int. 92–93, 165–172 (2016).
[3] H. D. I. Abarbanel, Analysis of Observed Chaotic Data. (Springer-Verlag, Inc., 1996).
[4] J. C. Sprott, Chaos and Time-Series Analysis. (Oxford University Press, 2003).
[5] G. Bandyopadhyay, and S. Chattopadhyay, Soft. Comp. (2008)
[6] A. B. Chelani, Atmos. Environ. 44, 4318–4324 (2010).
[7] V. Cuculeanu, C. Rada, and A. Lupu, Geophysique 52–53, 77–85 (2009).
[8] J. Chen, S. Islam, and P. Biswas, Atmos. Environ. 32, 1839–1848 (1998).
[9] L. Cao, Phys. D Nonlinear Phenom. 110, 43–50 (1997).
[10] N. A. Ghazali, N. A. Ramli, A. S. Yahaya, N. F. F. M. Yusof, N. Sansudin, and W. A. A. Madhoun. Environ. Monit Assess 165, 475–489 (2010).
[11] M. Muhamad, A. Z. Ul-saufie, and S. M. Deni, J. Environ. Sci. Technol. 8, 102–112 (2015).
[12] K. C. Tan, H. S. Lim, and M. Z. M. Jafri, Atmos. Pollut. Res. 1–14 (2016).
[13] K. Kocak, L. Saylan, and O. Sen, Atmos. Environ. 34, 1267–1271 (2000).

[14] M. Siek, M. Predicting Storm Surges. Chaos, Computational Intelligence, Data Assimilation, Ensembles. (Delft University of Technology, Netherlands, 2011).
[15] A. M. Fraser, and H. L. Swinney, Phys. Rev. A 33, 1134–1140 (1986).
[16] B. Sivakumar, J. Hydrol. 258, 149–162 (2002).
[17] M. T. Latif, L. Shun, and L. Juneng, Atmos. Environ. 61, 434–445 (2012).
[18] B. Ozbay, G. A. Keskin, S. C. Dogruparmak, and S. Ayberk, Atmos. Res. 102, 57–65 (2011).
[19] Y. Y. Toh, S. F. Lim, and V. G. Roland, Atmos. Environ. 70, 435–446 (2013).
[20] F. Inal, Clean-Soil, Air, Water 38, 897–908 (2010).
[21] N. Z. A. Hamid, M. S. M. Noorani, L. Juneng, and M. T. Latif, AIP Proc. 125, 125–131 (2013).
[22] N. Z. A. Hamid, and M. S. M. Noorani, Int. J. Math. Sci. Eng. 7, 661–666 (2013).
[23] N. Z. A. Hamid, M. S. M. Noorani, and N. H. Adenan, J. Phys. Conf. Ser. 890, 1–5 (2017).
[24] N. Z. A. Hamid and M. S. M. Noorani, Sains Malaysiana 46, 1333–1339 (2017).
[25] N. Z. A. Hamid, N. H. Adenan, N. H. and M. S. M. Noorani, AIP Proc. 1870, 1–8 (2017).
[26] N. Z. A. Hamid, and M. S. M. Noorani, Sains Malaysiana 43, 475–481 (2014)
[27] N. Z. A. Hamid, Int. J. of Eng. & Technol. 7(3.7), 590-593 (2018).