Prediction of PM₁₀ Time Series at Industrial Area Through an Improved Local Linear Approximation Method

Nor Zila Abd Hamid

*Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris,
35900, Tanjong Malim, Perak, Malaysia*

The main objective of this study is the development of prediction model of particulate matter (PM_{10}) pollution time series. Prediction model was developed through one of mathematical approach named chaotic approach. The approach started with the detection of chaos nature of PM_{10} pollution through phase space plot and Cao method. Next, the prediction is done through the local linear approximation method. In addition, an improvement of the local linear approximation method was also made. Both methods were applied to the PM_{10} pollution observed at industrial area of Shah Alam, located in Selangor state, Malaysia. Comparison through the calculation of mean absolute error, root mean squared error and correlation coefficients was made. Results showed that the improved method is better. Therefore, improvement of the local linear approximation method is worth. Since PM_{10} pollution is associated with the occurrence of haze, it is hoped that findings of this study will help stakeholders in having a better haze management.

Keywords: time series, PM_{10}, chaos, chaotic approach, prediction, local linear approximation method, industrial area.

I. INTRODUCTION

Air pollution contributes bad effect on health. Inhalation of air containing particulate matter (PM_{10}) causes many chronic diseases such as lung cancer, stroke, cardiovascular disease, bronchitis, asthma as well as leads to mortality (Id *et. al.,* 2017; Jie, 2017; Kamarehie *et. al.,* 2017). This is a burden on community and reduces country's productivity. Department of Environment Malaysia holding responsible to observe the level, quality as well as the trend of Malaysian air pollution. Types of air pollution that been observed are sulphur dioxide (SO_2), ozone (O_3), carbon monoxide (CO), nitrogen monoxide (NO) as well as PM_{10}. However, according to Department of Environment, PM_{10} pollution is dominant compared to others. Furthermore, PM_{10} pollution is always associated with the occurrence of haze. Therefore, the study of PM_{10} time series status is very important.

There are two types of time series' dynamics: deterministic and random. Time series with deterministic dynamics, can be predicted. Vice versa, time series with

random dynamics cannot be predicted. In between deterministic and random dynamics, there is another type of dynamics namely chaos. Any time series which has been classified as chaos dynamics can be predicted, however, due to some factors, only short-term prediction is allowed (Abarbanel, 1996; Sprott, 2003).

Before prediction can be carried out, the time series need to be tested first in order to determine whether its dynamics is either deterministic, random or chaos. Many methods have been introduced such as Lyapunov exponent, entropy, integral correlation and correlation dimension. Dynamics of PM_{10} time series in some countries such as Taiwan (Yu *et. al.,* 2013) and Pakistan (Saeed *et. al.,* 2017) have been determined as chaos through methods of correlation dimension, Hurst and Lyapunov exponent. In addition, phase space plot and Cao method as has been introduced by Cao (1997) are also can be used in order to classify the time series' dynamics. In Malaysia, Hamid and Noorani (2014) used both methods to classify the dynamics of PM_{10} time series in one of rural area namely Jerantut.

*Corresponding author's e-mail: nor.zila@fsmt.upsi.edu.my

Both methods also were applied towards various data such as series of traffic flow, sediment concentration and earthquake time series (Sivakumar, 2002; Frazier & Kockelman, 2004; Lakshmi & Tiwari, 2009). Therefore, in this study, phase space plot and Cao method were chosen. In Malaysia, prediction of PM_{10} pollution is usually carried out through multiple linear regression and neural networks method (for example, Abdullah *et al.*, 2017; Ul-saufie *et. al.,* 2012a; Ul-saufie *et. al.,* 2012b; Ul-saufie *et. al.,* 2013). Prediction models through both methods are dependent on meteorological factors such as rainfall, wind speed, relative humidity and temperature. However, in the case where the needed time series' information of all involved factors is not enough, other method is important to be developed in order to run the prediction. Therefore, in this study, local approximation method, a method based on chaotic approach was used. This method has its own advantages as prediction of PM_{10} is done simply by using time series from PM_{10} pollution only, without involving time series from other factors. Local approximation method has been used in Malaysia to predict O_3 time series (for example, see Hamid & Noorani, 2013; Zaim & Hamid, 2017; Hamid, 2018a), temperature time series (Hamid, 2018b, Hamid, 2018c) and streamflow time series (Adenan *et. al.,* 2017). All studies gained satisfactory results where the predicted and observed time series are close to each other. Therefore, in this study, local approximation method was applied. For the difference, the method was applied towards PM_{10} pollution time series. There are two objectives of this study which are: i) to classify the dynamics of PM_{10} time series using phase space plot and Cao method and ii) to predict PM_{10} pollution using local approximation method. In addition, the improvement of local approximation method was made. In 2014, the first case study using chaotic approach on PM_{10} pollution was successfully carried out by Hamid and Noorani (2014). In this study, the same approach was applied. In Hamid and Noorani (2014), the effectiveness of chaotic approach has been proven on PM_{10} pollution time series in rural area. In difference, this present study applies chaotic approach on PM_{10} pollution time series in industrial area.

II. PM_{10} POLLUTION TIME SERIES

The PM_{10} time series data are obtained from Department of Environment Malaysia. The time series were observed at Shah Alam, an industrial area which located in Selangor state, Malaysia. The time series were observed hourly for one month in June 2014. The month of June and year of 2014 is selected since during this period, haze has occurred and the reading of PM_{10} is high.

The PM_{10} time series are presented in Figure 1. According to Afroz *et. al.,* (2013), the national ambient air quality standards for PM_{10} pollutant in ambient air has been determined as 150 µg/m3. From Figure 1, it can be observed that there are some hours with unhealthy reading (more than 150 µg/m3). Therefore, analysis and prediction of PM_{10} pollution during this period is very important.
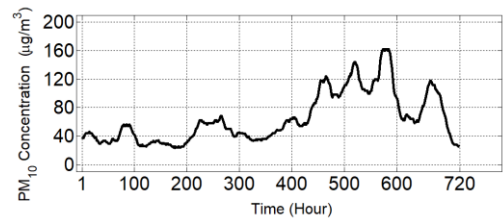

Figure 1: PM_{10} time series

III. CHAOTIC APPROACH

The PM_{10} time series from Figure 1 is recorded in the form of

$$X = \left(x_1, x_2, x_3, ..., x_{N-1}, x_N \right) \qquad (1)$$

where x_t is the PM_{10} time series at t-th hour. N is the number of total observation hour. PM_{10} time series were observed in June 2014. Therefore, there are 30 days of observation. In total, there are 720 hours. This means that there are 720 observations of PM_{10} pollution. The observed PM_{10} pollution was recorded as in (1) and divided into two parts namely X_{train} and X_{test}. X_{train} is a training time series to develop the prediction model, while X_{test} was used to test the performance of the models. In this study, time series from 1st June until 23rd June 2014 were used as X_{train} and the rest one week were used as X_{test}. The time series in (1) was transformed to a multi-dimensional vector with

equation:

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

There are two unknown parameters in (2) namely delay time, τ and embedding dimension, m. These parameters need to be calculated first. According to Velickov (2004) and Siek and Solomatine (2010), average mutual information method, introduced by Frazer and Swinney (1986) has performed well to calculate τ. Hence, the method was selected. The average mutual information is:

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

where $p(u_b)$ and $p(u_{b+T})$ are the probability of obtaining u_b and u_{b+T} in (1) while $p(u_b, u_{b+T})$ is a joint probability of $p(u_b)$ and $p(u_{b+T})$. By varying the value of T, $I(T)$ values were obtained. Next, plot of T against $I(T)$ was graphed. Parameter τ is chose based on the first minimum value of T. Furthermore, using the obtained τ value, the graph of $\{x_t, x_{t+\tau}\}$ namely phase space plot is developed. This plot was used to classify dynamics of the time series, whether it is random or chaos (Sivakumar, 2002). If an attractor exists, the dynamics of the time series are classified as chaos.

Another unknown parameter is m. Cao showed in Cao (1997) that Cao method: i) does not need other parameters; ii) use only parameter τ and iii) does not depend on the total time series. Therefore, Cao method is seeming simple and this method was applied to determine m. m from Cao method was determined by:

$$E1(m) = E(m+1)/E(m) \quad (4)$$

$$E(m) = \frac{1}{N-m\tau} \sum_{j=1}^{N-m\tau} \frac{\left\| \mathbf{Y}_j^{m+1} - \mathbf{Y}_n^{m+1} \right\|}{\left\| \mathbf{Y}_j^{m} - \mathbf{Y}_n^{m} \right\|} \quad (5)$$

where $\|\bullet\|$ is the maximum norm and $\mathbf{Y}_n^m$ is the nearest neighbour to $\mathbf{Y}_j^m$. Parameter m was determined when its respective value of $E1(m)$ stops changing. For a random time series, the value of $E1(m)$ does not stop changing with increasing m. Thus, the value of $E1(m)$ also helped

in determine the dynamics of the time series. Besides, $E2(m)$ from Cao (1997) was also applied in this study.

$$E2(m) = E*(m+1)/E*(m) \quad (6)$$

where

$$E*(m) = (1/U - m\tau) \sum_{j=1}^{U-m\tau} \left| x_{j+m\tau}^m - x_{w+m\tau}^m \right| \quad (7).$$

According to Cao method, the existence of $E2(m) \neq 1$ shows that the dynamics of observed time series is chaos.

IV. LOCAL LINEAR APPROXIMATION METHOD

Let

$$\mathbf{Y}_{j+1}^m = f\left(\mathbf{Y}_j^m \right) \quad (8).$$

This illustrates that the prediction of $\mathbf{Y}_{j+1}^m$ is depending on $\mathbf{Y}_j^m$. In order to run the prediction through (8), the method of f need to be specified. There is various type of f and in this study, the chosen method from chaotic approach is local linear approximation method. Prediction of future time series of $j+1$ through the local linear approximation method is done through the equation of:

$$\mathbf{Y}_{j+1}^m = A\mathbf{Y}_j^m + B \quad (9).$$

Where parameters A and B are constant and computed through the least square method. In addition, the method was improved. As in (9), parameters A and B are computed using X_{test}. In difference, for the improved method, each time a new prediction is run, X_{test} was updated. Therefore, for each new prediction, there was a new equation of:

$$\mathbf{Y}_{j+1}^m = A_n \mathbf{Y}_j^m + B_n \quad (10).$$

In other words, in (9), there is only one linear equation while in (10), for the total of X_{test} is n, then, there are n linear equations. The comparison of both (9) and (10) performance are computed through the mean absolute error mae, root mean squared error $rmse$ and correlation coefficient r. The values of mae and $rmse$ elaborates the difference between the observed and predicted time series.

The lower the mae and $rmse$, the better the method. On the other hand, r elaborates relationship between the observed and predicted time series. The r value is between -1 and +1 where the closer r to -1 or +1 showed the better agreement between observed and predicted time series.

V.   RESULTS AND DISCUSSION

In order to determine τ, the average mutual information method was applied. Figure 2 concludes that $\tau = 2$ since the first minimum T was two. Using the obtained $\tau = 2$, the graph of $\{x_t, x_{t+2}\}$ namely phase space plot is developed as in Figure 3. $\{x_t, x_{t+2}\}$ plot explains the revolution of observed PM_{10} pollution time series. Through Figure 3, it can be clearly observed that there exists an attractor which represented as incline line in the centre of the phase space. Referring to the study by Sivakumar (2002), the existence of attractor show that the dynamics of the time series is chaos. On the other hand, if the attractor does not exist, then the dynamics is not chaos. For further information, articles Chen *et. al.,* (1986) might be useful. Therefore, from Figure 3, the dynamics of the PM_{10} pollution time series is chaos.

Parameter m was determined where it's respective $E1(m)$ stops changing. From Figure 4, $E1(m)$ stops changing at four. Therefore $m = 4$. For a random time series, the value of $E1(m)$ does not stop changing with increasing m. Thus, from Figure 4 the observed $E1(m)$ showed that the dynamics of the PM_{10} time series is chaos. Besides, $E2(m)$ from Cao [8] also was applied in this study. According to Cao method, the existence of $E2(m) \neq 1$ shows that the dynamics of observed time series is chaos. $E2(m) \neq 1$ from $m = 1$ until $m = 8$. Therefore, the observed $E2(m) \neq 1$ showed that the dynamics of the PM_{10} time series is chaos.
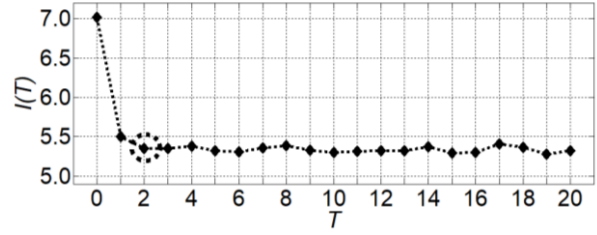

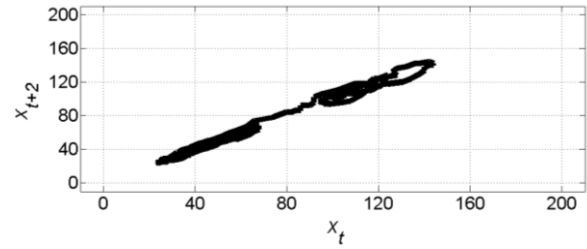Figure 2: Average mutual information method

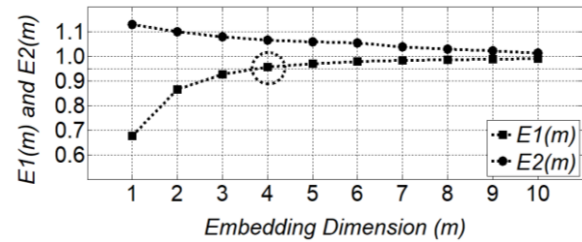Figure 3: Phase space plot

Figure 4: Cao method

With $\tau = 2$ and $m = 4$, equation (2) is developed and prediction is done through (9) and (10). The prediction period is one week, which equal to 168 hours. Performance of the model is reflected in the calculation of mae, $rmse$ and r. Table 1 presents the comparison of performance values between local linear approximation method (LLAM) and improved local linear approximation method (ILLAM). It can be seen that both models gain good prediction results in which the r's are above 0.8 and approaches one. r values exceeding 0.8 indicate that there is a strong relationship between the observed and predicted time series values. By comparing the performance indicator of both methods, the improved one was seen more powerful with lower mae and $rmse$ and higher r. The reduction of the mae and $rmse$ values indicates that the model is improving and increasing in the r approaching one indicates that the relationship between the observed and predicted time series is getting stronger. Therefore, the improved method is advantageous.

Figure 5 is the graph of prediction results for the

improved method. As can be observed in Figure 5, the model is able to predict high PM_{10} time series where the haze phenomenon is reported occur. Figure 6 is the scatter plot of the results. It can be clearly observed that through the improved method, the results are good where the observed and predicted time series are close to each other.
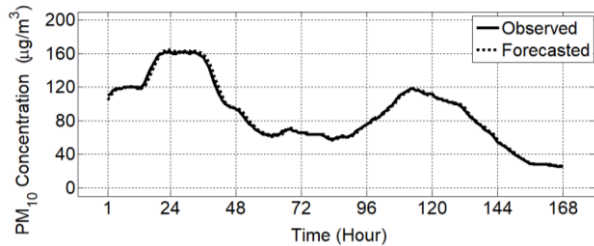

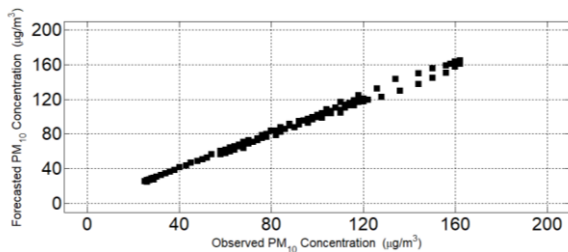Figure 5: Prediction results through ILLAM

Figure 6: Scatter plot of Prediction results through ILLAM

Table 1: Performance of the prediction models

	LLAM	ILLAM
mae	2.4881	1.8274
rmse	3.2090	2.4556
cc	0.9972	0.9981

Through original method, only one linear equation is developed to predict the entire X_{test}. Through the improved method, every time a new prediction is carried out, the X_{test} is updated. This method is more relevant because whenever new prediction is done, all required information changes.

VI. CONCLUSION

There are two objectives of this study which are: i) to classify the dynamics of PM_{10} time series using phase space plot and Cao method and ii) to predict PM_{10} pollution using local approximation method. In addition, the improvement of local approximation method was made. Results showed that the dynamics of PM_{10} time series at Shah Alam is chaos. Two prediction models were developed. The first model is the original model of local linear approximation method. The second model is the improvement of the original model. Comparison found that the improved model was better. It is hoped that the developed prediction models can assist stakeholders in having a better PM_{10} pollution management. In future, chaotic approach may be applied towards NO_x, CO and SO_2 as well as wind speed and relative humidity.

VII. ACKNOWLEDGEMENT

VIII. REFERENCES

Abarbanel, H. D. I. 1996, *Analysis of Observed Chaotic Data*. New York: Springer-Verlag.

Abdullah, S., Ismail, M. &Fong, S. I. Y. 2017, Multiple linear regression (MLR) models for long term PM_{10} concentration forecasting during different monsoon seasons. *Journal of Sustainability Science and Management*, 12(1), 60–69.

Adenan, N. H., Hamid, N. Z. A., Mohamed, Z. & Noorani, M. S. M. 2017, A pilot study of river flow prediction in urban area based on phase space reconstruction. *AIP Conference Proceedings*, 1870.

Afroz, R., Hassan, M. N. &Ibrahim, N. A. 2013, Review of air pollution and health impacts in Malaysia. *Environmental Research*, 92, 71–77.

Cao, L. 1997, Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D*,110,43–50.

Chen, J., Islam, S. &Biswas, P. 1998, Nonlinear dynamics of hourly ozone concentrations: nonparametric short-term prediction. *Atmospheric Environment*, 32(11), 1839–1848.

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

Frazier, C. & Kockelman, K. M. 2004, Chaos theory and transportation systems: An instructive example. *Transportation Research*, 1897, 9–17.

Hamid, N. Z. A. 2018a, Prediction of ozone pollution through chaotic approach.

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

Hamid, N. Z. A. 2018b, Application of chaotic approach in forecasting highland's temperature time series. *IOP Conference Series: Earth and Environmental Science*, 69(012107), 1–10.

Hamid, N. Z. A. 2018c, Environmental modelling through chaotic approach: A case study on ozone and temperature time series. *International Journal of Engineering & Technology*, 7(3.7), 590–593.

Hamid, N. Z. A. & Noorani, M. S. M. 2013, An improved prediction model of ozone concentration time series based on chaotic approach. *International Journal of Mathematical, Computational Science and Engineering*, 7(11), 206-211.

Hamid, N. Z. A. & Noorani, M. S. M. 2014, A pilot study using chaotic approach to determine characteristics and forecasting of PM_{10} concentration time series. *Sains Malaysiana*, 43(3), 475–481.

Id, Y. G., Lin, H., Liu, T., Lv, X. &Ma, W. 2017, The interaction between ambient PM_{10} and NO_2 on mortality in Guangzhou, China. *International Journal of Environmental Research and Public Health Article*, 2, 1-10.

Jie, Y. 2017, Air pollution associated with Sumatran forest fires and mortality on the Malay Peninsula. *Polish Jounal of Environmental Studies*, 26(1), 163–171.

Kamarehie, B., Ghaderpoori, M., Jafari, A., Karami, M. & Mohammadi, A. 2017, Estimation of health effects (morbidity and mortality) attributed to PM_{10} and $PM_{2.5}$ exposure using an air quality model in Bukan city, from 2015-2016 exposure using air quality model. *Environmental Health Engineering and Management Journal*, 4(3), 137–142.

Lakshmi S. S. & Tiwari, R. K. 2009, Model dissection from earthquake time series: A comparative analysis using modern non-linear forecasting and artificial neural network approaches. *Computers & Geosciences*,35, 191–204.

Saeed, S., Aziz, W., Rafique, M., Ahmad, I., Kearfott, K. J. &Batool, S. 2017, Quantification of non-linear dynamics and chaos of ambient particulate matter concentrations in Muzaffarabad city. *Aerosol Air Quality Research*, 17, 849–856.

Siek, M. & Solomatine, D. P. 2010, Nonlinear chaotic model for predicting storm surges. *Nonlinear Processes in Geophysics*, 17, 405–420.

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

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

Ul-saufie, A. Z., Yahaya, A. S., Ramli, N. A., Rosaida, N. &Hamid, H. A. 2013, Future daily PM_{10} concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). *Atmospheric Environment*, 77, 621–630.

Ul-saufie, A. Z., Yahaya, A. S., Ramli, N. A. &Hamid, H. A. 2012a, Future PM_{10} concentration prediction using quantile regression models. *2nd International Conference on Environmental and Agriculture Engineering*, 37, 15–19.

Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A. & Hamid, H. A. 2012b, Robust regression models for predicting PM_{10} concentration in an industrial area. *International Journal of Engineering and Technology*, 2(3), 364–370.

Velickov, S. 2004, *Nonlinear Dynamics and Chaos with Application to Hydrodynamics and Hydrological Modelling*. Delft University of Technology, Netherlands.

Yu, H., Lin, Y., Sivakumar, B. & Kuo, Y. 2013, A study of the temporal dynamics of ambient particulate matter using stochastic and chaotic techniques. *Atmospheric Environment*, 69, 37–45.

Zaim W. N. A. W. M. & Hamid, N. Z. A. 2017, Forecasting ozone pollutant (O_3) in Universiti Pendidikan Sultan Idris, Tanjung Malim, Perak, Malaysia based on monsoon using chaotic approach. *Sains Malaysiana*, 46(12), 2523–2528.