

PAPER • OPEN ACCESS

## Chaotic analysis and short-term prediction of ozone pollution in Malaysian urban area

To cite this article: Nor Zila Abd Hamid *et al* 2017 *J. Phys.: Conf. Ser.* **890** 012157

View the [article online](#) for updates and enhancements.

### You may also like

- [Fast range-corrected proton dose approximation method using prior dose distribution](#)  
Peter C Park, Joey Cheung, X Ronald Zhu et al.
- [Meteorological conditions contributed to changes in dominant patterns of summer ozone pollution in Eastern China](#)  
Zhicong Yin and Xiaoqing Ma
- [The current situation of surface layer ozone pollution in china's typical cities](#)  
Mengyao Lv, Chao Xie, Yuan You et al.



Free the Science Week 2023 April 2–9

Accelerating discovery through  
**open access!**

 [www.ecsdl.org](http://www.ecsdl.org) [Discover more!](#)

The banner features a dark background with a glowing blue digital interface. A hand is shown interacting with a central circular element that contains a white padlock icon. The interface is composed of various geometric shapes, lines, and data-like patterns, creating a futuristic and technological atmosphere.

# Chaotic analysis and short-term prediction of ozone pollution in Malaysian urban area

Nor Zila Abd Hamid<sup>1</sup>, Mohd Salmi Md Noorani<sup>2</sup> and Nur Hamiza Adenan<sup>1</sup>

<sup>1</sup>Department of Mathematics, Faculty of Science and Mathematics,  
Universiti Pendidikan Sultan Idris, 35900, Tanjong Malim, Perak, Malaysia

<sup>2</sup>School of Mathematical Sciences, Faculty of Science and Technology,  
Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia

E-mail:nor.zila@fsmt.upsi.edu.my

**Abstract.** This study focuses on the analysis and prediction of hourly ozone ( $O_3$ ) pollution in one of Malaysian urban area namely Shah Alam through chaotic approach. This approach begins by detecting the chaotic behavior of the  $O_3$  pollution using phase space plot and Cao method. Then, the local mean approximation method is used for prediction purposes. The  $O_3$  pollution observed at Shah Alam is detected as chaotic in behavior. Due to the chaotic behavior, only short-term prediction is allowed. Thus, the one-hour ahead prediction is done through the local mean approximation method. The prediction result shows that correlation coefficient value between the observed and predicted time series is near to one. This excellent prediction result shows in particular that the local mean approximation method can be used to predict the  $O_3$  pollution in urban area. In general, chaotic approach is a useful approach that can be used to analyze and predict the  $O_3$  pollution time series.

## 1. Introduction

Ozone ( $O_3$ ) is an air pollutant that endangers health [1,2]. Therefore, the development of prediction models of  $O_3$  pollution is important. Mathematically, the behaviors of the time series can be classified into deterministic and random. Deterministic time series is predictable while random time series is not predictable. Chaotic behavior is in between the deterministic and random behaviors [3]. Chaotic time series is predictable; however, due to the sensitive dependence on initial conditions, then, for a chaotic time series, only short-term prediction allowed [4]. There are various approaches that have been used by previous studies to test whether a time series of  $O_3$  is chaotic or not. By using the correlation dimension method, integral correlation, entropy and Lyapunov exponent, studies by [5–7] found that their observed  $O_3$  time series are chaotic. Recently, by using the exponents Lyapunov, Hausdorff dimensional method and phase space plot, [8] also proved that the behavior of  $O_3$  time series is chaotic. Phase space plot and Cao method [9] are able to classify the behaviors of the time series. However, these methods are rarely used on  $O_3$  time series although both have been proven effective over time series such as suspended sediment concentration, traffic flow and earthquakes (see [10–12]). Therefore, this study will apply both methods towards the chosen  $O_3$  pollution time series.

In the current research in Malaysia,  $O_3$  time series are predicted using neural networks and multiple linear regression method (see [13–15]). Prediction model through both methods are dependent on



meteorological factors such as water temperature, humidity, solar radiation and wind speed as well as gaseous factors such as the precursor gases of methane, carbon monoxide (CO) and nitrogen oxide (NO<sub>x</sub>). However, if the information of those factors is not sufficient, an alternative method is needed to run the prediction. Therefore, in this study, local approximation method, a method based on chaotic approach is used. This method has its own advantages as prediction process is done simply by using data from O<sub>3</sub> time series only, without involving data from other factors. The local approximation method has been used by [16] and [17] to predict O<sub>3</sub> time series in metropolitan areas. Both studies yielded very satisfactory results. Therefore, in this study, prediction models of O<sub>3</sub> time series is also carried out using the local approximation method. There is various sub method of local approximation method. The simplest and commonly used is the local mean approximation method. Thus, this method will be applied in this study.

The contributions of this study are: i) to apply the phase space plot and Cao method for detecting the presence of chaotic behaviors of O<sub>3</sub> time series and ii) to adapt the local mean approximation method to predict the observed time series. The pilot study towards O<sub>3</sub> times series in Malaysian background station has been conducted successfully by [18]. In this study, the same approach will be applied. However, this present study is focusing on predicting the O<sub>3</sub> pollution time series in Malaysian urban area.

## 2. Time Series Data

The observed secondary hourly data from 1<sup>st</sup> January to 31<sup>st</sup> March 2014 are obtained from the Department of Environment Malaysia. The O<sub>3</sub> pollution time series are recorded in ppb (part per billion) unit. The location where the O<sub>3</sub> time series was recorded is one of the government's primary school named SK TTDI Jaya, located in Shah Alam area. Shah Alam is an urbanized area located in Selangor state of Malaysia. Shah Alam area is approximately 55.2 km<sup>2</sup>. Shah Alam city filled with residential and commercial center. The main plant of the Malaysian car manufacturer Proton located at the industrial suburb of Shah Alam. The city began to grow after the Proton factory set up, which marked the beginning of the city as an industrial city. Network development factory and major highways along with commuter trains, buses and taxis contributed a lot to the quality of the air pollution in the area. Therefore, it is important to know the air quality in this area in the future.

## 3. Chaotic Approach

The O<sub>3</sub> time series from is recorded in the form of

$$X = (x_1, x_2, x_3, \dots, x_{N-1}, x_N) \quad (1)$$

where  $x_t$  is the O<sub>3</sub> time series at  $t$ -th hour and  $N$  is the total hours of observation.  $X$  was divided into two parts;  $X_{train}$  and  $X_{test}$ .  $X_{train}$  was used as a training data series to find the unknown parameters, while  $X_{test}$  was used to test the performance of the prediction model. This present study chose hourly O<sub>3</sub> data observed for 13 weeks, from 1<sup>st</sup> January 2014 until 17<sup>th</sup> March 2014 as  $X_{train}$  and the last two weeks was used as  $X_{test}$ . Thus, the total number of used data is 2160.

The time series in equation (1) will be reconstructed into the  $m$ -dimensional phase space of:

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

From equation (2), the value of delay time,  $\tau$  and embedding dimension,  $m$  must be determined. As the data are observed hourly (every hour), therefore the chosen  $\tau$  value is one. In calculating the value of  $m$ , it has been proved by [9] that Cao method has many advantages (i.e does not contain any subjective parameters and does not depend on the number of observed data). One may refer to [9] for further details on how to compute  $m$  using Cao method.

## 4. Prediction Models

Prediction process through the chaotic approach is interpreted through equation:

$$\mathbf{Y}_{j+1}^m = f(\mathbf{Y}_j^m) \quad (3).$$

Prediction of  $\mathbf{Y}_{j+1}^m$  is done based on the neighbors of  $\mathbf{Y}_j^m$ . The neighbors of  $\mathbf{Y}_j^m$  are labelled as  $\mathbf{Y}_{j'}^m$  where  $1 < j' < j$ . For the local mean approximation method, not all neighbors are used to predict  $\mathbf{Y}_{j+1}^m$ . With parameter  $k$  is the number of nearest neighbor, only  $k$  nearest neighbors are used.  $k$  nearest neighbors are neighbors with the minimum value of Euclidean distance  $\|\mathbf{Y}_{j'}^m - \mathbf{Y}_j^m\|$ . Parameter  $k$  is determined by trial and error method and in this study,  $k = 50$  is used. After  $\mathbf{Y}_j^m$  is determined, the one hour ahead  $\mathbf{Y}_{j+1}^m$  will be listed. The prediction of  $\mathbf{Y}_{j+1}^m$  is taken as the average of  $\mathbf{Y}_{j'+1}^m$  values:

$$\mathbf{Y}_{j+1}^m = \frac{\sum_{q=1}^k \mathbf{Y}_{j'_q+1}^m}{k} \quad (4).$$

The performance of the prediction models is calculated through the value of correlation coefficient  $r$ . The  $r$  was used to summarize the information about the relationship between the observed and predicted data. The  $r$  value is between -1 and +1. A value of 0 indicates that there is no relationship between the observed and predicted time series. Conversely, the closer  $r$  to -1 or +1, the stronger the relationship and explains that the real and predicted data series are close to each other.

## 5. Results and Discussion

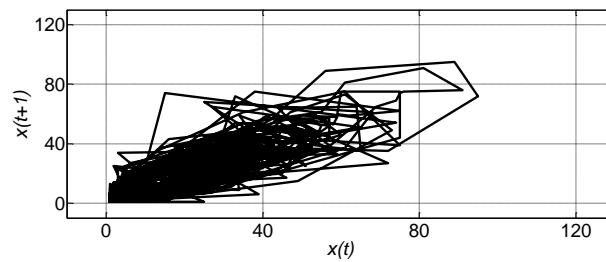
### 5.1. Chaotic Analysis

By using  $\tau = 1$ , the phase space plot was graphed. Figure 1 illustrates the phase space of  $\{x_t, x_{t+1}\}$ . The phase space plot represents the behaviors of the data. Through the reconstruction of the phase space, the movement of data from the initial position can be observed and consequently, the overall evolution of the data can be viewed through these phase space trajectories. As can be clearly seen in Figure 1, there exists a well-defined of attractor region in the middle of the plot where all points converge towards it. Referring to the study by [10], if there exists any well-defined region, the chaotic behaviors is present. On the other hand, if all points fill up the entire plot with no apparent structure of the attractor region, then the observed data are said to be random. Through Figure 1, it can be said that the chaotic behaviors is present in the observed  $O_3$  data series.

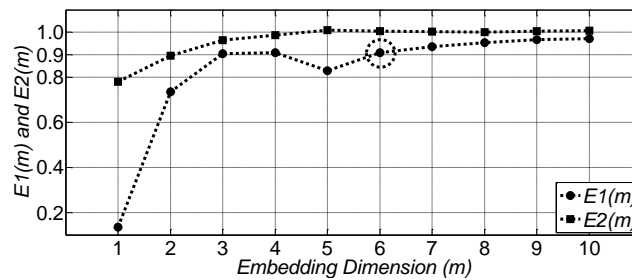
Figure 2 displays the results of  $E1(m)$  and  $E2(m)$  of the Cao method.  $E1(m)$  stops changing when the value of  $m$  is greater than the value of  $m_0 = 5$ . Therefore,  $m = m_0 + 1 = 6$  is the minimum embedding dimension. Furthermore, according to Cao method, if  $E1(m)$  saturates with increasing  $m$ , then chaotic behaviors is present in the data series. It can be observed that at  $m = 6$ , the value of  $E1(m)$  started to saturate.  $E1(m)$  continues to saturate with increasing  $m$ . Thus, the chaotic behavior is present. Moreover, the existence of  $E2(m) \neq 1$  at  $m = 1$  until  $m = 3$  assures the presence of the chaotic behaviors in the observed  $O_3$  pollution time series.

Recent investigation by [19–21] show that the  $O_3$  pollution in urban area has strong relationship with meteorological factors such as water temperature, sea breezes, suspended dust, relative humidity, solar

radiation, solar energy, wind direction and wind speed. The sensitivity analysis towards O<sub>3</sub> time series conducted by [22] found that the maximum and average temperature, solar radiation, maximum and average O<sub>3</sub> pollution, relative humidity, NO pollution, wind direction, PM<sub>10</sub> pollution as well as sunshine time influenced the O<sub>3</sub> time series. As  $m = 6$  is obtained from the Cao method, this suggests that at least six factors influence the observed O<sub>3</sub> time series in the urban area of Shah Alam. The list of factors detected and listed by recent studies above suggests that more than six factors influence the O<sub>3</sub> time series. Therefore, the present study finding is compatible and consistent. Thus, the result of  $m = 6$  obtained from the Cao method is consistent and reliable.

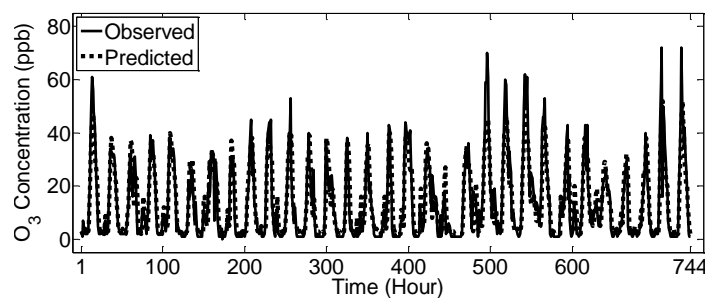


**Figure 1.** Phase space plot.



**Figure 2.** Result from Cao method.

### 5.2. Short-term prediction



**Figure 3.** Prediction results from the local mean approximation method.

Due to the sensitive dependence on initial conditions, then, for a chaotic time series, only short-term prediction allowed [4] for chaotic time series. In this present study, the one-hour ahead prediction is done for two weeks (336 hours). With  $\tau = 1$  and  $m = 6$ , the phase space of equation (2) is reconstructed and the prediction models are developed. Performance of the model is reflected in the calculation of correlation coefficient ( $r$ ). The prediction result of O<sub>3</sub> pollution model through the local mean approximation method is shown in Figure 4. The value of the performance indices is  $r = 0.9021$ . Thus, the model gains good results in which the  $r$ 's is above 0.9 and approaches 1.  $r$  value exceeding 0.9 indicates that there is a high correlation between predicted and observed values. Furthermore, from

Figure 3, it can be seen that the data trend (up and down) can be predicted well. Thus, the results demonstrate that the local mean approximation method is good in predicting the observed O<sub>3</sub> time series at urban area.

## 6. Conclusion

In this study, the chaotic behavior of hourly O<sub>3</sub> time series at an urban area of Shah Alam is detected through the phase space plot and Cao method. A short-term prediction model was built through the local mean approximation method, a method based on chaotic approach. The results demonstrate that the local mean approximation method is good in predicting the observed O<sub>3</sub> pollution at urban area with  $r$  value near to one. [23] has successfully predicted the PM<sub>10</sub> time series using the local approximation method. In future, chaotic approach is suggested to be applied on other pollutants time series such as NO and CO. Furthermore, the O<sub>3</sub> time series which used in this study is pure. The series currently consists of noise disturbance due to the presence of impurities. In the future, noise reduction techniques will be explored to get a time series with less noise disturbance.

## Acknowledgment

Thank you to the Department of Environment, Malaysia for providing the necessary data. This study is under the sponsorship of the research grants with the code 2015-0174-102-72 (RAGS/1/2015/SG0/UPSI/02/1), DIP-2014-034 dan FRGS/1/2014/ST06/UKM /01/1.

## References

- [1] Madaniyazi L, Nagashima T, Guo Y, Pan X and Tong S 2016 *Environ. Int.* **92-93** 165–172
- [2] Cakmak S, Hebborn C, Vanos J, Crouse D L and Burnett R 2016 *Environ. Pollut.* **214(2)** 589–99
- [3] Abarbanel H D I 1996 *Analysis of Observed Chaotic Data*. New York: Springer- Verlag 15-16
- [4] Sprott J C 2003 *Chaos and Time-Series Analysis* Oxford University Press 104-105
- [5] Cuculeanu V, Rada C and Lupu A 2009 *Geophysique.* **52-53** 77–85
- [6] Chattopadhyay G and Chattopadhyay S 2008 *Soft. Comput.* **12** 1007–12
- [7] Chelani A B 2010 *Atmos. Environ.* **44(34)** 4318–24
- [8] Petkov B H, Vitale V, Mazzola M, Lanconelli C and Lupi A 2015 *Commun. Nonlinear Sci. Numer. Simulat.* **26(1-3)** 238–49
- [9] Cao L. 1997 *Phys. D. Nonlinear Phenom.* **110(1-2)** 43–50
- [10] Sivakumar B 2002 *J. Hydrol.* **258** 149–62
- [11] Frazier C, Kockelman K 2004 *Transp. Res. Rec.* **1897(1)** 9–17
- [12] Lakshmi S S and Tiwari R K 2009 *Comput. Geosci.* **35** 191–204
- [13] Adyani N, Nor G, Ramli A, Shukri A, Faizah N, Yusof F, et al. 2010 *Environ. Monit. Assess.* **165** 475–89
- [14] Muhamad M, Ul-saufie A Z and Deni S M 2015 *J. Environ. Sci. Technol.* **8(3)** 102–12
- [15] Tan K C, Lim H S and Jafri M Z M 2016 *Atmos. Pollut. Res.* 1–14
- [16] Kocak K, Saylan L and Sen O 2000 *Atmos. Environ.* **34** 1267–71
- [17] Chen J, Islam S and Biswas P 1998 *Atmos. Environ.* **32(11)** 1839–48
- [18] Hamid N Z A and Noorani M S M 2013 *Int. J. Math. Comput. Sci. Eng.* **7(11)** 206–11
- [19] Talib M, Shun L and Juneng L 2012 *Atmos. Environ.* **61** 434–45
- [20] Ozbay B, Keskin G A, Dogruparmak S C and Ayberk S 2011 *Atmos. Res.* **102** 57–65
- [21] Toh Y Y, Lim S F and Roland V G 2013 *Atmos. Environ.* **70** 435–46
- [22] Inal F 2010 *Clean-Soil, Air, Water* **38(10)** 897–908
- [23] Hamid N Z A and Noorani M S M. 2014 *Sains Malaysiana* **43(3)** 475–81