MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 1, pp. 264–274 (2025)

# Predictive modeling of haze using chaos theory and deep learning algorithms

Hazlina D.[1], Ho M. K.[1], Nur Hamiza A.[2], Nor Zila A. H.[2], Sarah M.[1]

[1] *Asia Pacific University of Technology & Innovation, Kuala Lumpur, Malaysia*
[2] *University Pendidikan Sultan Idris, Perak, Malaysia*

(Received 19 November 2024; Revised 24 February 2025; Accepted 26 February 2025)

With the swift growth of urbanization and industrialization, fine particulate matter ($PM_{10}$) has escalated into a major global environmental crisis. $PM_{10}$ is often used as a haze indicator, severely affecting human health and ecosystem stability. Accurate prediction of $PM_{10}$ levels is crucial, but existing models face challenges in handling vast data and achieving high accuracy. This study investigates four years of $PM_{10}$ time series in industrial area in Malaysia. Paper aims to develop and compare haze predicting models using chaos theory, including the local linear approximation method (LLAM), local mean approximation method (LMAM); also several deep learning algorithms including convolutional neural network (CNN) and bidirectional LSTM. The performances of these models are evaluated using root mean square error (RMSE), correlation coefficient ($r$), and coefficient of determination ($R^2$). The result shows that Bi-LSTM provides the highest accuracy, LLAM outperforms CNN, while LMAM is reliable but less precise.

**Keywords:** *haze prediction; $PM_{10}$; chaos theory; deep learning.*
**2010 MSC:** 37M10 **DOI:** 10.23939/mmc2025.01.264

## 1. Introduction

Ministry of Health Malaysia defines haze as the degradation of quality of the air caused by the accumulation of fine suspended particles in the atmosphere [1]. These particles are present in greater concentrations, absorbing and scattering sunlight, which impedes vision. There are two main types of sources of haze: natural sources and human activity. While fuel combustion, open burning, agricultural practices, automobile emissions, and wildfires are examples of human activity, natural sources include volcanic eruptions, dust storms, and wildfires [2].

Particulate matter is made up of tiny solid or liquid particles that are small enough to breathe in and pose a major risk to one's health. Particles with a diameter of less than 10 micrometres can enter the respiratory system and the lungs. This risk is caused by fine particles, known as $PM_{2.5}$, with a diameter of less than 2.5 micrometres [3]. According to the Ministry of Health Malaysia [1], the health impacts of haze include throat and nose irritation, reduced lung function, respiratory infections, severe headaches, eye irritation, dizziness, vomiting, and nausea. Haze can negatively impact businesses and industries across a range of industries resulting in decreased productivity and losses from business and tourism, and increased costs for adaptation and mitigation [4]. The cost of haze-related hospital admissions has ranged from MYR 1.8 million to MYR 118.9 million, whereas the extra cost of illness has gone up from MYR 21 million to MYR 410 million [5].

$PM_{10}$ is defined as very small particulate matter that has a diameter of no more than ten micrometres. Compared to other particles, it makes up most of the haze and can originate from numerous causes, including smoke, dust, and vehicular emissions. An increase in $PM_{10}$ concentration in the air signifies higher levels of air pollution. Therefore, this paper will use $PM_{10}$ as a haze predictor. Numerous approaches had been employed to perform haze prediction, including univariate chaos theory [6–9]. Some studies applied several univariate machine learning techniques, such as deep learning, neural

This work was supported by grant from Asia Pacific University of Technology and Innovation (RDIG/04/2023).

network, and long short-term memory, Random Forest, XGBoost [9, 12, 13, 24]. Some studies include meteorological such as $O_3$, CO, $NO_2$, $SO_2$, $PM_{2.5}$ and $PM_{10}$ to predict the haze, namely multiple linear regression in [10, 11, 25] and deep learning techniques as in [14] and [23]. Wu et al. in contrast to this paper, study by Wu et al. [14] included other meteorological parameters and used LSTM to forecast the haze, whereas this paper used BiLSTM method.

This study aims to predict haze based solely on the $PM_{10}$ time series, employing two methods (i) chaos theory, specifically using local mean approximation method (LMAM) and local linear approximation method (LLAM), and (ii) deep learning techniques, namely Bidirectional Long Short Term Memory (BiLSTM) and convolutional neural networks (CNN).

Despite the growing concern over air quality and haze pollution in Malaysia, there remains a significant gap in developing robust predictive models capable of accurately forecasting haze events. Traditional statistical methods and air quality prediction models often fail to capture the complex, non-linear dynamics associated with haze formation and dispersion. These models typically rely on historical data and linear regression techniques, which may not adequately account for the chaotic nature of environmental variables influencing haze, such as meteorological conditions, industrial emissions, and transboundary pollution. Furthermore, there is a lack of comprehensive studies integrating chaos theory and advanced machine learning algorithms specifically tailored to the Malaysian context. Addressing this gap is crucial for improving the accuracy and reliability of haze forecasts, which are essential for public health and environmental management.

This study aims to bridge the identified research gap by developing and evaluating predictive models that leverage the strengths of chaos theory and deep learning models to forecast haze events in Malaysia. By integrating chaos theory-based models (LLAM and LMAM) with deep learning techniques (BiLSTM and CNN), this research seeks to enhance the predictive accuracy and understanding of haze dynamics. The significance of this study lies in its potential to provide more reliable and timely haze predictions, which can inform policy decisions and public health advisories. Improved predictive models can aid in early warning systems, allowing for preventive actions to minimize the negative impacts of haze on health, agriculture, and daily activities. Additionally, this study contributes to the scientific community by offering a novel methodological framework that combines chaos theory and machine learning, setting a precedent for future research in environmental prediction and air quality management.

This paper is structured as follows: Section 1 provides an introduction to the study, includes the research objectives and research gap. Section 2 details the methodology, divided into two subsections: (i) chaos theory approaches specifically the LLAM and LMAM models, and (ii) deep learning models, focusing on BiLSTM and CNN models. Section 3 presents the results and discussion, analyzing the performance of each predictive model. Finally, Section 4 summarizes the findings and implications of the research.

## 2. Methodology

### 2.1. Data pre-processing

In this paper, the hourly $PM_{10}$ time series data spanning from January 2018 to December 2021 from Klang, Selangor will be analyzed, encompassing a total of 35 052 observations. Table 1 shows the statistical analysis of the $PM_{10}$ dataset. The first 31547 data points will be used as the training set, while the subsequent 3 505 data points will serve as the testing set to evaluate the predictive model's performance. During training, the model's parameters are adjusted based on the training data to minimize the error in predictions. Without this data, there would be no basis for optimization. Especially in deep learning models, particularly neural networks, learn to recognize patterns and relationships in data. Training data provides the examples needed for the model to learn these patterns effectively. Also, training data is later used to evaluate the models' performance. By comparing the model's predictions on training data to the actual outcomes, we can assess how well the model is learning.

In the $PM_{10}$ dataset, there are 273 missing data points, which will be imputed using the series mean method, while noise reduction will be performed through spline interpolation method. Subsequently,

the data will be normalized using the min-max normalization method. Min-max normalization is a straightforward technique that rescales data values to a range between 0 and 1 by utilizing the minimum and maximum values of the original dataset. This method keeps the data points' relative order and spacing intact but can reduce variance and amplify the impact of outliers using the formula below,

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{1}$$

**Table 1.** Statistical analysis of hourly $PM_{10}$ concentration.

| Total data | Training size | Testing size | Mean | Standard deviation | Median | Mode | Skewness |
|---|---|---|---|---|---|---|---|
| 35052 | 31547 | 3505 | 37.681 | 27.307 | 32.996 | 37.681 | 13.723 |

### 2.2. Chaos theory

Time series can exhibit two types of dynamics: deterministic and random. Deterministic time series are predictable, whereas random time series are inherently unpredictable. Chaos represents an intermediate dynamic, blending elements of both predictability and randomness. The Cao method introduced by Cao [15] will be applied to determine if a time series exhibits chaotic behavior. One useful technique for examining the chaotic characteristics of nonlinear signals is the phase space reconstruction. The training dataset will be used in reconstructing the phase space and prediction which will be explained in detail in the next section.

#### 2.2.1. The phase space reconstruction

Phase space reconstruction's fundamental concept is to treat the time series as a component produced by a dynamic, nonlinear system. By examining the variation pattern of this component, the equivalent high-dimensional phase space of the system can be reconstructed [16]. A set of time series of one-dimensional data is $X = \{x_1, x_2, \ldots, x_N\}$ for $N$ observations. To reconstruct the phase space, two parameters are required: the time delay $\tau$ and the embedding dimension $m$. The time series $X$ can be reconstructed in a multi-dimensional phase space as shown in the equation below,

$$X_t = \left\{x_t, x_{t+\tau}, x_{t+2\tau}, \ldots, x_{t-(m-1)\tau}\right\}, \tag{2}$$

where $t = 1, 2, \ldots, N - (m-1)\tau$. Employing the time delay $\tau$ and embedding dimension $m$, the reconstructed phase space is presented below,

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_{N-(m-1)\tau} \\ x_{1+\tau} & x_{2+\tau} & x_{3+\tau} & \cdots & x_{N-(m-2)\tau} \\ x_{1+2\tau} & x_{2+2\tau} & x_{3+2\tau} & \cdots & x_{N-(m-3)\tau} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1-(m-1)\tau} & x_{2-(m-1)\tau} & x_{3-(m-1)\tau} & \cdots & x_N \end{bmatrix}. \tag{3}$$

Each column in Eq. (3) represents a vector and can be viewed as a phase point in the phase space, illustrating the mapping relationships among a phase point in the time series and other phase points in the vector. According to Eq. (3), this demonstrates that with $N$ data points, the embedding dimension $m$, and the delay time $\tau$, up to $N - (m-1)\tau$ vectors can be reconstructed.

#### 2.2.2. The time delay $\tau$ and embedding dimension $m$

The selection of embedding dimension $m$ and the time delay $\tau$ is critical during the phase space reconstruction of a time series, as it directly impacts the accuracy of the reconstructed system's chaotic characteristics. Inadequate selection of these parameters can result in reconstruction errors and subsequently lead to lower forecasting accuracy. Despite that, there are no definitive methods for determining the appropriate values of $m$ and $\tau$ [17]. In this paper, the average mutual information method introduced by Fraser and Swinney in [18] will be applied to calculate the value of $\tau$. This method determines the optimal embedding time delay for phase space reconstruction by selecting the time delay at which a local minimum is first reached by the mutual information function. Function $I(T)$ is introduced below, such that the function represents the mutual information function using various values of $\tau$,

where $T$ is the various values of $\tau$,

$$I(T) = \frac{1}{N}\sum_{t=1}^{N} p(x_t, x_{t+\tau}) \cdot \log\left[\frac{p(x_t, x_{t+\tau})}{p(x_t)p(x_{t+\tau})}\right] \tag{4}$$

given that $p(x_t)$ and $p(x_{t+\tau})$ are the probability of $x_t$ and time-shifted $x_{t+\tau}$, and also the $p(x_t, x_{t+\tau})$ is their joint probability. When $I(\tau)$ attains its first local minimum, the $\tau$ value at that point is defined as the optimum time delay.

Another important parameter needed to reconstruct the phase space is the embedding dimension $m$. The Cao method is employed in this study to determine $m$ due to two advantages: (i) it does not depend on the data size, and (ii) uses only parameter time delay $\tau$. The evaluation of the embedding dimension $m$ is based on the rate of change in the distance between nearest neighbors. Based on the Cao method, the distance between the nearest neighbors will converge when the value of $m$ increases, and the dimension at which this convergence occurs should be considered the embedding dimension [15]. The nearest neighbors of $X_t$ in $m$-dimensional phase space, denoted as $X_{n(i-m)}$ are identified. The mean function is described below to estimate the divergence of nearest neighbors,

$$E(m) = \frac{1}{N - m\tau}\sum_{t=1}^{N-m\tau} \frac{\|X_t(m+1) - X_{n(t,m)}(m+1)\|}{\|X_t(m) - X_{n(t,m)}(m)\|} \tag{5}$$

for some $t = 1, 2, \ldots, N - m\tau$ and $\|\cdot\|$ represents the Euclidean distance, while $X_t(m+1) = x_t, x_{t+\tau}, x_{t+2\tau}, \ldots, x_{t-m\tau}$ is the $t$-th reconstructed vector with dimension $m+1$, and $X_{n(t,m)}(m+1)$ is $X_{n(t,m)}(m)$ with dimension $m+1$. To study its variation from $m$ to $m+1$, the variable is defined in Eq. (6), such that when $E_1(m)$ converges at a particular $m_0$, $m_0$ is denoted as the minimum embedding dimension,

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

### 2.2.3. Chaotic behaviour analysis

In addition to selecting the embedding dimension, another function, $E_2(m)$ is defined by the Cao method to differentiate deterministic signals from stochastic ones, aiding in the investigation of whether a time series has chaotic behaviour [6],

$$E_2(m) = \frac{E^*(m+1)}{E^*(m)}, \tag{7}$$

where $E^*$ represents the mean distance between nearest neighbours in the $m$-dimensional space as prescribed

$$E^*(m) = \frac{1}{N - m\tau}\left|\sum_{t=1}^{N-m\tau} X_{t+m\tau} - X_{n(t,m)+m\tau}\right|. \tag{8}$$

For stochastic time series, data characteristics are independent of the dimension $m$. Therefore, $E_2(m) = 1$ for any $m$. Conversely, for deterministic time series, $E_2(m)$ is related to $m$. Thus, there will be instances where $E_2(m) \neq 1$. Additionally, $E_2(m)$ will converge to some extent while $m$ increases.

### 2.2.4. Prediction 1: Local linear approximation method (LLAM)

This study will employ the local linear approximation method (LLAM) and local mean approximation method (LMAM) for the prediction process. In LLAM, the nearest neighbors of $X_t$ need to be identified using the minimum Euclidean distance. Let the nearest neighbour of $X_t$ be $X_p$. Then for some $k$ nearest neighbors, $X_p$ is given by

$$X_p = \{X_{p_1}, X_{p_2}, \ldots, X_{p_k}\}. \tag{9}$$

Most existing research uses trial and error to identify the nearest neighbors. A study by Chen et al. [19] has selected the minimum number of nearest neighbors by calculating the attractor dimension. However, this study solely employs a nonlinear prediction method. Since no specific method is used to determine the value of $k$, this paper will apply $k = 50$. The value of $X_p$ and $X_{p+1}$ are fitted to

the linear equation $X_{p+1} = AX_p + B$. Then, the coefficients of $A$ and $B$ are computed using the least square method given that $X_{p+1} = \{X_{p_1+1}, X_{p_2+1}, \ldots, X_{p_k+1}\}$ and $X_p$ is provided in Eq. (9).

#### 2.2.5. Prediction 2: Local mean approximation method (LMAM)

This method will identify the neighbors of $X_t$ using the same approach described in the preceding section. Not all neighbors are considered for prediction, only $k$-nearest neighbors will be used. The estimation of $X_{t+1}$ is taken as the average of $X_{t'+1}$ as described below, where $X_{t'}$ are the neighbors such that $1 < j < j'$,

$$X_{t+1} = \frac{\sum_{q=1}^{k} X_{t'_q+1}}{k}. \tag{10}$$

After $X_{t'}$ is found, one-hour ahead value $X_{t'+1}$ will be listed.

### 2.3. Deep learning models

#### 2.3.1. Convolutional neural network (CNN)

A convolutional neural network (CNN) is a deep learning that can extract features and patterns from input data, making it effective in tasks, such as predicting continuous outputs from images, time series data, and sensor data. The model comprises multiple layers: convolutional layers, pooling layers, and fully connected layers. In the initial stages of a CNN, a convolutional layer applies a set of filters with a kernel size and stride to the input data. An activation function is used to introduce non-linearity to the extracted features. Meanwhile, a MaxPooling layer, as a type of pooling layer in CNN, reduces spatial dimensions and computational load by downsampling the feature maps while retaining the most important information [20]. This sequence of convolutional and pooling layers recurs to build a deep network that capable of capturing complex features. A one-dimensional vector is created by flattening the output of the final convolutional layer. This vector serves as the input to the fully connected layers, restructuring the spatial features into a suitable format for regression.

The flattened vector is passed through one or more fully connected layers to integrate the extracted features and predict the continuous target value. Typically, each layer contains neurons with specific numbers and an activation function. The final layer, a fully connected layer with a single neuron, outputs the continuous prediction. In this layer, no activation function is applied as it is a regression task requiring a continuous output. Table 2 shows the input parameters used in this study to build the CNN.

#### 2.3.2. Bidirectional long short-term memory (BiLSTM)

A bidirectional Long Short-Term Memory (BiLSTM) network is very good at modeling temporal dependencies in sequential data, that makes them useful for applications like natural language processing, time series prediction, and sensor data analysis. This model architecture typically consists of one or more bidirectional LSTM layers. Each layer processes the input sequence in two directions simultaneously: forward and backward. Usually, each direction comprises $n$ number of LSTM units. The information flow through the network is controlled by an activation function (often Tanh or Sigmoid) in each LSTM unit, and over-fitting can be avoided by applying a dropout rate, deactivating units at random during training. For more complex tasks, additional BiLSTM layers can be stacked. Each subsequent layer receives the concatenated output from the previous BiLSTM layer, effectively building a deeper representation of the temporal features [21].

The output from the final BiLSTM layer is then passed into one or more fully connected or dense layers. These layers integrate the temporal features and prepare them to predict the continuous target value. Each dense layer contains a certain number of neurons and activation functions. Then, the final layer is a fully connected layer with a single neuron that outputs the continuous prediction value. Similar with CNN, no activation function is applied in this layer. Table 3 shows the input parameters used in this study to build the BiLSTM.

Hyper-parameter tuning was applied to optimize the parameters in both BiLSTM and CNN. The hyper-parameters we tuned include learning rate, batch size, and network architecture leads to the

**Table 2.** Summary table of input parameters for CNN.

| Layer | Parameter | Description | Value |
|---|---|---|---|
| Convolutional | Filters | Number of convolutional filters to learn | 32 |
| | Kernel size | Size of convolutional kernel | 2 |
| | Activation | Activation function | Rectified linear unit (ReLU) |
| MaxPooling | Pool size | Size of the pooling window | 2 |
| | Optimiser | Algorithm to update the weights of the model | Adaptive moment estimation (Adam) |
| | Loss function | Measure how well the predictions | Mean square error (MSE) |

**Table 3.** Summary table of input parameters for BiLSTM.

| Layer | Parameter | Description | Value |
|---|---|---|---|
| Bi-LSTM layer 1 | Units | Number of LSTM units to learn | 100 |
| Bi-LSTM layer 2 | Units | Number of LSTM units to learn | 100 |
| MaxPooling | Pool size | Size of the pooling window | 2 |
| | Optimiser | Algorithm to update the weights of the model | Adaptive moment estimation (Adam) |
| | Loss function | Measure how well the predictions | Mean square error (MSE) |

number of layers and units per layer. The results in the paper are the optimal parameter values for the model. This setup is commonly used in deep learning models. Additionally, the parameter values are fined-tuned through hyper-parameter tuning to ensure the models' effectiveness in achieving high accuracy.

### 2.4. Evaluation of predictive models

The performance of the predictive models was rigorously evaluated using three key metrics: Root Mean Square Error (RMSE), Correlation Coefficient (CC), and Coefficient of Determination ($R^2$). These metrics provide a thorough assessment of the models' reliability and accuracy in predicting the $PM_{10}$ time series.

Root Mean Square Error (RMSE) is a standard measure used to quantify the difference between the predicted and observed values. It is calculated as follows,

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)}{N}}, \tag{11}$$

where $y_i$ represents the observed value, $\hat{y}_i$ is the predicted value, and $N$ is the number of observations. A lower RMSE value indicates a higher accuracy of the predictive model, as it signifies a smaller deviation between the predicted and actual values.

The strength and direction of the linear relationship between the observed and predicted values are measured by the correlation coefficient (CC). It is computed using the formula

$$\mathrm{CC} = \frac{\sum_{i=1}^{N} (y_i - \bar{y}_i)\left(y_i - \bar{\hat{y}}_i\right)}{\sqrt{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2 \left(y_i - \bar{\hat{y}}_i\right)^2}}, \tag{12}$$

where $\bar{y}$ and $\bar{\hat{y}}$ are the mean values of the observed and predicted data, respectively. The CC ranges from $-1$ to 1, with a value closer to 1 representing a strong positive linear relationship and higher predictive accuracy.

Coefficient of Determination ($R^2$) computes the variations in the observed data that are predictable from the independent variables. The $R^2$ is given by

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \tag{13}$$

Better predictive accuracy is indicated by an $R^2$ value that is closer to 1, denoting that a greater percentage of the variation in the observed data is explained by the model. In this study, two predictive

models based on chaos theory, namely LLAM and LMAM, and two deep learning models, BiLSTM and CNN, were applied. Their performances were evaluated using the aforementioned metrics. RMSE provided insights into the models' prediction accuracy, with lower values indicating minimal error, the CC highlighted the strength of the linear relationship between predicted and observed values, and $R^2$ illustrated the models' ability to explain the variance in the $PM_{10}$ time series data.

## 3. Results and discussion

### 3.1. Chaotic behaviour, time delay, and embedding dimension

The average mutual information method in Eq. (4) is used to obtain the time delay $\tau$. Figure 1 shows that the first minimum of $\tau$ is eleven, thus we conclude that the optimal value of $\tau = 11$. Using the obtained $\tau$, the graphs of phase space are plotted in two and three dimensions, as shown in Figure 2 and Figure 3. These figures demonstrate the presence of an attractor within the region, despite some outliers in the time series. Sivakumar in [22] stated that the time series seems to be chaotic when a well-defined attractor is present in a low-dimensional phase space, as shown by graphs in Figure 1 and Figure 2. Therefore, it is concluded that the $PM_{10}$ time series has chaotic behavior.

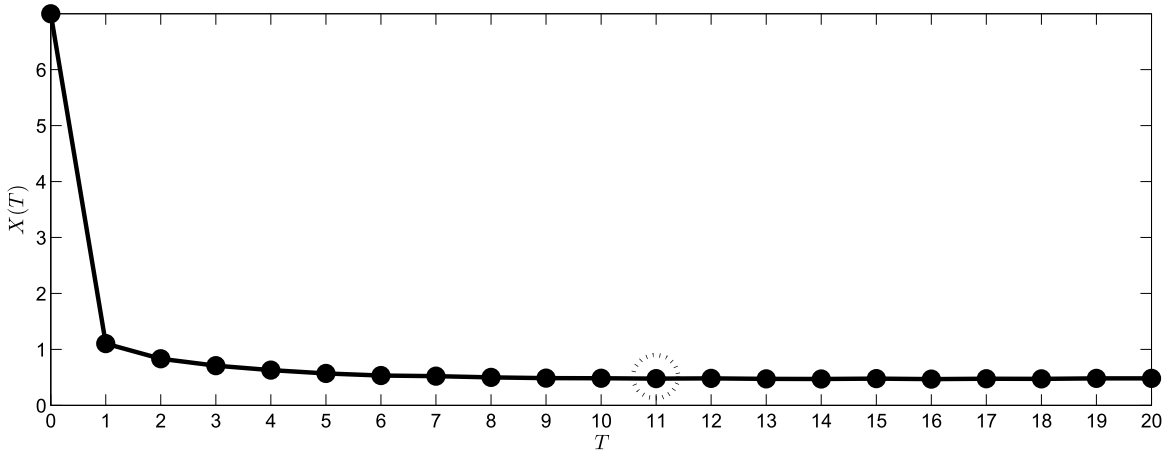


**Fig. 1.** Time delay $\tau$ using the Cao method.

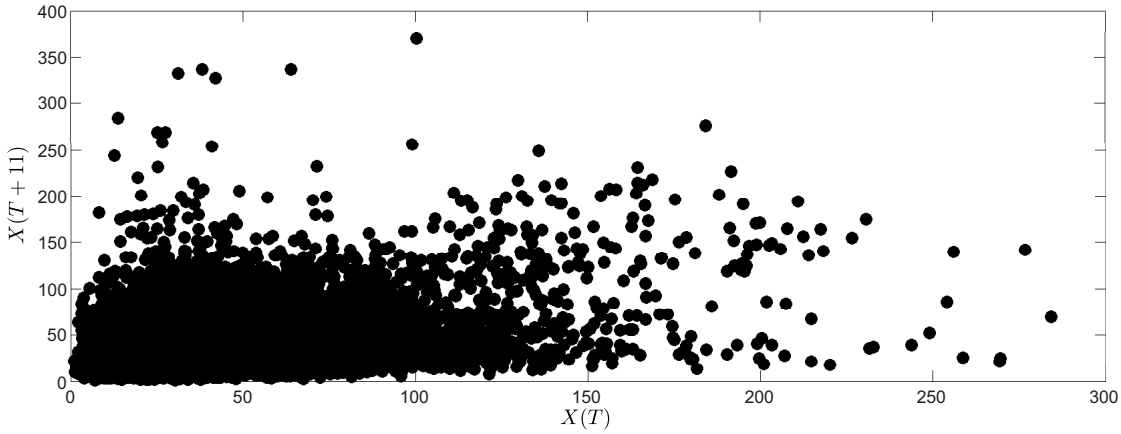


**Fig. 2.** Two dimensions attractor plot.

Figure 4 displays the results of $E_1(m)$ and $E_2(m)$ obtained using the Cao method. $E_1(m)$ begins to saturate when the value of $m$ exceeds $m_0 = 6$. Therefore, $m = m_0 + 1 = 7$ is identified as the minimum embedding dimension. Additionally, if $E_1(m)$ saturates when $m$ increases, it also indicates the presence of chaotic behavior in the time series [15]. Furthermore, if there exists at least one point where $E_1(m) \neq E_2(m)$ as shown in the figure, it confirms the chaotic behavior in the $PM_{10}$ time series. In Figure 4, there is at least one point where $E_2(m) \neq 1$ which guarantees the chaotic behavior in the $PM_{10}$ dataset.

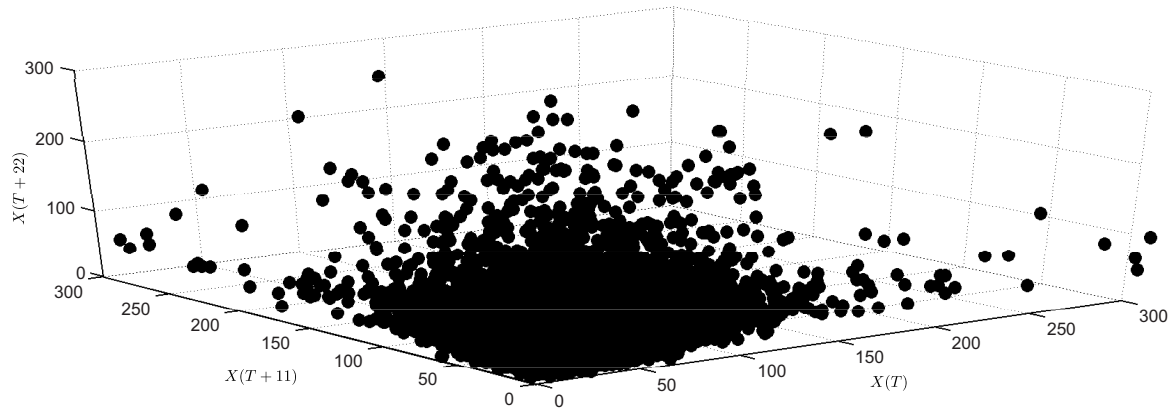


**Fig. 3.** Three dimensions attractor plot.

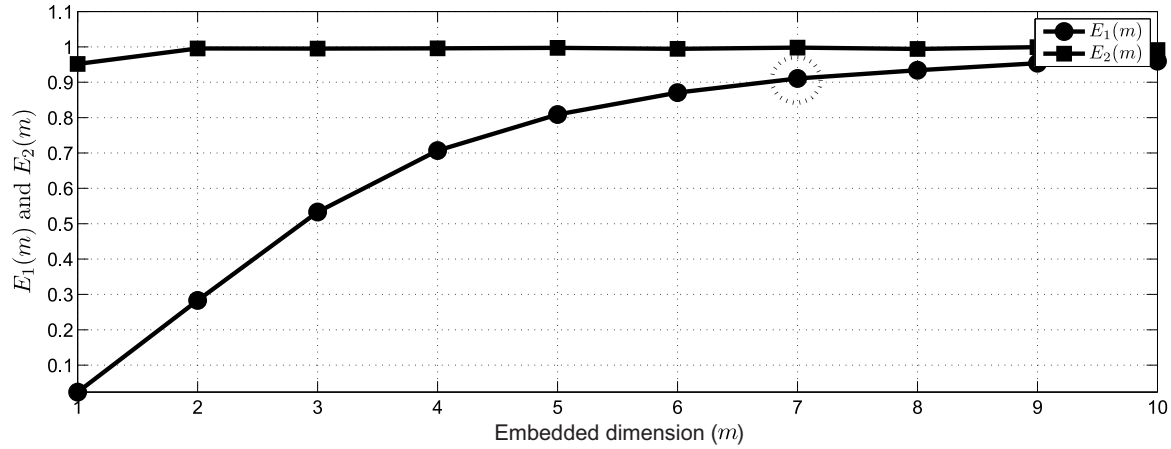


**Fig. 4.** Minimum embedding dimension using the Cao method.

### 3.2. Performance of predictive models

This study applied two predictive models based on chaos theory, namely LLAM and LMAM, as well as two deep learning models, CNN, and BiLSTM. Their performance will be evaluated using RMSE, CC, and $R^2$ as described in Table 4.

**Table 4.** Performance evaluation for the predictive models.

| Predictive models | RMSE | CC | $R^2$ |
|---|---|---|---|
| LLAM | 3.3337 | 0.9649 | 0.9310 |
| LMAM | 4.0612 | 0.9520 | 0.9063 |
| CNN | 3.3693 | 0.9623 | 0.9259 |
| BiLSTM | 3.1230 | 0.9677 | 0.9364 |

## 4. Discussion

LLAM achieves an RMSE of 3.3337, indicating strong predictive accuracy. A high CC of 0.9649 suggests a strong correlation between the predicted and observed values, while the $R^2$ of 0.9310 demonstrates that 93.1% of the variance in the target variable is explained by the model. These results suggest LLAM is highly effective in capturing the underlying dynamics of the $PM_{10}$ time series. LLAM provides robust performance with a balance of accuracy and explainability, leveraging the chaotic nature of the data to capture intricate patterns.

LMAM reports a higher RMSE of 4.0612, indicating slightly lower accuracy compared to LLAM. Its CC of 0.9520 and $R^2$ of 0.9063 further support this, showing that LMAM explains 90.63% of the variance. While still a strong performer, LMAM is marginally less effective than LLAM. Similar to LLAM, LMAM's higher RMSE indicates reduced accuracy, making it less suitable for datasets requiring finer predictive precision.

The CNN model achieves an RMSE of 3.3693, which is slightly higher than LLAM but comparable. The CC of 0.9623 and $R^2$ of 0.9259 suggest that CNN is capable of capturing nonlinear dependencies in the time series data, though it does not outperform LLAM in accuracy or explanatory power. CNN

demonstrates strong predictive power and is adept at handling the spatial features of the time series data, making it a valuable option for more complex tasks. However, the slightly higher RMSE indicates that CNN may not be as suited for data that involves chaotic behavior or requires precise long-term predictions, where chaos-based models excel.

BiLSTM delivers the lowest RMSE (3.1230), indicating superior accuracy compared to the other models. Its high CC (0.9677) and $R^2$ (0.9364) further underscore its ability to capture temporal dependencies in the time series, suggesting that it is particularly effective for sequential data. BiLSTM's ability to model both forward and backward dependencies in the time series results in excellent predictive accuracy, outperforming the other models in RMSE.

The comparison between chaos-based models (LLAM and LMAM) and deep learning models (CNN and BiLSTM) reveals interesting insights. LLAM and LMAM rely on the inherent chaotic nature of the data to reconstruct the phase space, which allows for more effective modeling of the intricate, nonlinear dynamics present in the time series. Despite this, the BiLSTM outperforms LLAM and LMAM in terms of RMSE, suggesting that deep learning architectures, particularly recurrent models like BiLSTM, are highly effective at capturing the temporal patterns and long-range dependencies in the data. However, chaos-based models still offer strong performance, especially LLAM, which surpasses CNN in all metrics. This highlights that, while deep learning models such as BiLSTM excel in accuracy, chaos theory-based models like LLAM are not only competitive but also offer advantages in terms of interpretability and the ability to handle chaotic time series data.

## 5. Conclusion

In summary, the performance evaluation of the predictive models based on RMSE, CC, and $R^2$ indicates that the BiLSTM model is the most effective in predicting $PM_{10}$ levels, followed closely by the LLAM model. Both models demonstrate high accuracy and reliability, with BiLSTM showing a slight edge due to its lower RMSE and higher CC and $R^2$ values. The CNN model also performs well, though not to the extent of BiLSTM and LLAM. The LMAM model, while still capable, lags behind the others in terms of accuracy and explanatory power. The analysis of predictive model performance highlights the superiority of the BiLSTM model in forecasting $PM_{10}$ levels, making it the most suitable choice for applications requiring high precision and reliability. The LLAM model also presents a robust alternative, particularly for scenarios where chaos theory-based approaches are preferred. While the CNN model shows promise, it does not match the effectiveness of the BiLSTM and LLAM models. The LMAM model, despite being the least accurate, still provides a viable option for $PM_{10}$ prediction. Overall, the findings underscore the importance of model selection in environmental forecasting, with BiLSTM emerging as the leading model for predicting particulate matter levels due to its superior performance across all evaluation metrics. The implications of this research are multifaceted, significantly impacting environmental management, public health, and policy-making in Malaysia. By developing predictive models that using chaos theory and machine learning, this study provides a more accurate and reliable framework for forecasting haze events. This enhanced predictive capability allows for the implementation of more effective early warning systems, enabling authorities to issue timely advisories and take proactive measures to protect public health and reduce the economic impacts of haze pollution.

### Acknowledgement

This research was funded by a grant from Asia Pacific University of Technology and Innovation (RDIG/04/2023).

[1] Kementerian Kesihatan Malaysia. Pelan tindakan pengurusan kesihatan akibat jerebu (2020).

[2] Latif M. T., Othman M., Kamin K. H. Fenomena Jerebu di Asia Tenggara: Punca dan Penyelesaian. Institut Alam Sekitar dan Pembangunan. **37**, (2017).

[3] United States Environmental Protection Agency. Particulate Matter (PM) Basics, https://www.epa.gov/pm-pollution/particulate-matter-pm-basics (2024).

[4] Quah E., Chia W.-M., Tan T.-S. Economic impact of 2015 transboundary haze on Singapore. Journal of Asian Economics. **75**, 101329 (2021).

[5] 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., Jing Xiang C., Ahamad F., Amil N., Darus N., Varkkey H., Tangang F., Jaafar A. B. Impact of regional haze towards air quality in Malaysia: A review. Atmospheric Environment. **177**, 28–44 (2018).

[6] Darman H., Abd Hamid N. Z. Predicting Haze Phenomenon using Chaos Theory in Industrial Area in Malaysia. Journal of Quality Measurement and Analysis. **20** (1), 159–169 (2024).

[7] Ul-Saufie A. Z., Yahaya A. S., Ramli N., Hamid H. A. 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 (2012).

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

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

[10] Abdullah S., Ismail M., Yuen Fong S. 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 (2017).

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

[12] Hasmarullzakim A. M., Abdullah A. Haze pattern prediction using deep learning. UTM Computing Proceedings. **3**, 51–55 (2018).

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

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

[15] Cao L. Practical method for determining the minimum embedding dimension of a scalar time series. Physica D: Nonlinear Phenomena. **110** (1–2), 43–50 (1997).

[16] Mei F., Wu Q., Shi T., Lu J., Pan Y., Zheng J. An Ultrashort-Term Net Load Forecasting Model Based on Phase Space Reconstruction and Deep Neural Network. Applied Sciences. **9** (7), 1487 (2019).

[17] Hajiloo R., Salarieh H., Alasty A. Chaos control in delayed phase space constructed by the Takens embedding theory. Communications in Nonlinear Science and Numerical Simulation. **54**, 453–465 (2018).

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

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

[20] Zafar A., Aamir M., Nawi N. M., Arshad A., Riaz S., Alruban A., Dutta A. K., Almotairi S. Comparison of Pooling Methods for Convolutional Neural Networks. Applied Sciences. **12** (17), 8643 (2022).

[21] Raihan A. S., Ahmed I. A Bi-LSTM Autoencoder Framework for Anomaly Detection – A Case Study of a Wind Power Dataset. 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE). 1–6 (2023).

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

[23] Chen J., Liu Z., Yin Z., Liu X., Li X., Yin L., Zheng W. Predict the effect of meteorological factors on haze using BP neural network. Urban Climate. **51**, 101630 (2023).

[24] Liu Z., Liu X., Zhao K. Haze prediction method based on stacking learning. Stochastic Environmental Research and Risk Assessment (2023).

[25] Liu S., Luo Q., Feng M., Zhou L., Qiu Y., Li C., Song D., Tan Q., Yang F. Enhanced nitrate contribution to light extinction during haze pollution in Chengdu: Insights based on an improved multiple linear regression model. Environmental Pollution. **323**, 121309 (2023).

## Прогнозне моделювання туману з використанням теорії хаосу та алгоритмів глибокого навчання

Хазліна Д.[1], Хо М. К.[1], Нур Хаміза А.[2], Нор Зіла А. Н.[2], Сара М.[1]

[1] *Азіатсько-Тихоокеанський університет технологій та інновацій, Куала-Лумпур, Малайзія*
[2] *Університет Пендідікан Султан Ідріс, Перак, Малайзія*

Зі швидким зростанням урбанізації та індустріалізації дрібні тверді частинки ($PM_{10}$) переросли в глобальну екологічну кризу. $PM_{10}$ часто використовується як індикатор туману, що серйозно впливає на здоров'я людини та стабільність екосистеми. Точне передбачення рівнів $PM_{10}$ має вирішальне значення, але існуючі моделі стикаються з проблемами в обробці великої кількості даних і досягненні високої точності. У цій статті досліджуються часові ряди $PM_{10}$ за чотири роки в промисловій зоні Малайзії. Метою цієї статті є розроблення та порівняння моделей прогнозування туману з використанням теорії хаосу, включаючи метод локальної лінійної апроксимації (LLAM), метод локальної середньої апроксимації (LMAM); також декілька алгоритмів глибокого навчання, включаючи згорткову нейронну мережу (CNN) і двонаправлену LSTM. Ефективність цих моделей оцінюється за допомогою середньоквадратичної помилки (RMSE), коефіцієнта кореляції ($r$) і коефіцієнта детермінації ($R^2$). Результат показує, що Bi-LSTM забезпечує найвищу точність, LLAM перевершує CNN, тоді як LMAM надійний, але менш точний.

**Ключові слова:** *передбачення туману; $PM_{10}$; теорія хаосу; глибокі навчання.*