Application of chaotic approach in forecasting highland’s temperature time series

Nor Zila Abd Hamid
Department of Mathematics, Faculty of Science and Mathematics, Sultan Idris Education University, 35900, Tanjong Malim, Perak, Malaysia

nor.zila@fsmt.upsi.edu.my

Abstract. This study focuses on application of chaotic approach in forecasting highland’s temperature time series. Chaotic approach begins by detecting the chaotic behaviour of the temperature time series using phase space plot and Cao method. Then, the local mean approximation method will be used for forecasting purposes. The temperature time series observed at Cameron Highlands is detected as chaotic in behaviour. Due to the chaotic behaviour, the forecasting is done through the local mean approximation method, a chaotic based method. The forecasting result shown that correlation coefficient value between the observed and forecasted time series is 0.9395. Correlation coefficient approaching to one shown that the forecasted result is excellent. This indicates that in particular, the local mean approximation method can be used to forecast the temperature time series in highland area. In general, chaotic approach is a useful approach that can be used to analyse and forecast the temperature time series.

1. Introduction
The physical change of a city is often associated with the construction activities and urban development. Activities such as manufacturing, vehicle exhaust emissions and air conditioning which carried out by humans have resulted in heat production which subsequently contributes to temperature rising. Study by Hashim [1] and Wai [2] discovered that phenomenon of climate changing has negative impacts on Malaysian communities. Recent studies by Lee et al. [3] and Maji et al. [4] found that changes and increasing in temperature can lead to mortality. This situation is particularly worrying since it can contribute to various health problems. Therefore, the forecasting model of temperature time series is very important to be built.

Before the forecasting model can be built, the behavior of the time series must be identified first. From mathematical modeling perspective, the behavior of the time series can be divided into two, deterministic and random. Deterministic time series can be forecasted while random time series are otherwise. Abarbanel [5] described another behavior namely chaotic. Chaotic behavior is between deterministic and random behavior. Sprott [6] stated that the time series with chaotic behavior is sensitive to the initial condition, which reflect that even though the chaotic time series can be forecasted, only short term forecasting is allowed.

In past studies, the time series of temperature is often forecasted using methods such as principal component analysis, multiple linear regression and neural networks. Studies such as [7]–[9] can be referred as examples. All these methods are known as multivariate techniques. Forecasting through the
techniques requires numbers of parameters. In the event that information about required parameters is insufficient, forecasting through multivariate techniques will not be effective. Thus, to overcome this problem, chaotic approach is introduced. Forecasting through chaotic approach is simply done by using available time series, without involving other parameters.

Chaotic approach has been successfully applied to build forecasting model on various variables, world widely. In Malaysia, application of chaotic approach has begun. Among them are studies on ozone modeling ([10]–[14]), river flow modeling [15] and PM$_{10}$ modeling [16]. However, in order to check the effectiveness of chaotic approach in modelling various time series, the research need to be continued. Therefore, in this study, chaotic approach will be applied to forecast the temperature time series. In this study, the effectiveness of the chaotic approach will be tested in modelling the temperature time series at one of Malaysian highland area namely Cameron Highlands. Throughout authors’ knowledge, forecasting of temperature time series in Malaysia using chaotic approach has never been explored. Therefore, this study will be implemented.

The objectives of this study are to: (i) determine the behavior of observed temperature time series through chaotic approach, (ii) calculate the number of variables that affects the temperature time series, (iii) elaborate the factors that affects the temperature time series and (iv) identify the performance of chaotic approach in forecasting the temperature time series.

In order to determine the behavior of observed temperature time series, two methods are chosen namely phase space plot and Cao method. In order to calculate the number of variables that affects the temperature time series, results from Cao method will be applied. Next, to elaborate the factors that affects the temperature time series, results from Cao method and past studies will be combined. In the end, the method of local mean approximation method or simply LMAM will be applied for the purpose of forecasting the temperature time series. The performance of the model will be measured through correlation coefficient value.

2. Materials and Methods
In this study, chaotic approach will be applied to forecast highland’s temperature time series. The following subsection will describe further about the temperature time series data used as well as the chaotic approach.

2.1. Time Series Data
The overall temperature time series data used in this research were observed by the Malaysian Meteorological Department at Cameron Highlands station located in Tanah Rata. Since the researcher is not collecting the data by herself, thus, the data are categorized as secondary data. The temperature time series data were recorded in °C unit, hourly. The temperature data were observed for three months, started from 12 am on 1st June 2016 and ended at 11pm on 31st August, 2016. Therefore, the total number of temperature time series data used is 2208 hours. The time series data are complete. Thus, there is no missing data. Figure 1 is the graph of completed temperature time series data while Table 1 explains statistics description of the temperature time series data.

The station where the temperature time series data were observed is located in Tanah Rata. Tanah Rata is located within the Cameron Highlands area and Cameron Highlands is located in the state of Pahang, Malaysia. The estimated area of Cameron Highlands is approximately 712 km$^2$. Cameron Highlands is one of the beautiful highlands in Malaysia. This area is often be a visitors attraction area, throughout the year. Furthermore, Cameron Highlands is often visited by public since it is one of the coolest area in Malaysia. Thus, the forecasting of temperature in this area is needed.
2.2. Chaotic Approach

Once the time series data was received from Malaysian Meteorological Department, the data are interpreted into a scalar form of:

\[ X = (x_1, x_2, \ldots, x_N) \]  

(1)

\( x_i \) is the data at \( i \)-th hour while \( N \) is the total hours. \( X \) was divided into two; \( X_{\text{train}} \) and \( X_{\text{test}} \). \( X_{\text{train}} \) was the training data which will be used to develop the forecasting model while \( X_{\text{test}} \) was used to measure the effectiveness of the forecasting model. In this study, \( X_{\text{train}} \) is from 1st June until 31st July 2016 while \( X_{\text{test}} \) is from 1st until 31st August 2016. Therefore, \( N = 2208 \) hours, \( X_{\text{train}} = 1464 \) hours and \( X_{\text{test}} = 744 \) hours. Data in (1) were reconstructed into the \( m \)-dimensional phase space of:

\[ Y_j^m = (x_j, x_{j+\tau}, x_{j+2\tau}, \ldots, x_{j+(m-1)\tau}) \]  

(2)

It can be seen that there are two parameters to be determined in Equation 2, which are delay time, \( \tau \) and embedding dimension, \( m \). In this study, \( \tau = 1 \) was used as the data are observed every one hour.
Next, the phase space will be plotted in $\{x, x_{\tau}, x_{2\tau}\}$. Since $\tau = 1$, therefore, the phase space will be plotted in $\{x, x_{\tau}\}$. Referring to some previous studies such as [17], the existence of any attractor indicates that the behaviour of the time series is chaotic.

In order to calculate the $m$ parameter, Cao method [18] is selected. This method is selected since the method does not depend on the number of data and the equation only require the value of $\tau$ parameter to calculate $m$. Cao method [18] involved equation of:

$$E1(m) = \frac{E(m+1)}{E(m)}$$

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

Symbol $\|\|$ is for the maximum norm and $Y_j^m$ is the nearest neighbour to $Y_j^m$. The value of minimum embedding dimension $m$ is selected at the value when $E1(m)$ begins to saturate. For a random time series data, $E1(m)$ will not reach saturation with increasing $m$. For a chaotic time series data, $E1(m)$ will reach saturation with increasing $m$. Therefore, if there exist saturation of $E1(m)$, then the behaviour of the time series is chaotic.

Next, Cao [18] introduced the parameter of $E2(m)$:

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

where

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

According to Cao [18], if all the values of $E2(m)$ are equal to one, then the behaviour of the observed time series is random. On the other hand, if there is at least one value of $E2(m)$ is not equal to one, then the observed time series is chaotic.

Forecasting through the LMAM method is carried out through the basic equation of:

$$Y_j^{m+1} = f(Y_j^m)$$

Study by Jayawardena [19] stated that forecasting $Y_j^{m+1}$ through LMAM is taken as the mean of the $Y_j^{m}$ values:
\[ Y_{j+1}^m = \frac{1}{k} \sum_{q=1}^{k} Y_{j_q}^m \]  

(8)

and \( k \) nearest neighbors \( Y_{j_q}^m \) are selected based on the minimum value of \( \| Y_{j'}^m - Y_{j}^m \| \), where \( j' < j \) and \( \| \cdot \| \) is the Euclidean distance.

In this study, the performance of the developed forecasting model is determined using the correlation coefficient value, \( r \). The value of \( r \) is the reflection of the strength of the relationship between the observed and the forecasted time series. The \( r \) value is between -1 and +1. The closer \( r \) to to -1 or +1, the stronger the relationship between the observed and forecasted time series, which further explains that the observed time series and their forecasted time series are close to each other.

3. Results and Discussion

3.1. Chaotic Behaviour

Before the forecasting model can be developed, the behavior of observed temperature time series must be determined first, and for that, two methods namely space phase plot and Cao method have been chosen. In implementing this step, the reconstruction of phase space is required. For that, the phase space parameters, \( \tau \) and \( m \) of Equation 2 need to be determined first. Using \( \tau = 1 \), the value of \( m \) through the Cao method is searched. Figure 2 shows the results from the Cao method for the time series of temperature at Cameron Highlands. From Figure 2, \( m = 4 \). Therefore, the combination of parameters are \( \tau = 1 \) and \( m = 4 \). This combination will be used to analyse the behaviour of the temperature time series as well as to develop the forecasting model.

The phase space plot of the observed temperature series in Cameron Highlands with \( \tau = 1 \) is as demonstrated in Figure 3. By analysing the plot, it is found that most points are converged to the middle of the phase space. This convergence is known as attractor [17]. The existence of this attractor indicates that the behaviour of the time series of temperatures in Cameron Highlands is chaotic. However, there are some points that are far from attractor. These points are called outliers, which may be caused by noise disturbances.

![Figure 2. Results from Cao method](image_url)
Figure 3. Results from phase space plot

Graph \( m \) against \( E1(m) \) was used to differentiate whether the behaviour of the time series is chaotic or random. For a random time series, the value of \( E1(m) \) will not reach saturation with the increase of \( m \). On the other hand, for a chaotic time series, the value of \( E1(m) \) will be saturated with an increase of \( m \). Based on Figure 2, \( E1(m) \) reaches saturation when \( m \) increase. Thus, the behaviour of Cameron Highlands’ temperature is chaotic. Furthermore, the existence of \( E2(m) \neq 1 \) denotes the presence of chaotic behaviour in the observed time series. As in Figure 2, \( E2(m) \neq 1 \) at \( m = 1 \). Thus, there exist at least one value where \( E2(m) \neq 1 \). Therefore, from \( E2(m) \), it also can be concluded that the behaviour of temperature time series at Cameron Highlands is chaotic. Results from both phase space plot and Cao method indicate that the behaviour of temperature time series at Cameron Highlands is chaotic. Therefore, the development of forecasting model through chaotic approach can be continued.

3.2. Factors Affect the Rise and Fall of Temperature Time Series

In order to investigate factors that affect the temperature series, past studies are reviewed. Research by Hashim [1] found that urbanization, industrialization and house development contributed to the rise and fall of temperature. Furthermore, Adiwijaya et al. [20] found that temperature was affected by environmental factors such as rainfall, solar radiation, wind speed, humidity and air pressure. From Cao method, \( m = 4 \) reflects that at least four factors affect the observed time series of the Cameron Highlands. Combination of findings by Hashim [1] and Adiwijaya et al. [20] proved that more than four factors affect the temperature series. Therefore, result of \( m = 4 \) from Cao method is compatible, consistent and reliable.

3.3. Forecasting Results

Chaotic time series is categorized to be sensitive to initial conditions. Therefore, for time series with chaotic behaviour, only short term forecasting is allowed [6]. Since the above results from phase space plot and Cao method found that the observed temperature time series is chaotic, then a one hour forecasting will be done. The forecasting period for \( X_{test} \) is one month and equivalent to 744 hours. The phase space from Equation 2 is reconstructed and the forecasting model is developed using a combination of parameters \( \tau = 1 \) and \( m = 4 \).

Figure 4 is the forecasting result of temperature time series in Cameron Highlands via the LMAM. From Figure 4, it is clearly seen that the forecasted time series and observed time series are close to each other. Subsequently, trends of rise and fall for both time series are almost equivalent. Thus, the chaotic basic forecasting method of LMAM is good in forecasting the temperature series in Cameron
Highlands. The performance of the forecasting model in this study is measured by calculating the value of correlation coefficient, \( r \). The obtained value is \( r = 0.9395 \). This value approaches one and this shown that temperatures in Cameron Highlands is well forecasted through LMAM.

![Figure 4. Forecasting results through LMAM with parameters \((\tau_1, m_4)\)](image)

4. Conclusion

The objectives of this study are to: (i) determine the behaviour of observed temperature time series through chaotic approach, (ii) calculate the number of variables that affects the temperature time series, (iii) elaborate the factors that affects the temperature time series and (iv) identify the performance of chaotic approach in forecasting the temperature time series.

Thus, from data analysis, it can be concluded that (i) the behaviour of observed temperature time series in Cameron Highlands is chaotic, (ii) the minimum number of variables that affects the temperature time series is four, (iii) factors affecting the rise and fall of temperatures are urbanization, industrialization and house development as well as rainfall, solar radiation, wind speed, humidity and air pressure, (iv) the performance of chaotic approach in forecasting the temperature time series is excellent with \( r = 0.9395 \). It is hoped that the developed forecasting model can help stakeholders in having a better temperature time series management.

5. Implication

Malaysian Academy of Sciences [21] has launched five strategic plans to be achieved by 2020 which are (i) utilize scientific thinking to determine the direction of Science, Technology and Innovation (STI) state, (ii) cultivate the culture of STI excellence, (iii) provide authoritative and timely STI inputs, (iv) promote science applications and applications for people's well-being and (v) facilitate implementation of innovation-based economic strategies. This study has developed mathematical model through chaotic approach for forecasting temperature time series in Malaysia. Thus, it is hoped that this study can help Malaysian Academy of Sciences in achieving their 4th strategic plan.

Malaysian Meteorological Department (MMD) [22] has also launched their own strategic plans to be achieved by 2020 which are: (i) improve the effectiveness of weather services to reduce disaster risk, (ii) strengthen flight meteorological services to ensure flight safety and wellbeing, (iii) empower the earthquake and tsunami services to reduce the risk of earthquake and tsunami, (iv) strengthening climate services for the nation's prosperity and (v) empowering human capital development. The contribution of this study is the development of mathematical model which will assist MMD to forecast and monitor future temperature time series. This model is expected to help MMD in achieving their 1st and 4th strategic plan.

The six strategic core of Ministry of Science Technology and Innovation (MOSTI) [23] are as follows: (i) the main policy of STI is relevant, (ii) creating value through research, development, commercialization and innovation, (iii) effective service delivery system, (iv) practical and STI
culture, (v) optimal success through strategic collaboration and (vi) effective source management. The development of a new mathematical model in this study is expected to help MOSTI’s 2nd and 4th strategic core.

6. Future Research
In future, chaotic approach can be expanded to forecast temperature time series at various areas categories such as lowlands, rural and urban area. In addition, chaotic approach can be applied to others environmental variables such as carbon monoxide, sea level, sea temperature and rainfall.

Acknowledgments
Highly appreciation goes to Malaysian Meteorological Department for supplying the data. Thank you to Sultan Idris Education University for sponsoring this study through research grant with code GPU 2016-0188-102-01.

References
[1] Hashim N M 2010 The impact of global warming trends on urban livability in Malaysia: An analysis Geogr. Online Malaysian J. Soc. Space 6 pp 72–88
[2] Wai N M, Camerlengo A, and Wahab A K A 2005 A study of global warming in Malaysia J. Teknol. 42 pp 1–10
[3] Lee J Y, Lee S H, Hong S C, and Kim H 2017 Projecting future summer mortality due to ambient ozone concentration and temperature changes Atmos. Environ. 156 pp 88–94
[4] Maji S, Ahmed S, Siddiqui W A, and Ghosh S 2017 Short term effects of criteria air pollutants on daily mortality in Delhi, India Atmos. Environ. 150 pp 210–219
[5] Abarbanel H D I 1996 Analysis of Observed Chaotic Data. New York: Springer-Verlag, Inc.
[6] Sprott J C 2003 Chaos and Time-Series Analysis. Oxford University Press
[7] Zounemat-Kermani M 2012 Hourly predictive Levenberg-Marquardt ANN and multi linear regression models for predicting of dew point temperature Meteorol. Atmos. Phys. 117 pp 181–192
[8] Napiorkowski M J, Piotrowski A P, and J. J N 2014 Stream temperature forecasting by means of ensemble of neural networks: Importance of input variables and ensemble size River Flow, pp. 2017–2025
[9] Wicki A and Parlow E 2017 Multiple regression analysis for unmixing of surface temperature data in an urban environment Remote Sens. 9 pp. 1–23
[10] Hamid N Z A, Noorani M S M, Juneng L, and Latif M T 2013 Prediction of ozone concentrations using nonlinear prediction Method AIP Proc. 125 pp. 125–131
[11] Hamid N Z A and Noorani M S M 2013 An improved prediction model of ozone concentration time series based on chaotic approach Int. J. Math. Comput. Sci. Eng. 7 pp. 206–211
[12] Hamid N Z A and Noorani M S M 2017 New improved chaotic approach model application on forecasting ozone concentration time series Sains Malaysiana 46 pp. 1333–1339
[13] Hamid N Z A, Adenan N H, and Noorani M S M 2017 Forecasting and analyzing high O3 time series in educational area through an improved chaotic approach AIP Conference Proceedings 1870 pp. 1–8
[14] Hamid N Z A, Noorani M S M, and Adenan N H 2017 Chaotic analysis and short-term prediction of ozone pollution in Malaysian urban area J. Phys. Conf. Ser. 890 pp. 1–5
[15] Adenan N H, Hamid N Z A, Mohamed Z, and Noorani M S M 2017 A pilot study of river flow prediction in urban area based on phase space reconstruction AIP Conference Proceedings 1870 pp. 1–4.
[16] Hamid N Z A and Noorani M S M 2014 A pilot study using chaotic approach to determine characteristics and forecasting of PM10 concentration time series Sains Malaysiana 43 pp. 475–481
[17] Sivakumar B 2002 A phase-space reconstruction approach to prediction of suspended sediment
concentration in rivers. *J. Hydrol.* **258** pp. 149–162

[18] Cao L 1997 Practical method for determining the minimum embedding dimension of a scalar time series *Phys. D Nonlinear Phenom.* **110** pp. 43–50

[19] Jayawardena A W 1997 Runoff forecasting using a local approximation method **239** pp. 167–171

[20] Adiwijaya, Wisesty U N, and Nhita F 2014 Study of line search techniques on the modified backpropagation for forecasting of weather data in Indonesia **Far East J. Math. Sci.** **86** pp. 139–148

[21] Malaysian Academy of Sciences, “ASM Annual Report 2016,” 2017.

[22] Malaysian Meteorological Department 2018 Strategic Plan Malaysian Meteorological Department (2016-2020) (Review Edition)

[23] Ministry of Science Technology and Innovation 2016 Strategic Plan Ministry of Science, Technology and Innovation (MOSTI) 2016-2020