INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

# A Chaotic Approach for Calculating Minimum Embedding Dimension

**Kamad Nath Shandilya**[1], **Satish Kumar**[2], **Sumita Singh**[3]*

**1** Research Scholar, Department of Physics, Patna University, Patna, Bihar, India
**2** Research Scholar, Department of Physics, Indian Institute of Technology, Patna, Bihar, India
**3** Professor, Department of Physics, Patna University, Patna, Bihar, India

Check for updates

OPEN ACCESS

**Received:** 01-10-2023

**Accepted:** 04-12-2023

**Published:** 05-01-2024

**Citation:** Shandilya KN, Kumar S, Singh S (2024) A Chaotic Approach for Calculating Minimum Embedding Dimension. Indian Journal of Science and Technology 17(1): 38-46. https://doi.org/10.17485/IJST/v17i1.2485

* **Corresponding author**.

**Funding:** None

**Competing Interests:** None



Published By Indian Society for Education and Environment (iSee)

**ISSN**
Print: 0974-6846
Electronic: 0974-5645

## Abstract

**Objective:** The aim of this study is to apply different methods to identify chaos in the temperature time series data of Delhi, India, spanning from 1995 to 2019 during the dry season. The goal is to assess the minimum embedding dimension and, consequently, determine the minimum number of factors influencing the temperature time series data. **Methods:** Lyapunov exponent and correlation dimension calculations have been employed to illustrate the chaotic nature of the data. A plot depicting E2(m) against m has been generated to differentiate data behaviour from stochastic to chaotic. To establish the minimum embedded dimension, an appropriate time lag ($\tau$) has been computed from the plot of Actual Mutual Information (AMI) versus time lag. The Cao method has been utilized to ascertain the minimum embedded dimension. **Findings:** The Lyapunov exponent value is found to be 0.01900, and the correlation dimension value is 0.90860. The positive Lyapunov exponent value and the non-integral correlation dimension value serve as evidence for the existence of chaos in the data. The minimum AMI value occurs at a time lag of 2 days. Utilizing this minimum AMI value and the Cao method, the minimum embedding dimension is determined to be 7. Therefore, the minimum number of parameters influencing the temperature is identified as 7. **Novelty:** Unlike many studies that solely rely on the Lyapunov exponent to detect chaos in time series data and use the reconstruction space method to determine embedding dimension, this study incorporates the Cao method for chaotic analysis of time series data. In addition to Lyapunov exponent and correlation dimension, the Cao method is employed to analyse chaotic behaviour and calculate the minimum embedding dimension, providing a comparative analysis between the two methods.

**Keywords:** Nonlinear dynamics and chaos; Time series data; Lyapunov exponent; Correlation Dimension; State space; Reconstruction space; Embedding space and embedding dimesion

## 1 Introduction

The existing literature has extensively explored chaotic systems, defined by the sensitive dependence on initial conditions in the solution of governing nonlinear differential equations [1]. Chaotic time series analysis has emerged as a popular method for studying these systems, particularly due to its requirement for time series data of only one variable [2].

Notably, E.N. Lorenz in 1963 conducted a study on the convective model of the atmosphere, establishing climate as a chaotic system [3]. Various studies have analyzed chaotic systems through nonlinear differential equations [3], logistic maps [4], and chaotic time series analysis [5–8]. Examples include the work of P. Indira, S.S.R Ibanathan, R.S. Selvaraj, and A.A. Suresh [5] on daily maximum temperature in Chennai, India, and A.T. Adewole, E.O. Falayi, T.O. Roy-Layinde, and A.D. Adelaja [6] analyzing air temperature, relative humidity, and wind speed in selected Nigerian stations.

Additionally, N.Z.A. Hamid, N.H. Adenan, N.B.A. Wahid, S.H.M. Saleh, and B. Bidin [7] utilized phase space reconstruction for chaotic analysis of temperature time series during the dry season in Shah Alam, Malaysia. M. Bahari and N.Z.A. Hamid [8] employed the Cao method to calculate the minimum embedding dimension for temperature time series data in Jerantut, Pahang, Malaysia.

Despite the abundance of studies affirming climate as a chaotic system, there is a notable gap in the literature regarding a comparative analysis of the various methods employed. The current study aims to address this gap by presenting a comprehensive investigation that incorporates Lyapunov exponent, correlation dimension, actual mutual information, and Cao method for analyzing chaotic behavior in atmospheric temperature time series data.

In subsection 3.1 we describe the calculation of the Lyapunov exponent as a measure of the divergence of trajectories in phase space, with a positive value indicating the presence of chaos. In subsection 3.2 the calculation of the correlation dimension is discussed, serving as the dimension of the phase space attractor, where a nonintegral value is considered evidence of chaos. subsection 3.3 elaborates on the use of actual mutual information (AMI) to determine the proper time lag for establishing the minimum embedding dimension. Cao method application is described in section 3.4, this subsection distinguishes deterministic data from stochastic data. Finally, subsection 3.5 details the calculation of *E1(m)* to determine the minimum number of factors influencing the temperature time series data.

By integrating these methods, the study aims to provide a more holistic understanding of chaotic behavior in atmospheric temperature, offering insights into the minimum factors influencing the variable.

### 1.1 Research Gap

Existing studies have primarily utilized either the Lyapunov exponent combined with correlation dimension or the Cao method to quantify chaos. However, a noticeable gap exists in the literature, where a comprehensive approach involving both the Cao method and traditional methods, such as the Lyapunov exponent and correlation dimension, is lacking. This study addresses this gap by incorporating both the Lyapunov exponent and correlation dimension, alongside the Cao method, as recommended in [7] for future research.

Furthermore, a common trend in previous research involves the use of the reconstruction space method for determining the embedding dimension. In contrast, this study employs the Cao method to ascertain the minimum embedding dimension, providing a departure from the conventional approach. This deviation from the norm is essential for a more comprehensive understanding of chaotic behavior in the context of time series data.

### 1.2 Objectives

The main objectives of the study

- Demonstrate the presence of chaos in the temperature time series data of Delhi, India, spanning from 1995 to 2019 during the dry season.
- Assess the minimum embedding dimension and, consequently, determine the minimum number of factors influencing the temperature time series data.

## 2 Temperature time series data

The temperature time series data utilized in this study was obtained from the secondary source https://academic.udayton.edu/kissock/http/Weather/gsod95-current/INDELHI.txt. The data spans the period from 1995 to 2019, focusing on the dry season, specifically from April to June.

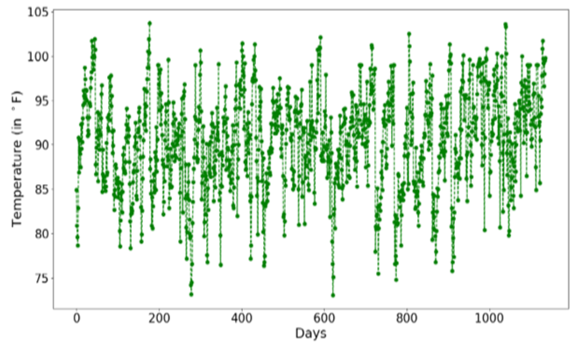


**Fig 1. Temperature vs days**

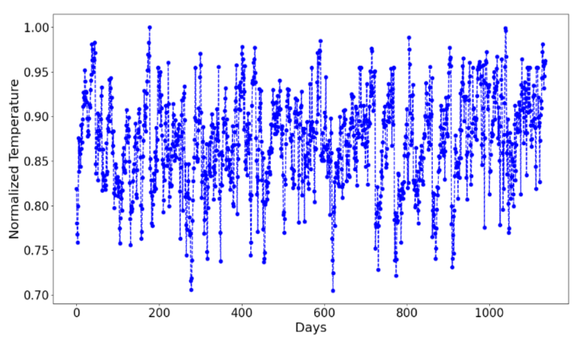


**Fig 2. Normalised temperature vs days**

# 3 Methodology

The identification of chaotic behavior in the time series data was based on the presence of a positive Lyapunov exponent, a non-integral value of correlation dimension[2,9], and the computation of the parameter *E2(m)* through the application of the Cao method. The determination of the minimum embedded dimension and, consequently, the smallest number of parameters influencing the time series variable was achieved using the Cao method[10].

## 3.1 Lyapunov exponent

The governing equations of a nonlinear dynamical system are inherently nonlinear, leading to solutions that exhibit high sensitivity to initial conditions. The trajectories of the system, even when initiated from closely positioned points in state space, undergo exponential divergence with respect to time. The measure of this divergence is quantified by the Lyapunov exponent. In a chaotic system, if two trajectories commence from points like $x_i$ and $x_j$, the separation between them diverges exponentially after n iterations in discrete time units.

Let $s_0 = \left|x_i - x_j\right|$ and after n iteration $s_n = \left|x_{i+n} - x_{j+n}\right|$ then

$$s_n = s_0 e^{\lambda n} \tag{3.1.1}$$

where $\lambda$ is the Lyapunov exponent. So, for positive value of $\lambda$ the separation between points rises exponentially, that is, two nearby trajectories diverge with respect to time.

From Equation (3.1.1), applying logarithm on both sides,

$$ln \frac{s_n}{s_0} = \lambda n \tag{3.1.2}$$

The slope of plot, $ln \frac{s_n}{s_0}$ vs $n$ drawn in Figure 5 using Equation (3.1.2), gives Lyapunov exponent $\lambda$.[2]

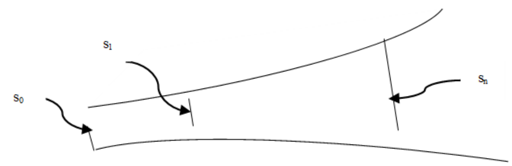


**Fig 3. Schematic view of divergence of trajectories**

## 3.2 Correlation Dimension

Correlation dimension is a method employed to calculate the fractal dimension of an attractor. Figure 4 visually illustrates an attractor situated in a two-dimensional state space. In this representation, dots symbolize the values of time series data for a single variable positioned on the attractor. To compute the correlation dimension, a circle of radius Ris selected such that $R > min.\left|x_i - x_j\right|$, where $x_i$ and $x_j$are two arbitrarily selected data points on the attractor.

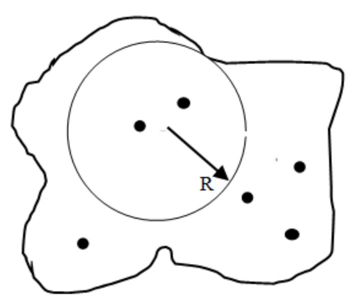


**Fig 4. Schematic view of two-dimensional phase space attractor**

Let the circle is centered on an arbitrarily selected i$^{th}$ point and $N_i$ are the number of points lying within the circle except i$^{th}$ point. Relative number of points on the attractor,

$$p_i(R) = \frac{N_i}{N-1} \tag{3.2.1}$$

Now correlation sum is quantified as

$$C(R) = \frac{1}{N}\sum_{i=1}^{N} p_i(R) \tag{3.2.2}$$

Or,

$$C(R) = \frac{1}{N}\sum_{i=1}^{N} \frac{N_i}{N-1} \tag{3.2.3}$$

If radius $R$ is chosen such that all the points are within the attractor, then

$$\textstyle\sum_{i=1}^{N} N_i = N(N-1) \tag{3.2.4}$$

Now, from Equation (3.2.5)

$$C(R) = 1 \tag{3.2.5}$$

The correlation sum, as derived from Equation (3.2.5), can be computed when the count of points within the circle is known. Due to the potential impracticality of manually counting points, an adjustment has been made to the expression of the correlation sum. This modification involves the introduction of the **Heaviside unit step function ,**

$$\Theta\left(R\left|x_i - x_j\right|\right) = 1,\ R > \left|x_i - x_j\right|$$

$$= 0,\ R < \left|x_i - x_j\right| \tag{3.2.6}$$

where, $x_i$ and $x_j$ are two arbitrary points within the circle.

Replacing $\sum_{i=1}^{N} N_i$ by $\sum_{i=1}^{N}\sum_{i\neq j,j=1}^{N}\Theta(R - \left|x_i - x_j\right|)$, Equation (3.2.3) becomes,

$$C(R) = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{i\neq j,j=1}^{N}\Theta(R - \left|x_i - x_j\right|) \tag{3.2.7}$$

Advantage of Equation (3.2.7) is, instead of counting the points lying within the circle, we only need the values of time series data lying on the attractor.

Now, correlation dimension $D_C$ is defined as,

$$C(R) = kR^{D_C} \tag{3.2.8}$$

Calculating $C(R)$ from Equation (3.2.8) and putting it into Equation (3.2.9) and then taking logarithm on both sides,

$$lnC(R) = lnk + D_C lnR \tag{3.2.9}$$

Plotting $ln\,C(R)$ with respect to $lnR$ as shown in Figure 6, the slope gives the correlation dimension.[2,9]

### 3.3 Actual Mutual Information (AMI)

To ensure the independence of each component in the constructed vectors, a suitable time delay is chosen. In this study, we adopted the method of mutual information to determine the appropriate time delay ($\tau$). Utilizing the histogram of the probability distribution of the data, let the probability that the data assumes a value inside the i$^{\text{th}}$ bin of the histogram is $p_i$ , probability that $x(t)$ is in i$^{\text{th}}$ bin and $x(t+\tau)$ in the j$^{\text{th}}$bin is $p_{ij}$, then the mutual information $I(\tau)$ for time delay $\tau$ is calculated as follows:

$$I(\tau) = \textstyle\sum_{i,j} p_{ij}(\tau)\, lnp_{ij}(\tau) - 2\sum_i p_i lnp_i \tag{3.3.1}$$

A plot of $I(\tau) \sim \tau$ has been drawn in Figure 7 and the value of $\tau$ is determined where $I(\tau)$ is minimum.[9]

### 3.4 Distinguishing data from random to chaotic

For random data, future and past values are independent and then $E2(m)$ is equal to 1 for any $m$. For chaotic data $E2(m)$ is certainly related to $m$, so, it cannot be constant for all $m$. There must be some $m$'s such that $E2(m) \neq 1$. To calculate $E2(m)$, we are using Cao method here.[8]

First, $E*(m)$ is calculated as following,

$$E*(m) = \frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}\left|x_{i+m\tau} - x_{n(i,m)+m\tau}\right| \tag{3.4.1}$$

Then,

$$E2(m) = \frac{E*(m+1)}{E*(m)} \tag{3.4.2}$$

Now the plot of $E2(m) \sim m$ has been drawn in Figure 8 and it is noticed that whether $E2(m) = 1$ or not. If $E2(m) \neq 1$ for at least single value of $m$ then the time series data is chaotic.[8,10]

### 3.5 Embedding Dimension

In this methodology, vectors are formulated from the time series data corresponding to a single variable within a system. The space where these vectors exist is termed the reconstruction space or embedding space, with its dimension referred to as the embedding dimension[2]. By determining the minimum embedding dimension, it becomes possible to identify the minimal number of parameters influencing the given time series data of the variable and, consequently, the system[7]. For this purpose, the Cao method is employed in this study[10].

Let $x_1, x_2, x_3, x_4 \quad \ldots\ldots$ $x_N$ are time series data of a variable of a system.

Vectors are constructed as,

$$y_i(m) = x_i, x_{i+\tau}, x_{i+2\tau}, x_{i+3\tau}\ldots\ldots x_{i+(m-1)\tau}); \tag{3.5.1}$$

where $m$ is the embedding dimension and $\tau$ is the time delay.

Parameter *a(i,m)* is calculated as,

$$a(i,m) = \frac{\left\| y_i\ (m+1) - y_{n(i,d)}\ (m+1) \right\|}{\left\| y_i\ (m) - y_{n(i,m)}\ (m) \right\|},\ i = 1, 2, \ldots\ldots, N - m\tau \tag{3.5.2}$$

$y_i\,(m+1)$ is the i^th^ reconstructed vector with embedding dimension $m$+1 i.e.

$$y_i\,(m+1) = \ (x_i, x_{i+\tau}, x_{i+2\tau}, x_{i+3\tau}\ldots\ldots x_{i+m\tau}); \tag{3.5.3}$$

$n\,(i,m)$is an integer, $1 \leq n\,(i,m) \leq N - m\tau$ and $y_{n(i,m)}\,(m)$ is the nearest neighbour of $y_i\ (m) \| y_k\ (m) - y_l\ (m) \|$ is the Euclidean distance in embedding space.

Now another parameter,$E\ (m)$is calculated as

$$E\ (m) = \frac{1}{N - m\tau} \sum_{i=1}^{N - m\tau} a(i,m) \tag{3.5.4}$$

Again, a parameter $E1\,(m)$ is calculated as,

$$E1\,(m) = \frac{E(m+1)}{E(m)}. \tag{3.5.5}$$

A plot of $E_1 \sim m$has been drawn in Figure 9 and the value of $m$ is determined when $E_1(m)$ saturates. If it saturates at $m_0$ then minimum embedding dimension will be $(m_0+1)$.[8,10]

## 4 Results and Discussion

Using Equation (3.1.2), thirty Lyapunov exponent values were computed with different initial values$x_i$ and $x_j$and varying lags. The largest positive Lyapunov exponent, determined as the slope of the plot in Figure 5, is $\lambda$=0.01900. Additionally, using Equation (3.2.9), the non-integral value of the correlation dimension, calculated from the slope of the plot in Figure 6, is 0.90860. These findings serve as compelling evidence indicating the presence of chaos in the temperature time series data. Reference[2] employed the same method to calculate Lyapunov exponent and correlation dimension, obtaining similar plots for comparison.

In Figure 6, it is evident that the value of C(R) becomes 1 for all R values greater than $|x_i - x_j|$, resulting in the saturation of the graph. Thus, Lyapunov exponent and correlation dimension, calculated from the plots in Figure 5 and in Figure 6, provide clear indications of the presence of chaos in the time series data of temperature. Furthermore, the values of E2(m), depicted as less than 1 in Figure 8, serve as additional evidence supporting the chaotic nature of the data.

The plot in Figure 7 has been drawn using Equation (3.3.1). It is observed that mutual information achieves its first minimum at $\tau$=2. According to reference[9], the appropriate time lag is selected when mutual information reaches its initial minimum. This proper choice of time lag enhances the determination of E2(m) and E1(m) with greater elegance. Taking $\tau$=2 and utilizing Equation (3.4.2), a plot of E2(m) against m has been depicted in Figure 8.

In the depicted plot, Figure 8, there exist values of E2(m) that are not equal to 1. This observation serves as proof that the time series data is indeed chaotic. Reference[8] and[10] employed the same methodology, resulting in similar plots for E2(m).

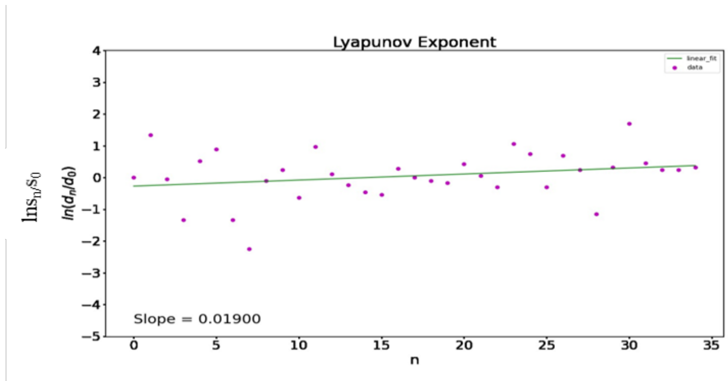


**Fig 5. P lot between $ln\frac{s_n}{s_0}$ and $n$.**

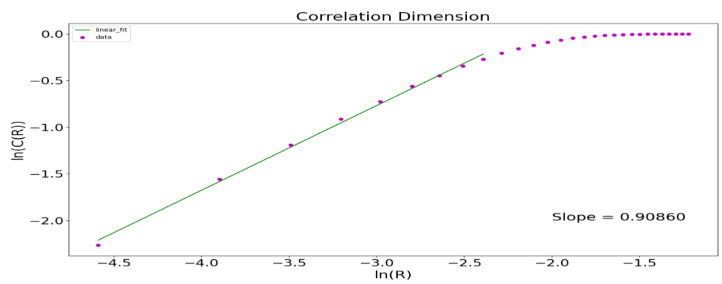


**Fig 6. Plot between$lnC(R)$ and $lnR$.**

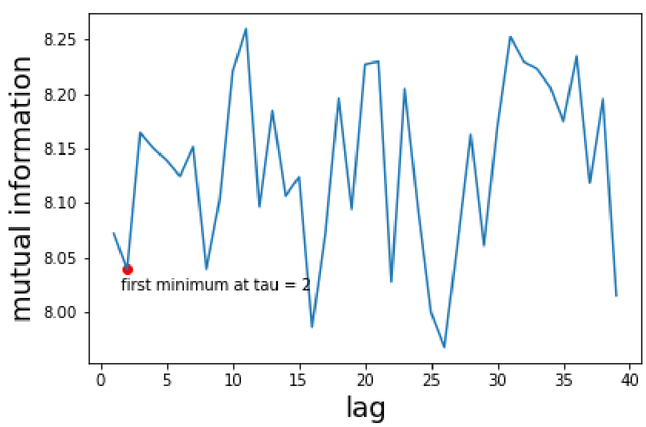


**Fig 7. Plot between mutual information and lag (days)**

By setting $\tau$=2 and utilizing Equation (3.5.5), the subsequent plot, illustrated in Figure 9, represents E1(m) as a function of m.

In Figure 9, it is observed that E1(m) ceases to vary after m=6, leading to the determination of the minimum embedding dimension as 7. This refers that the least number of parameters capable of influencing the temperature is 7. References [8] and [10] employed a similar method for plotting E1(m) and obtained comparable results.

The Cao method, utilized in this study to ascertain the minimum embedding dimension, offers several advantages. Notably, it is independent of the quantity of available data, effectively distinguishes chaotic data from stochastic data, and is computationally efficient. Moreover, the Cao method does not rely on any subjective parameters except for the time delay

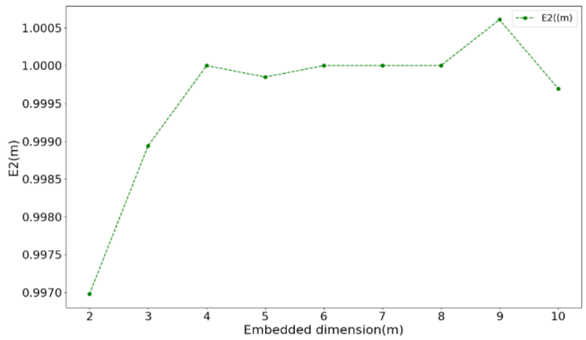


**Fig 8. Plot between** $E2(m)$ **and** $m$**.**

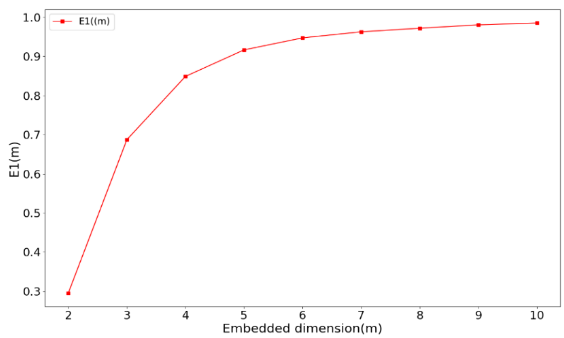


**Fig 9. Plot between** $E1\,(m)$ **and** $m$

in embedding [10]. The evidence supporting chaos in the time series data is reinforced by the positive value of the Lyapunov exponent and the nonintegral value of the correlation dimension. Additionally, the calculation of the parameter E2(m) in the Cao method serves as a robust indicator, providing further evidence of chaos in the studied data.

## 5 Conclusion

The Lyapunov exponent, computed as 0.00017 from the slope of the plot between $ln\frac{s_n}{s_0}$vs $n$ in Figure 5, indicates a positive value. Simultaneously, the correlation dimension, determined as 0.9086 from the slope of the plot between ln(C(R) and ln(R) in Figure 6, exhibits a non-integral value. These findings strongly suggest the presence of chaos in the temperature time series data. Furthermore, the values of E2(m) in Figure 8, not equal to 1, provide additional compelling evidence of the chaotic nature of the data.

The optimal time delay of 2, identified from the plot in Figure 7 where the AMI was minimum, was utilized for further analysis. The minimum embedding dimension, calculated from the plot between E1(m) and m in Figure 9, indicated saturation at m=6. Consequently, the minimum embedding dimension was determined as m+1=7 [8]. So, in addition to detection of chaos in time series data, Cao method is also useful in determining minimum embedding dimension.

This minimum embedding dimension of 7 suggests the existence of at least seven factors significantly influencing the system. Therefore, when formulating a governing equation and applying a forecasting model, it is advisable to incorporate this minimum number of factors to more accurately capture the system's behavior. The identification of these influential factors can contribute to the development of a more refined prediction model. The potential extension of this research involves studying time series data of other climate parameters such as atmospheric pressure, relative humidity, precipitation, and wind speed at different meteorological stations in India.

## 6 Acknowledgement

The authors are thankful for Department of Physics, Patna University and also grateful to Prof. Vijay A. Singh for fruitful discussions.

## References

1) Arfken GB, Weber HJ, Harris FE. Mathematical Methods for Physicists: A Comprehensive Guide. 7th ed. ELSEVIER. 2012. Available from: https://doi.org/10.1016/C2009-0-30629-7.
2) Hilborn RC. Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers. New York, USA. Oxford University Press. 1994.
3) Lorenz EN. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*. 1963;20(2):130–141. Available from: https://doi.org/10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2.
4) Mihailović DT, Mimić G, Arsenić I. Climate Predictions: The Chaos and Complexity in Climate Models. *Advances in Meteorology*. 2014;2014:1–14. Available from: https://doi.org/10.1155/2014/878249.
5) Indira P, Inbanathan SSR, Selvaraj RS, Suresh AA. Forecasting Daily Maximum Temperature of Chennai using Nonlinear Prediction Approach. *Indian Journal of Science and Technology*. 2016;9(39):1–6. Available from: https://sciresol.s3.us-east-2.amazonaws.com/IJST/Articles/2016/Issue-39/Article1.pdf.
6) Adewole AT, Falayi EO, Roy-Layinde TO, Adelaja AD. Chaotic time series analysis of meteorological parameters in some selected stations in Nigeria. *Scientific African*. 2020;10:1–21. Available from: https://doi.org/10.1016/j.sciaf.2020.e00617.
7) Hamid NZA, Adenan NH, Wahid NBA, Saleh SHM, Bidin B. Chaos Theory Modelling for Temperature Time Series at Malaysian High Population Area during Dry Season. *Environment and Ecology Research*. 2021;9(4):152–158. Available from: https://doi.org/10.13189/eer.2021.090402.
8) Bahari M, Hamid NZA. Analysis and Prediction of Temperature Time Series Using Chaotic Approach. In: IGEOS International Geography Seminar 2018 , 3–4 December 2018, Malaysia;vol. 286 of IOP Conference Series: Earth and Environmental Science. IOP Publishing. 2019;p. 1–9. Available from: https://iopscience.iop.org/article/10.1088/1755-1315/286/1/012027/meta.
9) Zonghua L. Chaotic time series analysis. *Mathematical Problems in Engineering*. 2010;2010:1–32. Available from: https://doi.org/10.1155/2010/720190.
10) Cao L. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena*. 1997;110(1-2):43–50. Available from: https://doi.org/10.1016/S0167-2789(97)00118-8.