

Determining the Chaotic Dynamics of Hydrological Data in Flood-Prone Area

Adib Mashuri, Nur Hamiza Adenan*, Nor Zila Abd Hamid

Department of Mathematics, Universiti Pendidikan Sultan Idris, Malaysia

Received July 30, 2019; Revised October 4, 2019; Accepted December 15, 2019

Copyright©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Flood-prone areas are associated with hydrological time series data such as rainfall, water level and river flow. The possibility to predict flood is to relate all the three data involved. However, in order to develop a multivariable prediction model based on chaos approach, each datum needs to identify chaotic dynamics. As such, the Sungai Galas, Dabong in Kelantan, Malaysia which is a flood disaster area has been selected for the analysis. Rainfall, water level and river flow data in this area were collected to be analysed using the Cao method to identify the presence of chaotic dynamics. The hydrological data is uncertain, which is difficult to predict because the data involved is located in the area of flood disaster. The analysis showed the presence of chaotic dynamics on rainfall, water level and river flow data in the Sungai Galas which involved uncertain data located in flood affected areas by using Cao method. Therefore, a multivariable flood prediction model can be implemented using a chaos approach.

Keywords Cao Method, Chaos Approach, Chaotic Dynamic, Flood Area, Hydrological Time Series Data

1. Introduction

This study is to identify the presence of chaotic dynamics of hydrological data located in flood prone areas for the purpose of constructing multivariable flood prediction model based on a chaos approach. The hydrological data involved are rainfall time series data, water level time series data and river flow time series data at Sungai Galas, Dabong in Kelantan, Malaysia. This area is a region where flood happens that causes uncertain hydrological data. The uncertain data refer to the data that are fluctuating as seen in Fig. 1.

There are several methods that can help in identifying the presence of chaotic dynamics i.e.; phase space plot method [1], Cao method [2] and exponent Lyapunov [3].

However, this study only chooses Cao method [4] to identify the dynamics of all the hydrological data involved. This is because the Cao method has successfully identified the presence of chaotic dynamics for hydrological data [5], [6]. In this study, the presence of the most rapidly detected dynamics will be identified by comparing the results of the study on the analysis using the Cao method to the hydrological data involved. Additionally, the study on this chaotic dynamics has recently been implemented in Malaysia. This research only involves river flow [1], ozone [7], suspended particles [8] and rainfall [9] independently. However, for the next study, the only chaotic dynamically identifiable data will be used in further studies to develop a multivariable prediction model based on a gradual approach. The construction of this multivariable prediction model is to predict flood in area of Sungai Galas.

2. Data

Rainfall, water level and river flow time series that are hydrological data from Sungai Galas are used in the analysis. The area has been selected for this study because the area was involved during the worst flood disaster in 2014 [10]. Time series data were taken from 2 stations which were (1) station 5320038 (rainfall observation station) and (2) station 5320443 (water level and river flow station). All data were obtained from the Department of Irrigation and Drainage Malaysia (DID). There were 2068 data being used for analysis, taken from February 1, 2005 to September 30, 2010. Table 1 shows the percentage of missing data. The missing data were determined by using the mean.

Table 1. Percentage of Missing Data

Time Series Data	Percentage of Missing Data
Rainfall	0.87%
Water level	0.00%
River flow	5.66%

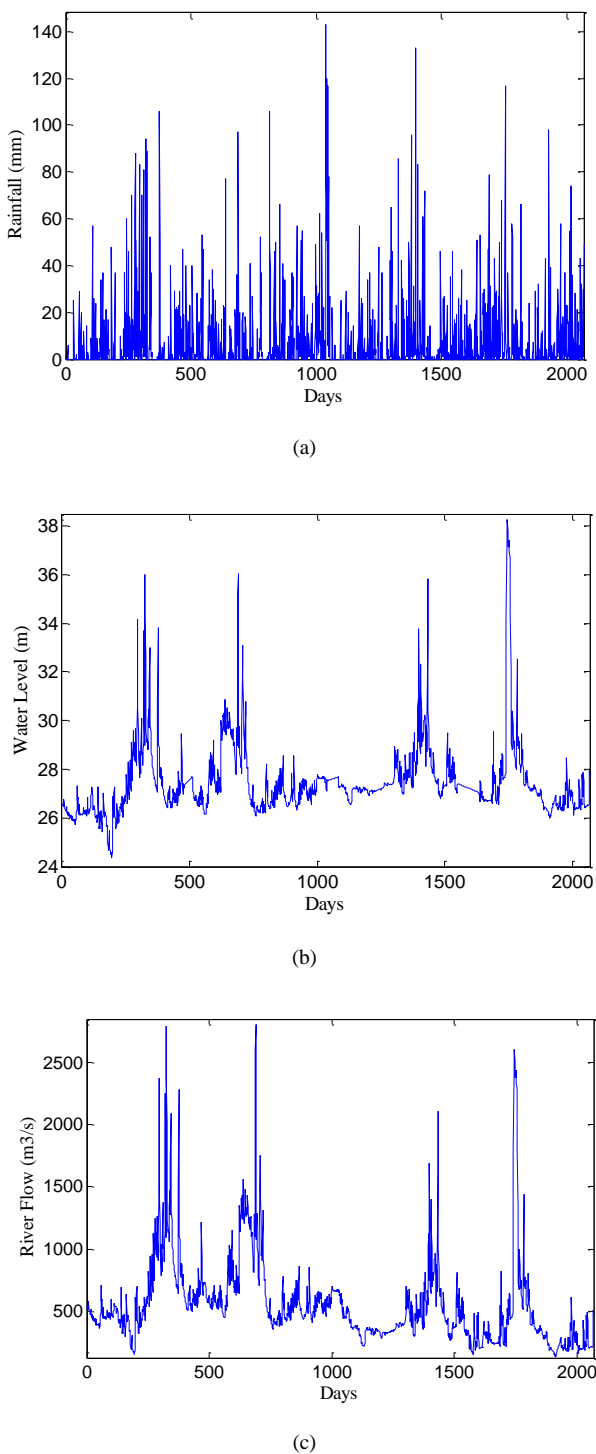


Figure 1. Hydrological data at Sungai Galas at Dabong, Kelantan: (a) rainfall, (b) water level and (c) river flow

Fig. 1 shows the distribution of data involved in this analysis. Referring to Fig. 1 (a), the maximum value of the rain was 140mm and there was also non-rainy day. Meanwhile, Fig. 1 (b) shows that the water level data had a value of over 36m for four times over the period of 2005 to 2010. According to DID, the Sungai Galas level is classified as danger level when it reaches 38m. Next, Fig. 1 (c) shows the flow of Sungai Galas which had the same fluctuation as water level data. If viewed more closely, the river flow reached over $1500 \text{ m}^3 \text{ s}^{-1}$. Hence, uncertain data are involved which are referred to the high and low peaks in each hydrological datum involved.

The dynamics of the data were divided into two types; namely, deterministic and random. The chaotic dynamics data were located between the deterministic and random [11]. The deterministic data moved from the starting point and were in accordance to the rules. Meanwhile, the chaotic dynamics data were seen as being random but actually in accordance to the rules. It can be noted here that the hydrological data involved looked like random but analysis was needed to identify the presence of chaotic dynamics on the hydrological data involved.

3. Methodology

The Cao method [4] was chosen to analyze the presence of chaotic dynamic on the hydrological time series data involved. There are two advantage methods compared to other methods; (1) this method does not depend on the number of parameters required except the delay time (τ) and (2) this method does not depend on the number of data. There are two parameters involved in this method; $Q1(d)$ and $Q2(d)$. Both parameters need to be plotted together with embedding dimension (d) to identify the presence of chaotic dynamics. There are two indications of chaotic dynamics available for the analysis of hydrological data involved using this method. The first indicator is that if $Q1(d)$ is saturated when embedding dimension (d) is increased, then the time series are chaotic dynamics. Whereas, chaotic dynamic is present if $Q2(d) \neq 1$ is present for any embedding dimension (d), which is the second indicator. The basis for the use of this Cao method is the phase space reconstruction. The phase space reconstruction involves each time series datum to be used. All hidrological data involved need to be reconstructed into d – dimensional. Assuming the data is:

$$x_1, x_2, \dots, x_N \quad (1)$$

x_1 refers to the first data while x_2 will be the second data. Next x_N is the data that refer to the last number of data N . Subsequently, the data is reconstructed into d -dimensional as in the following equation:

$$Y_i(d) = (x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau}), i = 1, 2, \dots, N - (d-1)\tau \quad (2)$$

where d is embedding dimension and τ is time delay. The value of d varies while the value τ is from the calculation of Average Mutual Information (AMI) [12]. $Y_i(d)$ is the result of the reconstruction of the vector i th for each embedding dimension. After the phase space reconstruction is performed, the value $Q1(d)$ is calculated by using the equation:

$$Q1(d) = \frac{Q(d+1)}{Q(d)}, \text{ and} \quad (3)$$

$$Q(d) = \frac{1}{N-d\tau} \sum_{n=1}^{N-d\tau} \frac{\|Y_i(d+1) - Y_{j(i,d)}(d+1)\|}{\|Y_i(d) - Y_{j(i,d)}(d)\|} \quad (4)$$

with $\|\cdot\|$ is Euclidian distance between $Y_{j(i,d)}(d+1)$ and the nearest neighbour for $Y_i(d)$. $Y_i(d+1)$ is the result of the phase space reconstruction of vector i th for $d+1$ dimension with :

$$Y_i(d+1) = (x_i, x_{i+\tau}, \dots, x_{i+(d)\tau}). \quad (5)$$

where $j(i,d)$ has value $1 \leq j(i,d) \leq N-d\tau$. In identifying the presence of chaotic dynamics, the graph $Q1(d)$ against d needs to be plotted in order to determine chaotic dynamics. Subsequently, computations of $Q2(d)$ are continued in order to confirm the existence of chaos in the time series data involved. The calculations of $Q2(d)$ are:

$$Q^* = \frac{1}{N-d\tau} \sum_{i=1}^{N-d\tau} |x_{i+d\tau} - x_{j(i,d)+d\tau}| \text{ and} \quad (6)$$

$$Q2(d) = \frac{Q^*(d+1)}{Q^*(d)}. \quad (7)$$

In order to identify the presence of chaotic dynamics, graph $Q2(d)$ against d needs to be plotted.

4. Results and Discussion

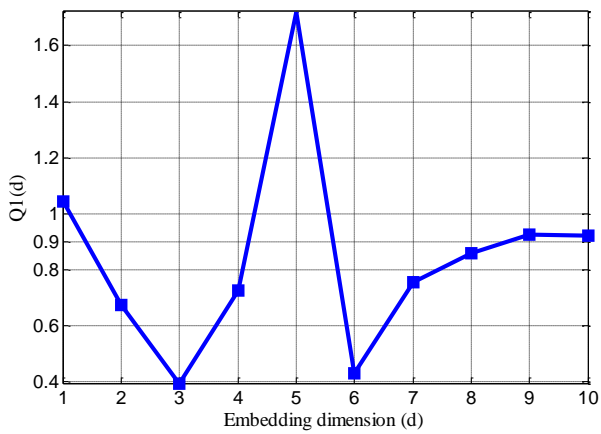
The results of this study are divided into 3 parts; (1) AMI calculation result for determination of time delay τ value, (2) result for $Q1(d)$ indicator, and (3) result for $Q2(d)$ indicator for all three hydrological data involved. Table 2

shows the result of AMI calculation for the determination τ . AMI's first minimum value is the optimal value for τ . Referring to Table 2, the optimum values τ for rainfall data and water levels data are 10 with each having the first minimum values of AMI at 3.897 and 3.347. Meanwhile, AMI's first minimum value of river flow data is 3.419 with $\tau = 7$. All these τ values are used for the phase space reconstruction as in equation (2). Furthermore, the result of phase space reconstruction for each hydrological time series data will be used to calculate the parameters $Q1(d)$ and $Q2(d)$.

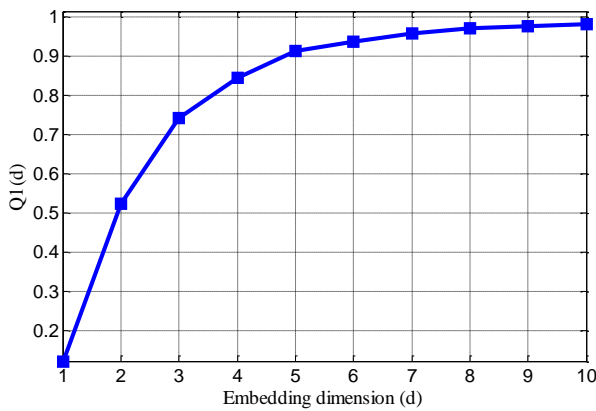
Table 2. Average Mutual Information (AMI)

Time delay (τ)	Average Mutual Information (AMI)		
	Time Series Data		
	Rainfall	Water level	River flow
1	4.263	3.811	3.850
2	4.136	3.586	3.633
3	4.085	3.488	3.515
4	4.015	3.467	3.500
5	3.988	3.447	3.473
6	3.958	3.420	3.436
7	3.907	3.389	3.419
8	3.904	3.350	3.429
9	3.901	3.349	3.384
10	3.897	3.347	3.369
11	3.923	3.354	3.371
12	3.873	3.321	3.337
13	3.880	3.304	3.347
14	3.884	3.282	3.300
15	3.904	3.290	3.318

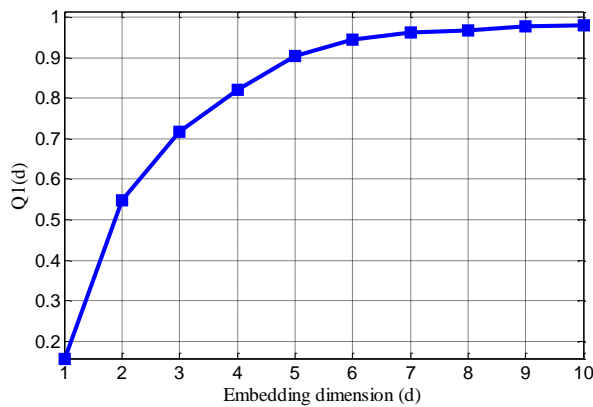
Fig. 2 shows the results for calculations $Q1(d)$ for each hydrological data used. Each calculation $Q1(d)$ is plotted together with the increasing value of d . Referring to the figure $Q1(d)$, the value increases as the value d increases for each hydrological data involved. However, at one stage, the $Q1(d)$ value will reach the saturated value. Fig. 2 (a) shows the $Q1(d)$ value of the rainfall data begins to saturate when the value $d = 9$ which is between 0.9 to 1.0. Although the value $Q1(d)$ fluctuates, it will eventually saturate. Whereas, Fig. 2 (b) and Fig. 2 (c) showed the $Q1(d)$ value starts to saturate when $d = 5$ when approaching the value of 1 for water level and river flow data. Therefore, all the hydrological data analysed are chaotic dynamics with all $Q1(d)$ values continuing to saturate with the increase of d .



(a)



(b)

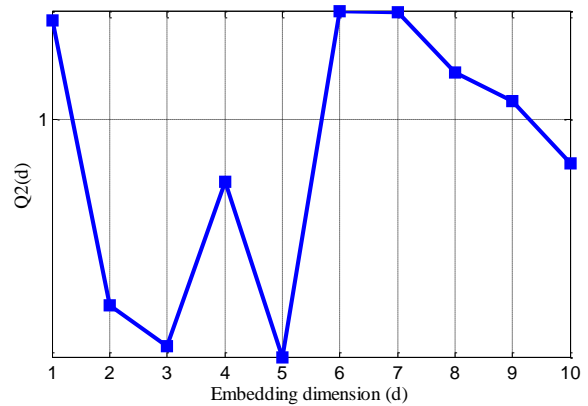


(c)

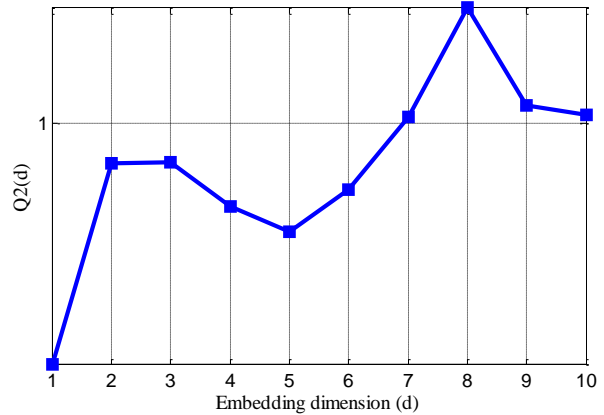
Figure 2. $Q1(d)$ versus embedding dimension d for (a) rainfall, (b) water level and (c) river flow.

Next, Fig. 3 shows the findings for $Q2(d)$ estimation of the analysed hydrological data. All values $Q2(d) \neq 1$ for each d . Hence, it is believed that the analysed hydrological data in Sungai Galas are chaotic dynamics. Although data on the same date used the determination of chaotic existence, the rainfall data can provide chaotic

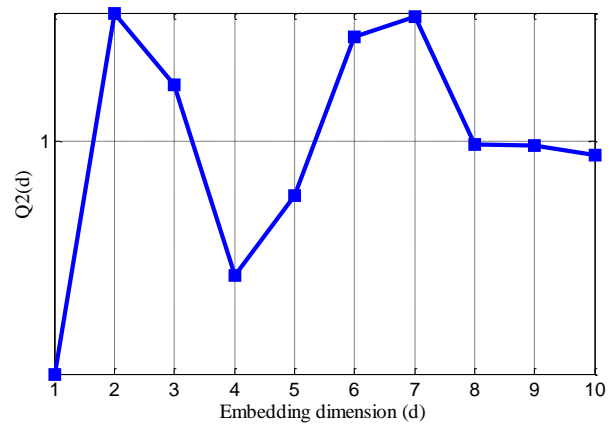
dynamics indication slower than water level data and river flow by using $Q1(d)$.



(a)



(b)



(c)

Figure 3. $Q2(d)$ versus embedding dimension d for (a) rainfall, (b) water level and (c) river flow

Application of Cao method for rainfall time series data was also conducted in Changchun, China [13]. The results showed that both graphs $Q1(d)$ against d and $Q2(d)$ against d gave the same results with the analysis of rainfall

data at Sungai Galas, in which $Q1(d)$ was saturated when d increased and $Q2(d) \neq 1$ for at least one value of d . Thus, the Cao method has successfully identified the presence of chaotic dynamics in rainfall data.

Meanwhile, the graph $Q1(d)$ also continued to be saturated with the increased value of d for data analysis in Sungai Muda [14]. Furthermore, the findings of this study also showed that river flow data was used to provide $Q2(d) \neq 1$ for at least one value of d . Therefore, the findings were in line with the analysis of chaotic dynamics in Sungai Muda [14] where the river flow was chaotic dynamics by using $Q2(d)$.

5. Conclusions

The purpose of analyzing the presence of chaotic dynamics for hydrological time series data which involve rainfall data, water level data and river flow data is to build a multivariable flood prediction model by using a chaos approach. Therefore, the first step before the prediction is to identify the presence of chaotic existence on every hydrological data involved.

Sungai Galas was chosen as a study area as this area was involved during the worst flood disaster in 2014. Therefore, it is seen on the data involved that it is uncertain and always fluctuating. The hydrological time series data are located in the same area of Sungai Galas, Kelantan in Malaysia. Basically, there are two parameters from Cao methods involved which are $Q1(d)$ and $Q2(d)$. For parameter $Q1(d)$ water level data and river flow data show quick respond compared to rainfall data in order to determine the chaotic dynamics. Meanwhile, all hydrological data shows $Q2(d) \neq 1$ for at least one value of d for parameter $Q2(d)$ and this result can be evidence for the presence of chaotic dynamics for all data involved. The results for both analyses showed the presence of chaotic dynamics for rainfall data, water level data and river flow data in Sungai Galas which involved uncertain data as the area was located in flood affected areas using the Cao method. Next, with the intention of construction a multivariable flood prediction model by using a chaos approach can be implemented. Basically, the multivariable flood prediction model is based on phase space reconstruction using all hydrological data involved in this study.

In addition, the implication for this study could contribute to identifying the chaotic dynamics in order to understand the hydrological system in flood area especially in Sungai Galas. By comparing all hydrological data involved, it is proved that Cao method could identify the chaotic dynamics for uncertain data, in flood area and use daily data. Besides that, Cao method could be suggested to be used in other hydrological data such as sediment transport, evaporations and water quality with the purpose of determining the chaotic dynamics in hydrological

system.

Acknowledgements

The authors thankfully acknowledged the financial support provided by the Ministry of Education Malaysia (2019-0009-102-02: FRGS/1/20/2018/STG06/UPSI/02/3) as well as Department of Irrigation and Drainage Malaysia for providing the hydrological data.

REFERENCES

- [1] N. H. Adenan, M. S. M. Noorani. Multiple Time-Scales Nonlinear Prediction of River Flow using Chaos Approach, *Jurnal Teknologi*, Vol. 78, No. 7, 1-7, 2016.
- [2] W. N. A. W. M. Zaim, N. Z. A. Hamid. Peramalan Bahan Pencemar Ozon (O₃) di Universiti Pendidikan Sultan Idris, Tanjung Malim, Perak, Malaysia mengikut Monsun dengan Menggunakan Pendekatan Kalut, *Sains Malaysiana*, Vol. 46, No. 12, 2523–2528, 2017.
- [3] X. Li, J. Sha, Y. Li, Z. L. Wang. Comparison of Hybrid Models for Daily Streamflow Prediction in a Forested Basin, *J. Hydroinformatics*, Vol. 20, No. 1, 191–205, 2018.
- [4] L. Cao. Practical Method for Determining the Minimum Embedding Dimension of a Scalar Time Series, *Phys. D Nonlinear Phenom.*, Vol. 110, No. 1–2, 43–50, 1997.
- [5] D. X. She, X. Yang. A New Adaptive Local Linear Prediction Method and Its Application in Hydrological Time Series, *Math. Probl. Eng.*, Vol. 2010, 1–16, 2010.
- [6] N. H. Adenan, N. Z. A. Hamid, Z. Mohamed, M. S. M. Noorani. A Pilot Study of River Flow Prediction in Urban Area Based on Phase Space Reconstruction, *The 24th National Symposium on Mathematical Sciences (SKSM24)*, vol. 1870, 040011, 2017.
- [7] N. Z. A. Hamid. Application of Chaotic Approach In Forecasting Highland's Temperature Time Series, *IOP Conf. Ser. Earth Environ. Sci.*, Vol. 169, No. 1, 012107, 2018.
- [8] N. Z. A. Hamid, M. S. M. Noorani. Suatu kajian perintis menggunakan pendekatan kalut bagi pengesanan sifat dan peramalan siri masa kepekatan PM₁₀, *Sains Malaysiana*, Vol. 43, No. 3, 475–481, 2014.
- [9] M. L. Sapini, N. S. Adam, N. Ibrahim, N. Rosmen, N. M. Yusof. The Presence of Chaos in Rainfall by using 0-1 Test and Correlation Dimension, *AIP Conference Proceedings*, Vol. 1905, No. 1, p. 050040, 2017.
- [10] Bernama, Terkini: Banjir di Kelantan, Terengganu, Pahang tambah buruk, *Nasional, Berita Harian*, 23 Dec 2014.
- [11] H. D. I Abarbanel. *Analysis of Observed Chaotic Data*. New York: Springer New York, 1996.
- [12] N. Z. A. Hamid, N. H. Adenan, M. S. M. Noorani. Forecasting and Analyzing High O₃ Time Series in

Educational Area through an Improved Chaotic Approach, AIP Conference Proceedings, Vol. 1870, 2017.

- [13] W. L. U. Ouyang, X. Xin, Y. Zhang, W. Cheng, T. Yu. Monthly Rainfall Forecasting using EEMD-SVR Based on Phase-Space Reconstruction, *Water Resour. Manag.*, Vol. 30, No. 7, 2311–2325, 2016.
- [14] N. H. Adenan, M. S. M. Noorani. Peramalan Data Siri Masa Aliran Sungai di Dataran Banjir dengan Menggunakan Pendekatan Kalut. *Sains Malaysiana*, Vol. 44, No. 3, 463–471, 2015.