DATA ANALYTICS AND APPLIED MATHEMATICS (DAAM)
e-ISSN: 2773-4854
VOL. 3, ISSUE 1, 013 – 019
DOI: https://doi.org/10.15282/daam.v3i1.7075

ORIGINAL ARTICLE

The prediction of Malaysian Borneo tide water level using a Chaotic approach

NM Ali[1,*], NZ Abd Hamid[1] and I. Shahida[2]

[1]Department of Mathematics, Faculty of Sciences and Mathematics, Sultan Idris Education University, 35900 Tanjong Malim, Perak, Malaysia.
[2]Marine & Electrical Engineering Technology Section, Malaysian Institute of Marine Engineering Technology, Universiti Kuala Lumpur, 32200 Lumut, Perak

ABSTRACT – Predicting sea level is crucial since the rising of sea levels can cause flood, inundation and coastal erosion. This study investigates the potential of the chaotic model to forecast the future sea level in Malaysian Borneo (Sabah). The studied data is an hourly time series of water level collected at the particular benchmark location (station 74003) in the district of Sandakan, Sabah. This study employed 5136 time series data from June 2017 to November 2017. The research goals are to discover the existence of chaotic dynamics by reconstructing the multi-dimensional phase space and Cao method, and to predict future water level data using the local linear approximation approach. The outcome reveals that the correlation coefficient betwixt actual and forecasted data is 0.9228 which is close to 1. It thus reveals the reliability of the prediction method for forecasting the sea level time series data, as well as encouraging indication that the chaotic approach is appropriate to a time series of Malaysian sea level. In implication, these findings are expected to help agencies, particularly the Malaysian Department of Survey and Mapping (JUPEM), organise better sea-level management.

ARTICLE HISTORY
Received: 31/10/2021
Revised: 14/02/2022
Accepted: 17/02/2022

KEYWORDS
Chaotic approach
Cao method
Local linear approximation
Phase space approach
Sea level prediction

INTRODUCTION

Many academics have stated that sea levels have increased dramatically during the last century and that they are rising even faster currently than in the previous two millenniums. One of the most important environmental problems in the twenty-first century is the possible consequence of accelerating sea level rise [1,2]. Sandakan has the most low-lying land area of the 23 districts in Malaysia's state of Sabah. Future sea level rise is anticipated to wreak havoc on Malaysia's east coast coastal systems, causing inundation, floods, and erosion, notably along the Sandakan coast [3]. The population and amenities of the Sandakan coastal area include a water village settlement, port and harbour facilities, a coastal highway, and major industrial buildings [4]. Thus, the improvement of the sea level prediction models is necessary to enhance the management of the sea level and also evolve and employ ocean-based alternative energy technologies in Sandakan.

In Malaysia, methods such as Autoregressive Integrated Moving Average [5], Regression Vector Support Machine [6], and Trend Analysis[7,8] are employed to predict sea level time series data. Most approaches, however, require some factors that influence sea level rises, such as sea surface temperature, salinity, density, surface air pressure, wind speed, and total cloud cover [5]. Furthermore, the methods used involve long-term data, which poses the potential of data loss. Therefore, an alternative approach is necessary to address this issue. As a result, a chaotic approach is proposed. The chaotic approach analyses and predicts time series data with only one variable and is also reliable for short-term prediction [9].

In recent years, the chaos method gained higher consideration and has been considered for use in mathematical modelling. The chaotic method was successfully applied in modelling sea level [10], ozone [11], particulate matter [12] and river flow [13]. In Malaysia, various studies have been carried out successfully using a chaotic approach, such as time series of ozone[9,14], carbon monoxide[15], temperature [16], and water level [17,18]. So far as the authors are aware, a lack of study on sea level data through the chaotic approach has been done in Malaysia. As a result, this research will help to improve the use of the chaotic approach for mathematical modelling in Malaysia.

In general, chaotic analysis is divided into two particular parts: identifying the chaotic nature of sea level data and forecasting the time series of sea level data. The Cao method and the phase space diagram have been used to identify the presence of chaotic dynamics in sea level data. The prediction model would be fabricated using the chaotic approach once the existence of a chaotic nature was established. To forecast the observed data, the basic technique of the chaotic approach, namely the local linear approximation approach, has been used.

*CORRESPONDING AUTHOR | NM Ali | nazliahbintimohdali@gmail.com

METHODOLOGY

Time series data

The interest area of this study is Sandakan, Sabah tidal gauge station 74003. Figure 1 shows the study region, which is denoted by the green oval.

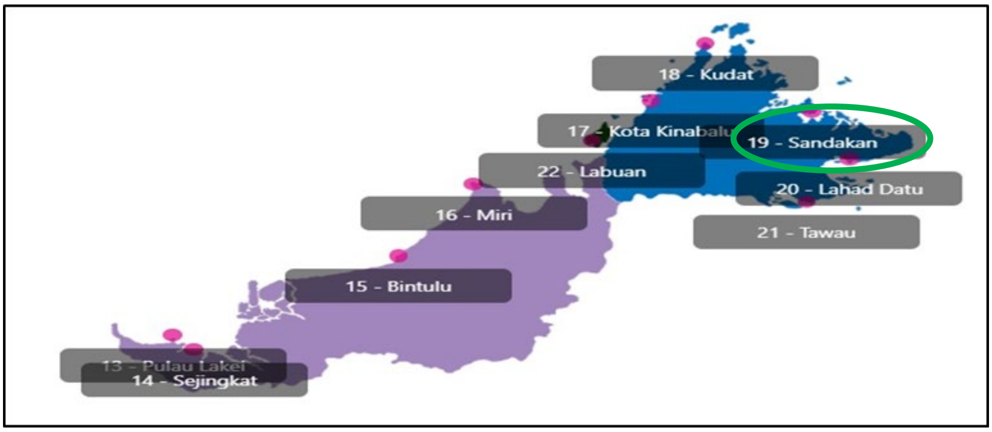


Figure 1. Existing tidal gauge station in East Malaysia.

The time series of sea level data was observed and recorded hourly and were collected over six months, beginning on 1st June 2017 and ending on 31st November 2017. The 5136 data were presented in the scalar form of a one-dimensional vector:

$$X = \{x_1, x_2, x_{3,\dots,} x_{5136}\} \tag{1}$$

Training data was taken from the first 5 months of data, which totalled 4392 records where:

$$X_{training} = \{x_1, x_2, x_{3,\dots\dots} x_{4392}\} \tag{2}$$

Meanwhile, the last 744 data (November),

$$X_{test} = \{x_{4393}, x_{4394}, x_{4395,\dots\dots} x_{5136}\} \tag{3}$$

were preserved to test the prediction performance.

Chaotic approach

The construction of a phase space diagram can reveal the existence of the chaotic dynamics in data. It is depicted by the presence of an attractor on the plot. The phase space is reconstructed by using (2). According to Takens [19], the reconstruction of the phase space is in the form of:

$$Y_i^d = \{x_i, x_{i+\tau}, x_{i+2\tau}, x_{i+3\tau}, \dots, x_{i+(d-\tau)}\} \tag{4}$$

Two parameters i.e. the delay time, τ and the minimum embedding dimension, d must first be determined. The average mutual information method was used in this study to identify the appropriate time delay of the dynamic system meanwhile, the Cao method is used to determine the minimum embedding dimension.

Average Mutual Information method

The Average Mutual Information function (AMI), as defined by Fraser and Swinney [20], is as follows:

$$I(T) = \frac{1}{N}\sum_{a=1}^{N} p(u_a, u_{a+T}) \log_2\left[\frac{p(u_a, u_{a+T})}{p(u_a)p(u_{a+T})}\right] \tag{5}$$

$p(u_a)$ and $p(u_a+T)$ are the probability to obtain u_a and u_a+T respectively on $X_{training}$, whereas $p(u_a,u_a+T)$ is the joint probability of $p(u_a)$ and $p(u_a+T)$. The graph T is plotted against $I(T)$ and τ is the first minimum T that indicates the minimum value $I(T)$.

Cao method

The value d must be identified after the value τ has been determined. According to Regonda et al. [21], d is the bare minimum number of factors needed in order to describe the dynamics of the data. The Cao method was chosen for computing d in this study because, in addition to discovering the parameter d , it can also be used to determine the existence of chaotic nature [22]. The formula is defined as follows:

$$E1(d)=\frac{E(d+1)}{E(d)} \tag{6}$$

where

$$E(d)=\frac{1}{N-d\tau}\sum_{n=1}^{N-d\tau}\frac{\left\|\mathbf{Y}_n^{d+1}-\mathbf{Y}_{jj}^{d+1}\right\|}{\left\|\mathbf{Y}_n^{d}-\mathbf{Y}_{jj}^{d}\right\|}$$

If $E1(d)$ saturates once the value of d is exceeding d_0, thus d_0+1 considered to be the minimum embedding dimension required [14]. Cao [22] also introduces the following equation:

$$E2(d)=\frac{E(d+1)}{E(d)} \tag{7}$$

where

$$E(d)=\frac{1}{N-d\tau}\sum_{n=1}^{N-d\tau}\left|x_{n+d\tau}^{d}-x_{jj+d\tau}^{d}\right|$$

The chaotic dynamics exist in the observational data series if there contains at least one d in which $E2(d)\neq 1$.

Prediction model

In this study, the Local Linear Approximation Model (LLAM) was developed to forecast sea level time series. This model is successfully employed by Domenico et al. [10] and Zakaria et al. [17] in the research on river flow and sea level data, accordingly. For LLAM, the training set (2) is used to form a linear equation $Y_{n+1}=AY_n+B$. The parameters A and B are computed using the least square method. To predict x_{n+1} ,the equation $x_{n+1}=Ax_n+B$ is used with the actual value of x_n . For instance, to predict x_{4310},x_{4309} is used ($x_{4310}=Ax_{4309}+B$).

RESULTS AND DISCUSSION

Chaotic dynamics

To depict the chaotic behaviour of the data, the phase space diagram of $\{x(t),x(t+\tau)\}$ was fabricated using $\tau=3$. Figure 2 reveals the reproduction of Sandakan sea level data applying $\{x(t),x(t+3)\}$. The graph displays the occurrence of an attractor in the phase space diagram. The presence of an attractor proves the existence of chaotic dynamics in time series data [13]. Thus, this implies the time series observed are chaotic.

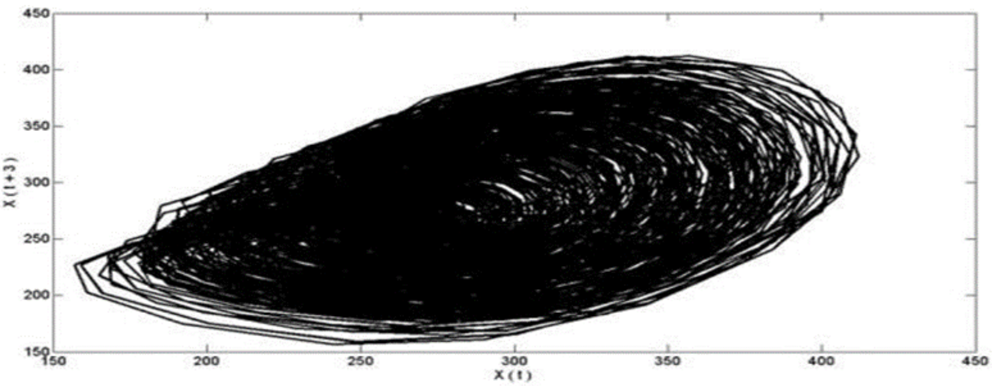


Figure 2. Phase space plot.

Moreover, this finding can be validated by the determination of $E2(d)$ through Cao method. Figure 3 displays the result $E2(d)$ which indicates the existence of $E2(d) \neq 1$. The presence of $E2(d) \neq 1$ demonstrates the existence of a chaotic nature in the empirical time series of sea level data [22].

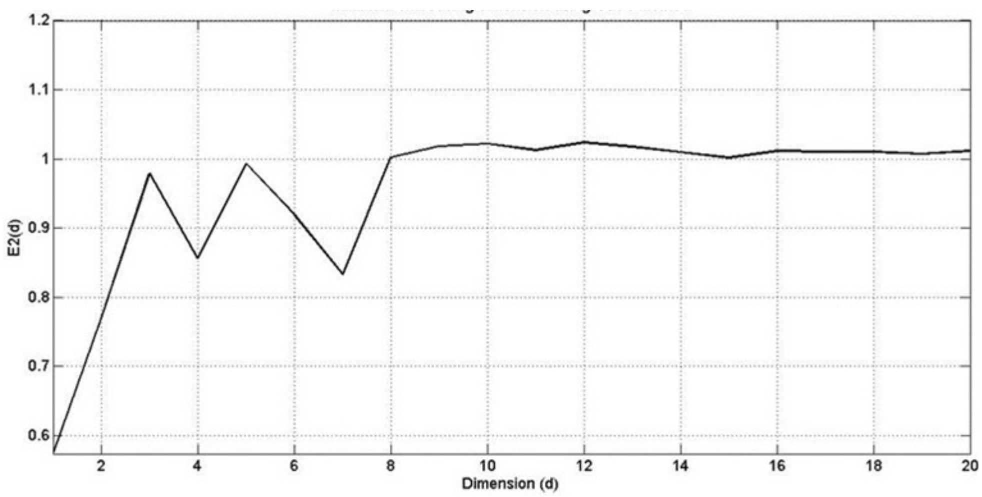


Figure 3. $E2(d)$.

Average Mutual Information

As reported by Siek [23], the AMI approach is an effective method for calculating the value of τ whereby τ is the first minimum T that indicates the minimum value $I(T)$. Figure 4 demonstrates the time delay, $\tau = 3$.

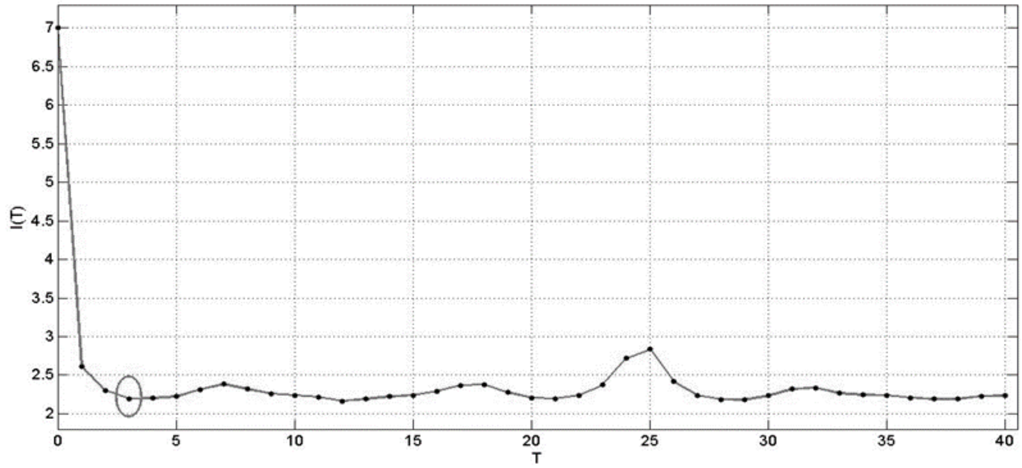
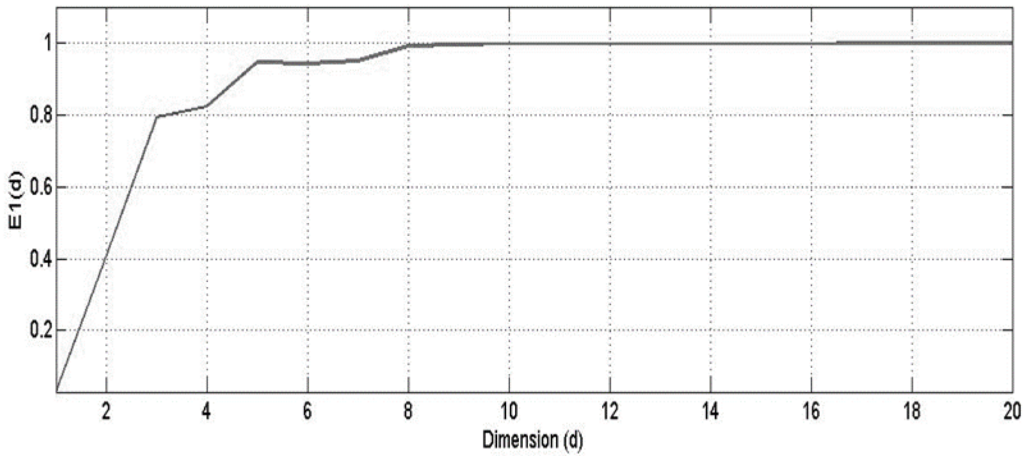


Figure 4. Average Mutual Information result ($\tau = 3$).

Cao method

Figure 5 illustrates the result $E1(d)$. It is observed that after $d_0 = 5$, $E1(d)$ begins to saturate within the value 0.9 to 1.0. Thus, the minimum embedding dimension is $d = 6$ ($d_0 + 1$). As stated by Regonda et al. [21], d is the minimum number of factors that influence the dynamics of a time series. Hence, the findings show that at least six factors influence the observed time series of sea level in Sandakan.



.

Figure 5. Embedding dimension, $E1(d)$.

Prediction result

The LLAM model was fabricated by reconstructing equation (1) with $\tau = 3$ and $d = 6$. This study performs one hour ahead prediction for one month (744 hours). The forecasting results are compared with the observed data (2) to assess the performance of the constructed model. The performance of the prediction model can be verified using the performance indicator, i.e. the coefficient of correlation (cc). Figure 6 depicted the prediction results of the sea level model through the LLAM. Apparently, this implies that the data trend can be predictable. The value of the cc is 0.9228 which is approaching 1. This explains that there exists a very strong relationship between experimental and forecasted values [24]. Hence, these findings prove that the LLAM is reliable and applicable for predicting sea level data.

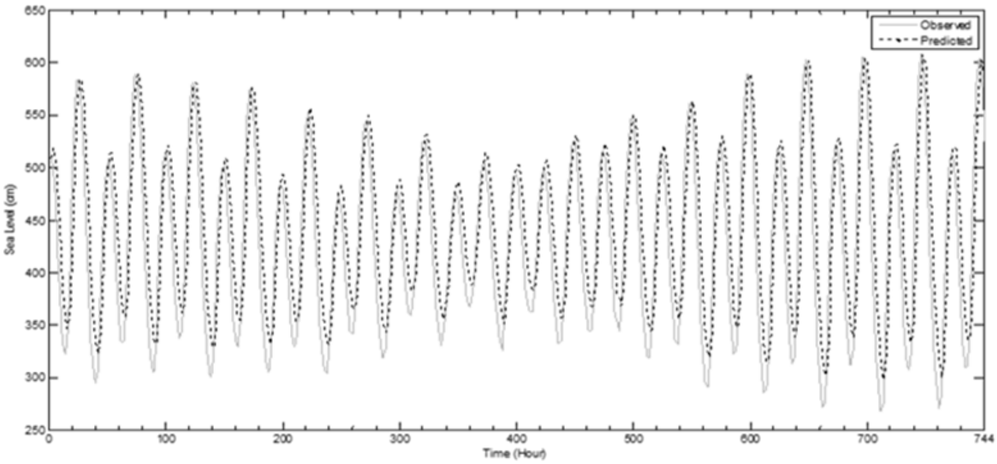


Figure 6. Observed versus predicted.

CONCLUSION

This finding reveals that the chaotic method is a proven method for forecasting the Sandakan sea level data series. The analysis fulfils all of the study's objectives, including identifying the existence of chaotic dynamics in the sea level data series at a particular area and forecasting sea level in a specific area through the chaotic model. The optimal delay time and embedding dimension for phase space reconstruction were defined using average mutual information and the Cao method. The phase space plot and the Cao method are used to detect chaotic behaviour in this study. The chaotic model, i.e. the LLAM has been employed to forecast the sea level data. The coefficient of correlation was used to evaluate the performance of the LLAM. The value of cc=0.9228 proves that the LLAM accurately predicts the observed Sandakan sea level time series. Conclusively, the current approach can theoretically be used for future prediction in a variety of other fields.

The study is expected to help communities along the coast become more aware of devastating disasters such as flood events, coastal erosion, inundation, and casualties. It is expected that in future this study will support agencies, particularly the Malaysian Department of Mapping and Survey (JUPEM) and local entities, in measuring and managing sea levels.

ACKNOWLEDGEMENT

Sincere thanks to the Malaysian Department of Mapping and Survey (JUPEM) for supplying the required sea level time series data.

REFERENCES

[1] R.J. Nicholls and A. Cazenave, “Sea-level rise and its impact on coastal zones,” Science, vol. 328, no. 5985, pp. 1517–1520, 2010.

[2] A. Cazenave and W. Llovel, “Contemporary sea level rise,” Ann. Rev. Mar. Sci., vol. 2, no. 1, pp. 145–173, 2010.

[3] N.A. Awang, A.M. Shah, A. Ahmad, Y.A. Benson and M.R.A. Hamid, “Sea level rise impacts and adaption measures for Sandakan, Sabah,” in Proceedings of the 11th International Conference on Hydroscience & Engineering, 2014, pp. 1017–1026.

[4] D. Anthony, A. Shaaban, T.H. Aung, E.Saleh, R.A. Hamid and A.Osman, “Sea level changes along the Coast of Sandakan Town,” in Proceedings of the 36th IAHR World Congress, Malaysia, 2015.

[5] A.L. Balogun and N. Adebisi, “Sea level prediction using ARIMA, SVR and LSTM neural network: assessing the impact of ensemble ocean-atmospheric processes on models’ accuracy,” Geomat. Nat. Hazards Risk, vol. 12, no. 1, pp. 653–674, 2021.

[6] V. Lai, M. Malek, S. Abdullah, S. Latif and A. Ahmed, “Time-series prediction of sea level change in the East Coast of peninsular Malaysia from the supervised learning approach,” Int. J. Des. Nat. Ecodyn., vol. 15, no. 3, pp. 409–415, 2020.

[7] Q.H. Luu, P. Tkalich and T. W. Tay, “Sea level trend and variability around Peninsular Malaysia,” Ocean Sci., vol. 11, no. 4, pp. 617–628, 2015.

[8] A.B.A. Radzi and H.B. Ismail, “Trend analysis of sea level rise for,” in International Conference on Emerging Trends in Engineering and Technology (ICETET’2013), Malaysia, 2013, pp. 7–8.

[9] N.Z. Abd Hamid, M.S. Md Noorani and N.H. Adenan, “Chaotic analysis and short-term prediction of ozone pollution in Malaysian urban area,” J. Phys. Conf. Ser., vol. 890, pp. 012157, 2017.

[10] M.D. Domenico, M.A. Ghorbani, O. Makarynskyy, D. Makarynska and H. Asadi, “Chaos and reproduction in sea level,” Appl. Math. Model, vol. 37, no. 6, pp. 3687–3697, 2013.

[11] K. Koçak, L. Şaylan and O. Şen, “Nonlinear time series prediction of O3 concentration in Istanbul,” Atmos. Environ., vol. 34, no. 8, pp. 1267–1271, 2000.

[12] A.B. Chelani and S. Devotta, “Nonlinear analysis and prediction of coarse particulate matter concentration in ambient air,” J. Air Waste Manag. Assoc., vol. 56, no. 1, pp. 78–84, 2006.

[13] R. Khatibi, B. Sivakumar, M.A. Ghorbani, O. Kisi, K. Koçak and D. Farsadi Zadeh, “Investigating Chaos in river stage and discharge time series,” J. Hydrol., vol. 414–415, pp. 108–117, 2012.

[14] W.N.A.B. Wan Mohd Zaim and N.Z. Abd Hamid, “Peramalan bahan pencemar ozon (o3) di Universiti Pendidikan Sultan Idris, Tanjung Malim Perak, Malaysia mengikut monsun dengan menggunakan pendekatan Kalut,” Sains Malays., vol. 46, no. 12, pp. 2523–2528, 2017.

[15] K.C. Jusoh, N. Z. Hamid and S. Side, “Forecasting through the Chaotic approach on Carbon Monoxide time series in industrial area,” Journal of Science and Mathematics Letters, vol. 9, pp. 55–62, 2021.

[16] N.Z.A. Hamid et.al, “A pilot study using chaos theory to predict temperature time series in Malaysian semi urban area,” Turk. J. Comput. Math. Educ. (TURCOMAT), vol. 12, no. 3, pp. 997–1003, 2021.

[17] N.H. Zakaria, N.H. Adenan, N.S. Karim and A. Mashuri, “Prediction of water level time series data for dam at Selangor using Chaotic approach and local linear approximation method,” Journal of Science and Mathematics Letters, vol. 9, pp. 10–17, 2021.

[18] A. Mashuri, “Chaotic identification of hourly and daily water level time series data in different areas of elevation at Pahang river,” Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 3, pp. 2427–2435, 2021.

[19] F. Takens, "Detecting strange attractors in turbulence," Dynamical Systems and Turbulence, vol. 898, 1980.

[20] A.M. Fraser and H. L. Swinney, “Independent coordinates for strange attractors from mutual information,” Phys. Rev. A Gen. Phys., vol. 33, no. 2, pp. 1134–1140, 1986.

[21] S. Regonda, B. Rajagopalan, U. Lall, M. Clark, and Y.I. Moon, “Local polynomial method for ensemble forecast of time series,” Nonlinear Process. Geophys., vol. 12, no. 3, pp. 397–406, 2005.

[22] L. Cao, “Practical method for determining the minimum embedding dimension of a scalar time series,” Physica D, vol. 110, no. 1–2, pp. 43–50, 1997.

[23] M. Siek, *Predicting Storm Surges. Chaos, Computational Intelligence, Data Assimilation, Ensembles*. Netherlands, 2011.

[24] P. Schober, C. Boer, and L. A. Schwarte, “Correlation coefficients: appropriate use and interpretation,” Anesth. Analg., vol. 126, no. 5, pp. 1763–1768, 2018.