Journal of Quality Measurement and Analysis *JQMA* 20(3) 2024, 23-34
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2003.2024.02

PERFORMANCE COMPARISON OF HAZE PREDICTION USING CHAOS THEORY AND MULTIPLE LINEAR REGRESSION

(Perbandingan Prestasi Peramalan Jerebu Menggunakan Teori Kalut dan Regresi Linear Berganda)

HAZLINA DARMAN* & NOR ZILA ABD HAMID

ABSTRACT

Forecasting haze is essential for protecting the environment, the economy, and public health. It assists authorities in taking preventative action to lessen the adverse effects of haze episodes and boost community resistance to air pollution. The goal of this study was to create a model for haze prediction by using two methods, multiple linear regression and chaos theory. In this study, chaos theory forecasts haze using univariate time series which is PM_{10}, whereas multiple linear regression (MLR) utilizes multivariate time series for its predictions, namely ambient temperature, wind speed, ozone, nitrogen dioxide, carbon monoxide, and sulphur dioxide. Data for this study will be collected during the southwest monsoon from an industrial area in Klang, Selangor. The results of these two models will be compared to determine which model gave better performance. With these predictive models, policymakers and relevant authorities can receive timely alerts, allowing them to implement preventive measures that can reduce the impact of haze on public health and the environment.

Keywords: haze forecasting; chaos theory; multiple linear regression; sustainability development goals

ABSTRAK

Peramalan jerebu adalah penting untuk alam sekitar, ekonomi, dan kesihatan manusia,. Ia dapat membantu pihak berkuasa untuk mengambil langkah proaktif dalam mengurangkan kesan negatif kejadian jerebu dan meningkatkan daya tahan masyarakat terhadap kesan pencemaran udara. Kajian ini bertujuan membangunkan model peramalan jerebu menggunakan dua kaedah, iaitu regresi linear berganda dan teori kalut. Dalam kajian ini, teori kalut menggunakan univariat siri masa PM_{10} manakala regresi linear berganda menggunakan beberapa pemboleh ubah bebas untuk peramalan jerebu seperti: suhu atmosfera, kelajuan angin, ozon, nitrogen doksida, karbon monoksida, dan sulfur dioksida. Dalam kajian ini, data akan diambil dari kawasan perindustrian Klang, Selangor semasa musim Monsun Barat Daya. Prestasi kedua-dua model ini akan dibandingkan untuk menentukan model yang memberikan peramalan yang lebih baik. Dengan menggunakan model peramalan ini, pihak berkuasa penggubal dasar dan pihak berkuasa yang berkaitan boleh menerima makluman tepat pada masanya, membolehkan mereka melaksanakan langkah pencegahan yang boleh mengurangkan kesan jerebu terhadap kesihatan awam dan alam sekitar.

Kata kunci: teori kalut; peramalan jerebu; regresi linear berganda; matlamat pembangunan mampan

1. Introduction

Ministry of Health Malaysia (2020) defined haze as a state of deteriorating air quality caused by the presence of fine particulate matter in the air as reported in Health Management Actions Due to Haze Plan. Due to the great concentration, these particles scatter and absorb sunlight, making visibility worse. Haze and fog are weather phenomena that are closely related, consist of water droplets suspended in the air close to the earth's surface and haze refers to particles suspended in the air that reduce visibility by scattering light (Li *et al.* 2019). The southwest monsoon season brings frequent haze, particularly during hot and dry weather each year, from the end of May to the middle of September, The El Nino's dry spells lengthen the period of unhealthy air quality (Latif *et al.* 2018). Natural sources and human activity are the two main sources of haze. Dust storms, wildfires, and volcanic eruptions are examples of natural origins of haze. In the meanwhile, human activities include agricultural activities, open burning, fuel combustion, and automobile emissions (Latif *et al.* 2017).

There are several detrimental effects that haze can have on the environment and human health. According to the Ministry of Health Malaysia (2020), haze can cause respiratory system infections, lower lung function, irritation of the eyes, irritation of the nose and throat, severe headaches, nausea, and vomiting. Haze is also linked to acute respiratory, psychological, and neurological issues, as well as cardiovascular problems and increased mortality (Cheong *et al.* 2019). The economic impact of haze on Malaysia can be substantial. The 2013 haze episode alone resulted in an estimated economic cost of MYR 1.49 billion, equivalent to 0.48% of the GDP (Varkkey & Copeland 2020). Hospital admissions due to haze cost between MYR1.8 and MYR118.9 million, while the cost of disease had increased from MYR21million to MYR410 million (Latif *et al.* 2018). Haze can cause disruptions to several industries and businesses, including a decline in business and tourism, a decrease in productivity, and additional expenses for mitigation and adaptation. (Quah *et al.* 2021).

Haze contains particles of $PM_{2.5}$, PM_{10}, Sulphur Dioxide (SO_2), ozone (O_3), Carbon Monoxide (CO), and Nitrogen Dioxide (NO_2). PM_{10} refers to fine particulate matter with a diameter of 10 micrometers or smaller (Talhelm 2024). Out of all the particles in haze, it makes up the greatest portion and the sources include smoke, dust, and vehicle emissions. Higher levels of air pollution are indicated by rising PM_{10} concentrations in the atmosphere. Studies by Abd Abdullah *et al.* (2017; 2020), Abd Hamid and Md Noorani (2014), Ul-Saufie *et al.* (2012) and Shang *et al.* (2021) had used PM_{10} to develop the haze predicting model in their papers since it is the most dominant component in haze. Thus, this study will also use PM_{10} time series to predict haze.

Several approaches have been explored to improve the accuracy and effectiveness of haze prediction models, such as chaos theory (Darman & Abd Hamid 2024; Abd Hamid 2020; Abd Hamid & Md Noorani 2014; Pino-Vallejo *et al.* 2018), long-short term memories (Wu *et al.* 2021; Liu *et al.* 2019), neural network (Zhang *et al.* 2021; Pinghua 2018), deep learning (Idris & Yassin 2021; Hasmarullzakim & Abdullah 2018), and multiple linear regression (Chen & Wang 2019; Abdullah *et al.* 2017).

This paper predicted haze by employing two methods (i) chaos theory, and (ii) multiple linear regression. Chaos theory required only PM_{10} time series, while multiple linear regression predicted PM_{10} using several meteorological factors, namely Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), Sulphur Dioxide (SO_2), ozone (O_3), wind speed, and ambient temperature.

1.1. *Haze and Sustainable Development Goals (SDGs)*

Haze is closely connected to several of the Sustainable Development Goals (SDGs) established by the United Nations. Firstly, haze contributes to air pollution, which directly impacts SDG 3: Good Health and Well-being. The pollutants in haze can lead to respiratory problems and other health issues, affecting the well-being of individuals and communities. Additionally, haze is linked to SDG 13: Climate Action.

Haze is commonly produced by activities, such as forest burning, which emit greenhouse gases and contribute to climate change. By addressing the issue of haze, countries can take steps towards mitigating climate change and achieving the targets set under SDG 13. Furthermore, haze also has implications for SDG 15: Life on Land. The burning of forests during haze episodes can result in habitat destruction and loss of biodiversity, impacting ecosystems and the preservation of terrestrial life. By addressing haze, efforts can be made to protect and restore ecosystems, aligning with the objectives of SDG 15.

Overall, addressing the issue of haze is crucial for achieving multiple SDGs, including those related to health, climate action, and the preservation of biodiversity.

1.2. *Dataset*

In this study, haze will be predicted using the PM_{10} time series where the data was taken from Klang, Selangor, Malaysia. 29 Klang areas have been gazetted as industrial areas, according to the Selangor State Council's official website (Dewan Negeri Selangor 2016). Since the 1980s, Klang has been known to encounter high particle concentrations. Soleiman *et al.* (2003) evaluated the extreme haze occurrences that the Klang Valley experienced over the years. According to their paper, Klang has a significant risk of pollution due to the city's fast industrial development and urbanization. The Klang region will therefore be an ideal location to predict the haze.

The Department of Environment (2019) shows illustrates the trend in the annual average PM_{10} concentration levels in ambient air from 2010 to 2019. According to land use groups, industrial areas had the highest concentration of PM_{10}, followed by urban, suburban, and rural areas. Researching haze prediction in this field is crucial because industrial sites often contribute to the emission of pollutants into the atmosphere. These areas were frequently populated by factories, power plants, and other industrial facilities that generated pollutants such as particulate matter (PM2.5 and PM_{10}), sulphur dioxide (SO2), and nitrogen oxide (NO). These pollutants have the potential to combine with other elements in the atmosphere to produce haze, which is bad for both the air quality and human health. Department of Environment, Malaysia provided datasets for air pollutants taken from 1st May 2017 to 30th September 2017, including 3671 hourly observations of PM_{10} time series. This particular time frame was selected because of the southwest monsoon and El Nino events, which will cause a high concentration of haze. Table 1 provides the statistical analysis of the PM_{10} time series, while Figure 1 illustrates the trend of PM_{10} during the southwest monsoon in Klang, Selangor.

Table 1: Statistical analysis of PM_{10} in Klang during southwest monsoon

Total data	Number of train data	Number of test data	Minimum value	Maximum value	Mean value	Standard deviation	Median	Mode	Skewness
3672	3336	336	0.58	353.547	38.329	0.364	35.36	38.329	2.658

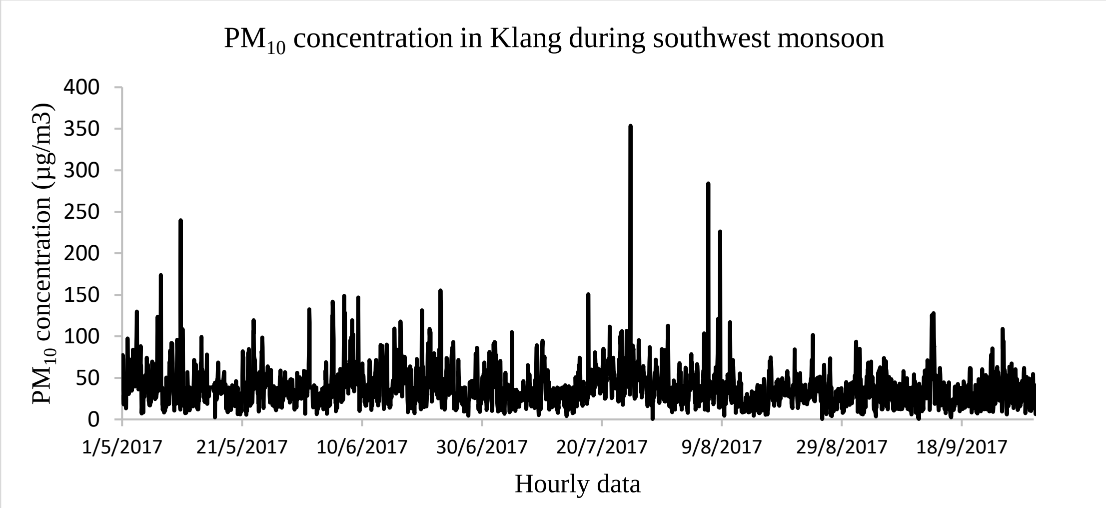

Figure 1: PM_{10} during southwest monsoon in Klang, Malaysia

2. Chaos Method

The two steps of the chaos technique were the phase space reconstruction and the prediction process. Phase space will be reconstructed using the phase space approach and the Cao method. While the Cao method can distinguish the nature of the time series, the phase space approach is easier to use. As a result, this study will use these two approaches.

2.1. *Reconstruction of phase space*

The one-dimensional scalar time series is provided below:

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

where X was the PM_{10} time series for a total of N observations. X was divided into two parts: training dataset, X_{train} and testing dataset, X_{test}. The equations are shown below.

$$X_{train} = \{x_1, x_2, x_3, \dots, x_l\} \tag{2}$$

$$X_{test} = \{x_{l+1}, x_{l+2}, x_{l+3}, \dots, x_N\} \tag{3}$$

The chaos parameters such as embedded dimension, m and time delay, τ will be obtained from the training dataset while the prediction process will use the testing dataset. In this study, the first 3,336 hourly data points will be used to train the model, and the subsequent 336 hourly data will be utilized to test the model's performance. Two parameters, time delay, τ and embedding dimension, m need to be determined first before the phase space is reconstructed. The training dataset X_{train} will be reconstructed into m-dimensional phase space as shown below.

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

To obtain a time delay τ, this paper applied the average mutual information method which was introduced by Fraser and Swinney (1986). Eq. (5) introduced function $I(T)$ that denotes the mutual information using different values of τ and *T* is various value of time delayτ as shown below:

$$I(T) = \frac{1}{N}\sum_{t=1}^{N} p(x_t, x_{t+\tau})\, log\left[\frac{p(x_t, x_{t+\tau})}{p(x_t)p(x_{t+\tau})}\right] \tag{5}$$

where x_t is the original time series, $x_{t+\tau}$ is time series shifted by τ, $p(x_t)$ and $p(x_{t+\tau})$ are the probability of x_t and $x_{t+\tau}$ accordingly and $p(x_t, x_{t+\tau})$ is their joint probability. By varying the τ, *I*(*T*) is attained. Graph *T* against *I*(*T*) is then plotted, and the first minimum value of *T* will determine the time delay τ. Using the obtained τ, the phase space graph is plotted to distinguish if the time series is chaotic by checking the existence of an attractor (Sivakumar 2002).

Cao method was introduced by Cao (1997) which provides the following benefits: (i) independent of the data size, (ii) does not rely on other parameters except time delay τ. The embedding dimension, *m*, is estimated using the Cao method. To investigate the variation from *m* to *m* + 1, the variable below is defined:

$$E_1(m) = \frac{E(m+1)}{E(m)} \tag{6}$$

and mean is selected to estimate the divergence of nearest neighbours given by

$$E(m) = \frac{1}{N-m\tau}\sum_{j=1}^{N-m\tau}\frac{\left\|Y_j^{m+1} - Y_N^{m+1}\right\|}{\left\|Y_j^{m} - Y_N^{m}\right\|} \tag{7}$$

where $\|\bullet\|$ denotes the Euclidean distance and $\left\|Y_j^m\right\|$ was the nearest neighbour to Y_j^m. Parameter *m* was determined when $E_1(m)$ converges to a certain minimum embedding dimension. Aside from determining the embedding dimension, Cao method defines another function, $E_2(m)$ to check whether the time series is chaotic. If exist cases that $E_2(m) \neq 1$, then the dynamic of the time series is chaotic. Its variation from *m* to *m* + 1 is defined as

$$E_2(m) = \frac{E^*(m+1)}{E^*(m)} \tag{8}$$

where E^* is the mean separation of the nearest neighbour in the *m*-dimensional space, such that

$$E^*(m) = \frac{1}{N-m\tau}\sum_{j=1}^{N-m\tau}\left|x_{j+m\tau}^m - x_{N+m\tau}^m\right| \tag{9}$$

2.2. *Prediction: Local Mean Approximation Method (LMAM)*

In this study, Local Mean Approximation Method (LMAM) will be used to predict the haze. Phase space had been reconstructed as shown in section 2.1 by applying Eq. (4) and the prediction process is defined:

$$Y_{j+1}^m = f\left(Y_j^m\right) \tag{10}$$

The prediction of Eq. (11) was done based on the neighbours of Y_j^m. The neighbours were labeled as $Y_{j'}^m$ for $1 < j < j'$ and the neighbours with the smallest value of Euclidean distance

$\left\|Y_{j'}^{m} - Y_{j}^{m}\right\|$ will be used for prediction, which was denoted as k nearest neighbours. We use the trial and error method to determine the value of k. In this paper, we let parameter k equal to 50. After $Y_{j'}^{m}$ was determined, the one hour ahead $Y_{j'+1}^{m}$ will be listed. The prediction of Y_{j+1}^{m} was taken as the mean of $Y_{j'+1}^{m}$ values as shown below:

$$Y_{j+1}^{m} = \frac{\sum_{q=1}^{k} Y_{j_q'+1}^{m}}{k} \tag{11}$$

3. Multiple Linear Regression

Multiple linear regression (MLR) is a statistical method used to model the relationship between a dependent variable and several independent variables. It expands on the concept of simple linear regression, which involves only one independent variable, to accommodate multiple predictors. The goal of multiple linear regression is to create a mathematical model that explains the relationship between the dependent variable and the independent variables, allowing for the prediction of the dependent variable based on the given values of the predictors. The general equation of MLR is as follows:

$$y = \beta_0 + \sum_{i=1}^{N} \beta_i X_i + \varepsilon \tag{12}$$

where y is the dependent variable, X_i is the independent variables, β_0 is the constant term, β_i are the regression coefficients, and ε is the stochastic error associated with regression. MLR assumes that the residuals follow a normal distribution have zero mean are uncorrelated and have constant variance. In this study, the dependent variable will be PM_{10} (in $\mu g/m^3$ unit) with six independent variables of SO_2 (in ppm unit), NO_2 (in ppm unit), O_3 (in ppm unit), CO (in ppm unit), windspeed (in meter/second unit), and ambient temperature (in Celsius).

3.1. *Variable of Inflation (VIF)*

Multicollinearity occurs when the independent variables in a regression model are correlated with each other. To check for multicollinearity, method Variable of Inflation (VIF) will be used. If the average VIF is below 10, then no multicollinearity exists (Ul-Saufie *et al.* 2012). The VIF is given as below:

$$VIF = \frac{1}{1-R_i^2} \tag{13}$$

where R_i^2 is the coefficient of determination with i -th predictor.

3.2. *Durbin Watson test*

Durbin Watson test is used to measure the autocorrelation in residuals. The value ranges between 0 and 4 and if DW = 2, then no autocorrelation in residual. The formula is given by

$$d = \frac{\sum_{i=1}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \tag{14}$$

where n is the number of observations and e_i is the i-th residual from the regression model.

4. Result and Discussion

4.1. *Chaos method*

Average mutual information method was applied to determine the time delay τ. Figure 2 concludes that $\tau = 8$ since the first minimum T was eight. Using $\tau = 8$, the attractor plot of PM_{10} time series was formed in two dimensions and three dimensions as shown in Figure 3 and Figure 4. From these figures, it can be clearly seen that an attractor existed in the region which indicated chaos dynamics in PM_{10} time series. The presence of a well-defined attractor in a low-dimensional phase space indicates that the data may exhibit chaotic behavior.

Embedded dimension m was determined when its respective $E_1(m)$ saturated. From Figure 5, $E_1(m)$ saturated at value six, thus the embedded dimension m = 6. In contradiction, for random time series, the value of $E_1(m)$ will not saturate with increasing of m. Besides, according to Cao method, if existed any $E_2(m) \neq 1$, then the dynamic of PM_{10} time series was chaos, which can be seen from Figure 5. With $\tau = 8$ and $m = 6$, Eq. (12) was developed and prediction was performed using Local Mean Approximation Method (LMAM). The prediction period was for 14 days, which equalled to 336 hours.

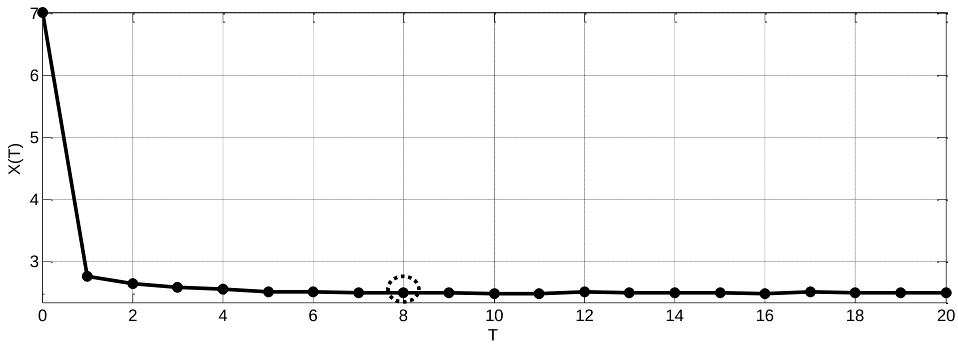

Figure 2: Time delay τ using average mutual information method

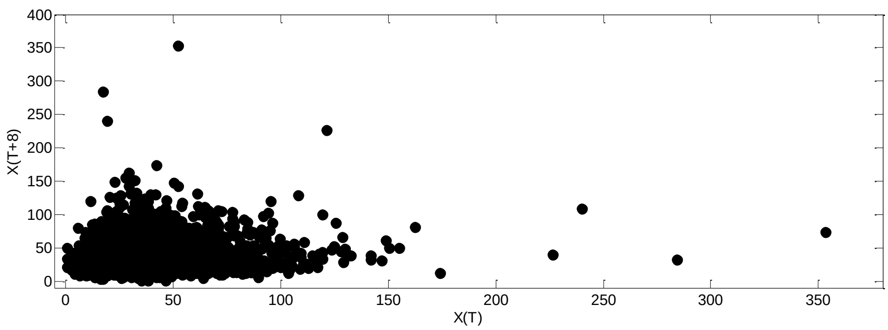

Figure 3: Attractor plot in two dimensions

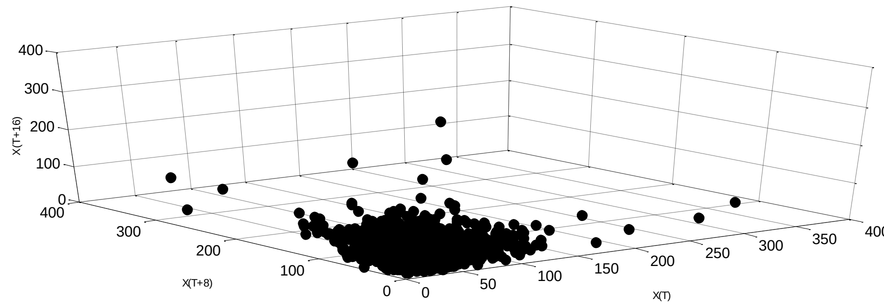

Figure 4: Attractor plot in three dimensions

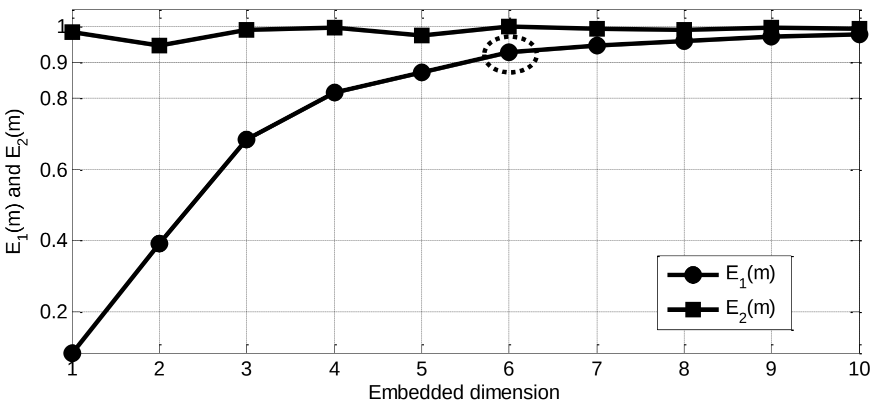

Figure 5: Embedding dimension using Cao method

4.2. *Multiple linear regression method*

The developed multiple linear regression model was given as below:

$$Y = -4.631 + 189.094X_1 + 850.015X_2 - 62.557X_3 + 13.379X_4 \\ -0.055X_5 + 0.634X_6 \tag{15}$$

where Y was PM_{10}, X_1 was SO_2, X_2 was NO_2, X_3 was O_3, X_4 was CO, X_5 was wind speed, and X_6 was the ambient temperature. The coefficient of determination, R^2 was 0.210 which indicated 21% of variations of PM_{10} were explained by all the independent variables, while another 79% of variations were explained by other factors. The Durbin-Watson test was 0.924 which fell in an acceptable range of 0 to 4. This indicated the model did not have any first-order autocorrelation problem. Moreover, the range of VIF was lower than 10 as 1.006 to 1.572, which indicated no multicollinearity in variables. The correlation of each independent variable with the PM_{10} is given in Table 2 below:

Table 2: Correlation and p-values between independent variables and PM_{10}

Meteorological factors	SO2	NO2	O3	CO	Windspeed	Ambient temperature
Correlation	0.076	0.4	-0.161	0.367	-0.033	-0.077
p-value	< 0.001	< 0.001	0.088	< 0.001	0.132	< 0.001

From Table 2, there was a very weak positive correlation between PM_{10} and SO_2, a weak positive correlation of NO_2 and CO with PM_{10}, also a very weak negative correlation of ozone, windspeed and ambient temperature with PM_{10}. The *p*-values also showed that SO_2, NO_2, CO, and ambient temperature were significant factors in predicting PM_{10} time series. From Eq. (16), with one unit increase of PM_{10}, there will be 189.094 units increase in SO_2, 850.015 units increase in NO_2, 13.379 units increase in CO, 0.634 unit increase in ambient temperature, 62.557 units decrease in O_3, and 0.055 unit decreased in wind speed.

4.3. *Performance measure*

The performance of both methods, the Chaos method and MLR was measured using absolute accuracy error (AAE) and correlation coefficient (CC). The values of AAE described the difference between the observed and predicted data. The smaller the MAE, the better the method. The CC explained the relationship between the observed and predicted data. The value ranged between -1 and +1. The closer the value to -1 or +1, the stronger the relationship between the observed and predicted data. In contradiction, if the value is close to 0, the relationship between the observed and predicted data is weak.

Performance of predicting PM_{10} for both chaos method and multiple linear regression method were measured using correlation coefficient (CC) and absolute accuracy error (AAE) as shown in Table 3 below:

Table 3: Performance measure between chaos method and multiple linear regression

Performance measure	Correlation coefficient (CC)	Absolute accuracy error (AAE)
Chaos theory	0.6420	9.6594
Multiple linear regression	0.4913	11.350

As can be observed in Table 3, the chaos method produced a higher correlation coefficient (CC) of 0.6420 compared to multiple linear regression, 0.4913. The higher value of CC indicated better prediction because the relationship between the observed and predicted PM_{10} was stronger. Again, the chaos method predicted PM_{10} better compared to multiple linear regression because it produced a smaller absolute accuracy error.

5. Conclusion

It was proven that PM_{10} time series has chaotic behaviour from the Cao method that had been performed earlier. This can be proven by the attractor plot in Figure 3 and Figure 4. Therefore, the chaos method was suitable for predicting PM_{10} using LMAM. Meanwhile, MLR was a good method to determine the significant factors that affect PM_{10}. However, the chaos method performed better compared to MLR in predicting haze because it produced higher CC and smaller AAE. From Figure 6 and Figure 7, we can also conclude that prediction model using chaos method fitted better with the observed PM_{10} compared to MLR. The observed data in the graph is the testing data that is shown by Eq. (3). For future study, a predicting model that apply chaos method which incorporates all meteorological factors can be developed that might produce better prediction. The result of this study can be used by the Department of Environment, Malaysia to give early warnings to people on haze occurrence or they can take preventive actions before haze occur.

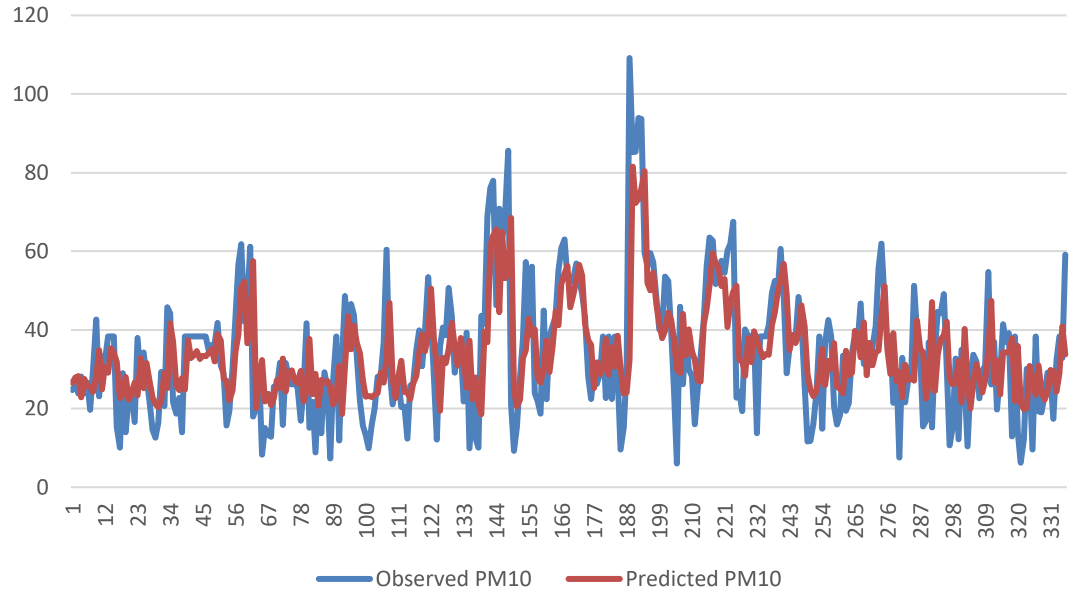

Figure 6: Comparison of predicted PM_{10} with observed PM_{10} using chaos method

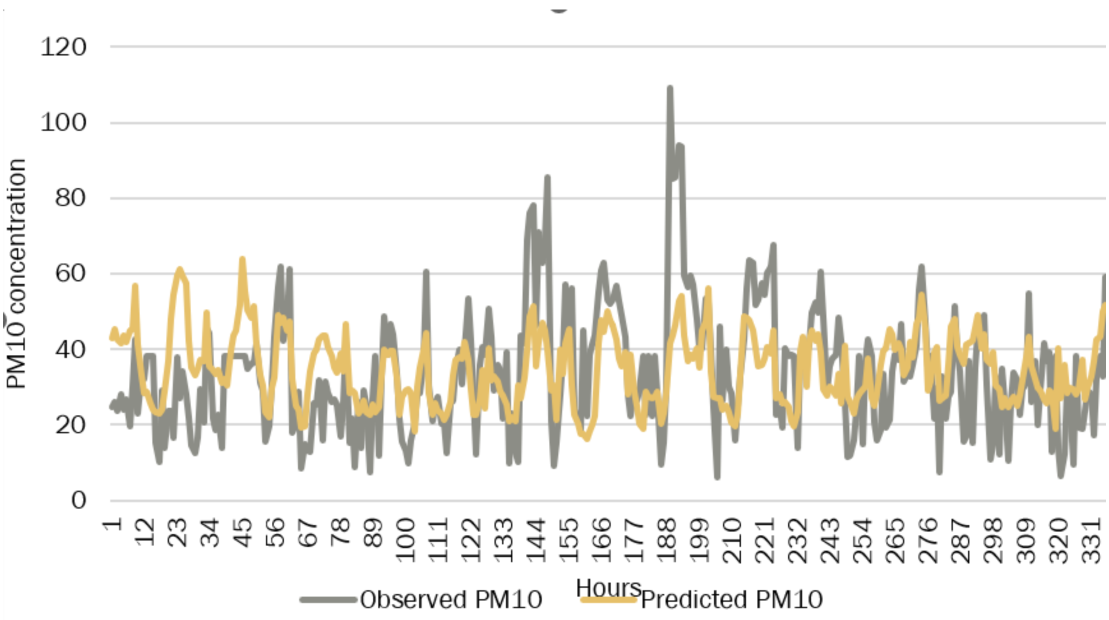

Figure 7: Comparison of predicted PM_{10} with observed PM_{10} using multiple linear regression

Acknowledgement

Thank you to the Department of Environment, Malaysia for providing datasets on air pollutants. This study is under the sponsorship of University Pendidikan Sultan Idris (UPSI).

References

Abd Hamid N.Z. 2020. Prediction of PM_{10} time series at industrial area through an improved local linear approximation method. *ASM Science Journal* **13**: 1-7.

Abd Hamid N.Z. & Md Noorani M.S. 2014. Suatu kajian perintis menggunakan pendekatan kalut bagi pengesanan sifat dan peramalan siri masa kepekatan PM10. *Sains Malaysiana* **43**(3): 475-481.

Abdullah S., Ismail M. & Fong S.Y. 2017. Multiple linear regression (MLR) models for long term PM10 concentration forecasting during different monsoon seasons. *Journal of Sustainability Science and Management* **12**(1): 60–69.

Abdullah S., Napi N.N.L.M., Ahmed A.N., Mansor W.N.W., Mansor A.A., Ismail M., Abdullah A.M. & Ramly Z.T.A. 2020. Development of multiple linear regression for particulate matter (PM_{10}) forecasting during episodic transboundary haze event in Malaysia. *Atmosphere* **11**(3): 289.

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.

Chen J. & Wang J. 2019. Prediction of PM2.5 concentration based on multiple linear regression. *2019 International Conference on Smart Grid and Electrical Automation (ICSGEA)*, pp. 457–460.

Cheong K.H., Ngiam N.J., Morgan G.G., Pek P.P., Tan B.Y.-Q., Lai J.W., Koh J.M., Ong M.E.H. & Ho A.F.W. 2019. Acute health impacts of the Southeast Asian transboundary haze problem—A review. *International Journal of Environmental Research and Public Health* **16**(18): 3286.

Darman H. & Abd Hamid N.Z. 2024. Predicting haze phenomenon using chaos theory in industrial area in Malaysia. *Journal of Quality Measurement and Analysis* **20**(1): 159-169.

Department of Environment. 2019. *Annual Environmental Quality Report 2019*.

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

Hasmarullzakim A.M. & Abdullah A. 2018. Haze pattern prediction using deep learning. *UTM Computing Proceedings* **3**.

Idris A.C. & Yassin H. 2021. Deep learning method for haze prediction in Singapore. *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pp. 1–6.

Dewan Negeri Selangor. 2016. Kawasan perindustrian di Selangor. https://dewan.selangor.gov.my/question/kawasan-perindustrian-di-selangor/ (1 July 2023).

Ministry of Health Malaysia. 2020. *Pelan Tindakan Pengurusan Kesihatan Akibat Jerebu*.

Latif M.T., Othman M. & Kamin K.H. 2017. *Fenomena Jerebu di Asia Tenggara: Punca dan penyelesaian*. UKM Bangi, Selangor: Insitute for Environment and Development (LESTARI).

Latif M.T., Othman M., Idris N., Juneng L., Abdullah A.M., Hamzah W.P., Khan M.F., Nik Sulaiman N.M., Jewaratnam J., Aghamohammadi N., Sahani M., Xiang C.J., Ahamad F., Amil N., Darus M., Varkkey H., Tangang F. & Jaafar A.B. 2018. Impact of regional haze towards air quality in Malaysia: A review. *Atmospheric Environment* **177**: 28–44.

Li W., Liu X., Zhang Y., Tan Q., Feng M., Song M., Hui L., Qu Y., An J. & Gao H. 2019. Insights into the phenomenon of an explosive growth and sharp decline in haze: A case study in Beijing. *Journal of Environmental Sciences* **84**: 122–132.

Liu X., Liu Q., Zou Y. & Liu Q. 2019. A LSTM-based approach to haze prediction using a self-organizing single hidden layer scheme. In Yang C.N., Peng S.L. & Jain L. (eds.). *Security with Intelligent Computing and Big-data Services. SICBS 2018. Advances in Intelligent Systems and Computing, vol 895.* Cham: Springer.

Pinghua Z. 2018. Empirical study on haze forecasting model based on neural network. *IOP Conference Series: Materials Science and Engineering* **439**: 032073.

Pino-Vallejo M., Tierra A., Haro A. & Perugachi N. 2018. Prediction of concentrations of PM2.5 in Downtown Quito using the chaos theory. *AIP Conference Proceedings* **2003**: 020013.

Quah E., Chia W.-M. & Tan T.-S. 2021. Economic impact of 2015 transboundary haze on Singapore. *Journal of Asian Economics* **75**: 101329.

Shang K., Chen Z., Liu Z., Song L., Zheng W., Yang B., Liu S. & Yin L. 2021. Haze prediction model using deep recurrent neural network. *Atmosphere* **12**(12): 1625.

Sivakumar B. 2002. A phase-space reconstruction approach to prediction of suspended sediment concentration in rivers. *Journal of Hydrology* **258**(1–4): 149–162.

Soleiman A., Othman M., Samah A.A., Sulaiman N.M. & Radojevic M. 2003. The Occurrence of Haze in Malaysia: A Case Study in an Urban Industrial Area. In Rao G.V., Raman S. & Singh, M.P. (eds.). *Air Quality*, pp. 221-238. Birkhäuser, Basel: Pageoph Topical Volumes.

Talhelm T. 2024. *Particle Air Pollution: PM10 vs PM2.5*. https://smartairfilters.com/en/blog/pm10-pm2-5-difference-particle-air-pollution/ (15 April 2024).

Ul-Saufie A.Z., Yahaya A.S., Ramli N. & Hamid H.A. 2012. Performance of multiple linear regression model for long-term PM10 concentration prediction based on gaseous and meteorological parameters. *Journal of Applied Sciences* **12**(14): 1488–1494.

Varkkey H. & Copeland A. 2020. Exchanging health for economic growth? Haze in the context of public health and political economy in Malaysia. *Journal of Social Health* **3**(1): 1-13.

Wu X., Liu Z., Yin L., Zheng W., Song L., Tian J., Yang B. & Liu S. 2021. A haze prediction model in Chengdu based on LSTM. *Atmosphere* **12**(11): 1479.

Zhang Z., Tian J., Huang W., Yin L., Zheng W. & Liu S. 2021. A Haze prediction method based on one-dimensional convolutional neural network. *Atmosphere* **12**(10): 1327.

Department of Mathematics,
Faculty of Science and Mathematics,
Universiti Pendidikan Sultan Idris
35900, Tanjung Malim, Perak DR, MALAYSIA
E-mail: hazlinadarman@gmail.com, nor.zila@fsmt.upsi.edu.my*

Received: 17 August 2023
Accepted: 31 July 2024

*Corresponding author