J. Ind. Geophys. Union (July 2017)
v.21, no.4, pp: 277-284

A complete chaotic analysis on daily mean surface air temperature and humidity data of Chennai

A. Antony suresh* and R. Samuel Selvaraj
Department of Physics, Presidency College, Chennai, Tamil Nadu, India
*Corresponding Author: antony.suresh4@gmail.com

ABSTRACT
In this article are presented results from analysis of daily mean surface air temperature and humidity data applying nonlinear techniques. The data are collected for Chennai, India during January 1988–December 2013. The phase space, which illustrates the progress of the behavior of a nonlinear dynamical system, is reconstructed by Takens delay embedding theorem. The delay time and embedding dimension are estimated using average mutual information (AMI) and false nearest neighbor (FNN) algorithm respectively. Based on these embedding parameters (delay time τ and embedding dimension m) the correlation dimension for various embedding dimension and largest lyapunov exponent are estimated. Finally, the phase space reconstruction algorithm is employed to make a short-term prediction of the chaotic time series, whose governing equations of the system are unknown. The predicted values are in good agreement with the observed ones within 7 days, but they appear much less accurate beyond that limit (7 days). These results indicate that chaotic characteristics clearly exist in the air temperature and humidity data; techniques based on nonlinear dynamics can therefore be used to analyze and predict the air temperature.

Key words: Chaos, Phase space reconstruction, Hurst exponent, Lyapunov exponent and Poincare map.

INTRODUCTION

Atmospheric processes, such as air temperature and humidity, are usually nonlinear and complex (Selvam, 2012). The underlying complexity with in temperature and humidity makes the investigation as one of the indefinable tasks. The advancement in chaos theory offers new ways to draw the hidden information in random-like data. However, the methods originally branches from nonlinear dynamics and chaos theory, are now used to identify pure deterministic nonlinear mechanisms. Chaotic mechanisms are also used to identify deterministic elements which are mixed with other stochastic elements in the data. Although, chaotic time series illustrates the characteristic of dynamical systems as random, in the embedding phase space they represent deterministic behavior (Zhang and Man, 1998). Over the past two decades, distinguishing deterministic chaos and noise has become an important problem in many diverse fields, e.g., weather forecast (Lorenz, 1993), sunspot prediction (Park et al., 1996), hydrology (Sivakumar, 2000 and 2004; Rodriguez et al., 1989; Elshorbagy et al., 2002), traffic flow (Nair et al., 2001), foreign exchange rate (Das and Das, 2007), economics (Chen, 1998), etc., this is due to the availability of numerical algorithms for quantifying chaos using experimental time series.

The concept Phase-space reconstruction, embedding of a single-dimensional time series in a multi-dimensional phase-space to characterize the underlying dynamics, could be an useful methodology for forecasting. In recent time Phase-space reconstruction has found its applications in various fields such as lake volume (Abarbanel and Lali, 1996), rainfall (Berndtsson et al., 1994; Rodriguez et al., 1989), rainfall-runoff (Sivakumar et al., 2001b), etc. In this paper, we use nonlinear time series techniques to analyze the temperature and humidity data of Chennai, India. The results indicate that chaotic characteristics obviously exist in the temperature and humidity data; a technique which is based on phase space dynamics, can be used to analyze and predict the temperature and humidity.

Analysis of Nonlinear Time Series

Reconstruction of phase space

The fundamental idea of the phase space reconstruction is that, a time series contains the information about unobserved state variables, which can be used in the prediction of the present state. For a scalar time series x_t, where t = 1, 2, 3, . . ., may be used to construct a vector time series that is equivalent to the original dynamics from a topological point of view. The phase space can be reconstructed using the method of delays (Cao, 1997; Abarbanel et al., 1990; Frede, and Mazzega, 1999; Fraser and Swinney1986; Takens, 1981). The state space needs to form a coordinate system to confine the structure of orbits in phase space, this can be made by using lagged variable; $x_{t+\tau}$, where τ is the delay.

$$Y_t = \{x_t, x_{t+\tau}, x_{t+2\tau}, \ldots\ldots, x_{t+(m-1)\tau}\} \quad (1)$$

Where τ, m is referred to as the delay time and embedding dimension respectively.

Determining Time-Delay and embedding dimension

Time–delay embedding is the best empirical method for analyzing a dynamical system (Packard et al., 1980). It has been shown that under reasonable conditions, a time delay embedding preserves the quantities of the dynamical system in which we are interested (Takens, 1981; Whitney, 1936; Andrew and Harry 1986).

The proper choice of the time delay τ, is needed for reconstructing the trajectory in phase space from the chaotic scalar time series data. In order to characterize the chaotic systems and to obtain the quantities such as lyapunov exponent and other generalised entropies for a measured scalar time series, which is generated by a chaotic system, an appropriate state vector needs to be constructed with suitable time delay τ. For an infinite noise free data set the value of the delay time τ is in principle almost arbitrary (Takens, 1981), however for finite amount of data the choice of τ determines the quality of reconstructed trajectory in phase space and thereby one obtains for the generalized entropies, exponents and dimensions. The problem of proper choice of τ has been tackled by Fraser and Swinney (Andrew and Harry, 1986; Liebert and Schuster, 1989). Mutual information is a tool to measures of independence between data samples, the value of τ that produces the first local minimum of mutual information be used for phase space construction (Shaw, 1985), the first minimum of the mutual information can be selected as time delay. Average mutual information (AMI) is a theoretic method to connect two sets of measurements with each other criterion. The average mutual information between x(t) and x(t+ τ) can be calculated by

$$I(\tau) = \sum_{x(t),x(t+\tau)} \Pr(x(t), x(t+\tau)\log_2\left[\frac{\Pr(x(t), x(t+\tau)}{\Pr((x(t))\Pr(x(t+\tau))}\right] \qquad (2)$$

I(τ) determines the average amount of information shared by two values in the time series. When the value of T increases, x(t) and x(t+τ) becomes independent and I(t) will tends to zero (Abarbanel and Lali, 1996).

Similarly, determining an optimum embedding is a significant process, the precision of τ and m is directly related with the accuracy of invariables of the described characteristics of the strange attractors in phase space reconstruction. Time series which is reconstructed in minimal embedding dimension m, and the reconstructed attractor is a one-to-one image of the attractor in the original phase space. As the embedding dimension increases, the attractor unfolds; the same points on the attractor will not cross itself. The attractor would be completely unfolded in dimension (Embedding Dimension), where number of nearest neighbours arising through projection is zero. If time series is reconstructed to a smaller embedding dimension than minimum embedding dimension, then the state space trajectories projection of the points might appear as near neighbourhoods of other points which they are not neighbours in actual. The benefit of these neighbours, among other things, is that they allow the information on how phase space neighbourhoods evolve to be used to generate equations for the precise prediction of the time evolution of new points on or near the attractor (Abarbanel et al., 1990; Whitney, 1936; Farmer and Sidorowich,1987). The lowest dimension, in which none of the orbits in the attractor overlaps, is called the embedding dimension of that attractor. False nearest neighbour is an appropriate method for estimation of the optimum embedding dimensions, because this algorithm eliminates the incorrect neighbours. The points which are close to each other in one dimension, due to projection the points will be separated in higher embedding dimensions. The distance between two neighbor points amplify when going from dimension d to d+1 and it is a criterion for casting the embedding errors. This criterion is called false nearest neighbour, and should satisfy the following equation:

$$\left[\frac{R_{d+1}^2(t,r) - R_d^2(t,r)}{R_d^2(t,r)}\right]^{0.5} = \frac{\left|x(t+\tau) - x(t_r+\tau)\right|}{R_d(t,r)} > R_{tot} \qquad (3)$$

Where t and t_r are the times corresponding to neighbour point and the origin point, respectively. R_d is the distance in the phase space with embedding dimension d. R_{tot} is the tolerance threshold (Abarbanel et al., 1990).

Hurst exponent

Harold Edwin Hurst is known for introducing the Hurst exponent as a measure for the predictability of a time series (Hurst, 1951; Harris et al., 1987). Hurst exponent is not only used as a measure of long-term memory but also correlate the fractal dimension of the time series, and it has been used in many fields.

The Hurst exponent is determined using R/S analysis.

$$(R/S)_n \approx cn^H \qquad (4a)$$

$$\log(R/S)_n = \log c + H\log n \qquad (4b)$$

Where H is the Hurst exponent

Hurst exponent can change between 0 and 1. The Hurst exponent of 0.5 shows a true random walk. A value between 0 and 0.5 indicates non-persistent behavior, meaning that the data is not random but the current trend is unlikely to continue. A Hurst exponent between 0.5 and 1 proves that the data are more persistent and the current direction is likely to continue. Hurst exponent value is 0 means that the time series changes direction with every sample. A constant time series with non-zero gradient will result in a Hurst value of 1 (Edman, 1996).

Figure 1. Map showing the location of study area.

Poincaré map

The Poincaré map is a tool to observe the response of a nonlinear system. Analyzing high dimensional dynamical flow of the nonlinear system in the corresponding phase space is an important task. Typically, rather than analyzing the continuous flow in the $(d)^{th}$ dimension phase space, we observe the dynamics induced by the flow on a particular section of the phase space called Poincaré section (Basu, 2007). A Poincaré section is a hypersurface in the phase space, which is transverse to the flow of a given dynamic system. The intersection induces a set of points in $(d-1)^{th}$ dimension space. The projection of a Poincaré section on the X(T)-Y(T) plane is referred to as the Poincaré map of the dynamic system, where T is driving force. For chaotic motion, the return points in the Poincaré map form a geometrically fractal structure.

Largest Lyapunov Exponent

Lyapunov exponent is an appreciable quantitative measure of chaotic dynamics and in many cases it is the only evidence for chaos, Exponential divergence of nearby orbits in phase space is accepted as the hallmark of chaotic behaviour (Drazin, 1994; Ramasubramanian and Sriram, 2002). A system with at least one positive lyapunov exponent is defined to be chaotic. The magnitude of the exponent confirms the time scale, beyond which the system dynamics become unpredictable, that can be determined (Shaw, 1981). Lyapunov exponents are the average exponential rates of divergence or convergence of nearby orbits in phase space. Since nearby orbits correspond to nearly identical states, exponential orbital divergence means that the separation between the two orbits in that system will also be a function of time, whose initial difference will soon behave quite differently, so predictability is rapidly vanished. The mean exponential rate of divergence of two initially close orbits was described by

$$\lambda = \lim_{t \to \alpha} \frac{1}{t} \ln \frac{|\Delta x(X_0, t)|}{|\Delta X_0|} \tag{5}$$

This number, called the largest lyapunov exponent (λ) is used for distinguishing among the various types of orbits and provides a measure of the rate of this divergence (Froyland, 1992), the exponents measure the rate at which system processes create or destroy information (Shaw, 1981).

Chaos is basically deterministic; it is unpredictable beyond certain short intervals. In fact, the accurate prediction of a chaotic dynamical system is a function of the largest Lyapunov exponent (Abarbanel and Lali, 1996).

$$\Delta t_{\max} = \frac{1}{\lambda_{\max}}. \tag{6}$$

Study Area

Chennai is spread roughly from 12°50′ N to 13°17′ N latitude, and from 79°59′ E to 80°20′ E longitude. This area is one of the most highly populated urban sites and the fourth largest metropolis in India and 36^{th} largest urban area in the world, encompassing a total area of roughly $426 km^2$ (Figure 1). It is located on the south-eastern coast of India in north-eastern part of Tamil Nadu on a flat coastal plain known as the eastern coastal plains, with the Bay of Bengal to its east. For most part of the year, the weather is hot and humid.

Time series of daily mean temperature and humidity data of Chennai, were obtained from IMD. The 26-year period from 1988 to 2013, with 9055 data points is used in the study. Various researches over the past two decades have made a significant progress in the methods to identify chaos in a time series (Abarbanel and Lali, 1996; Berndtsson et al.,1994; Li and Liu, 2000) are used in this studies.

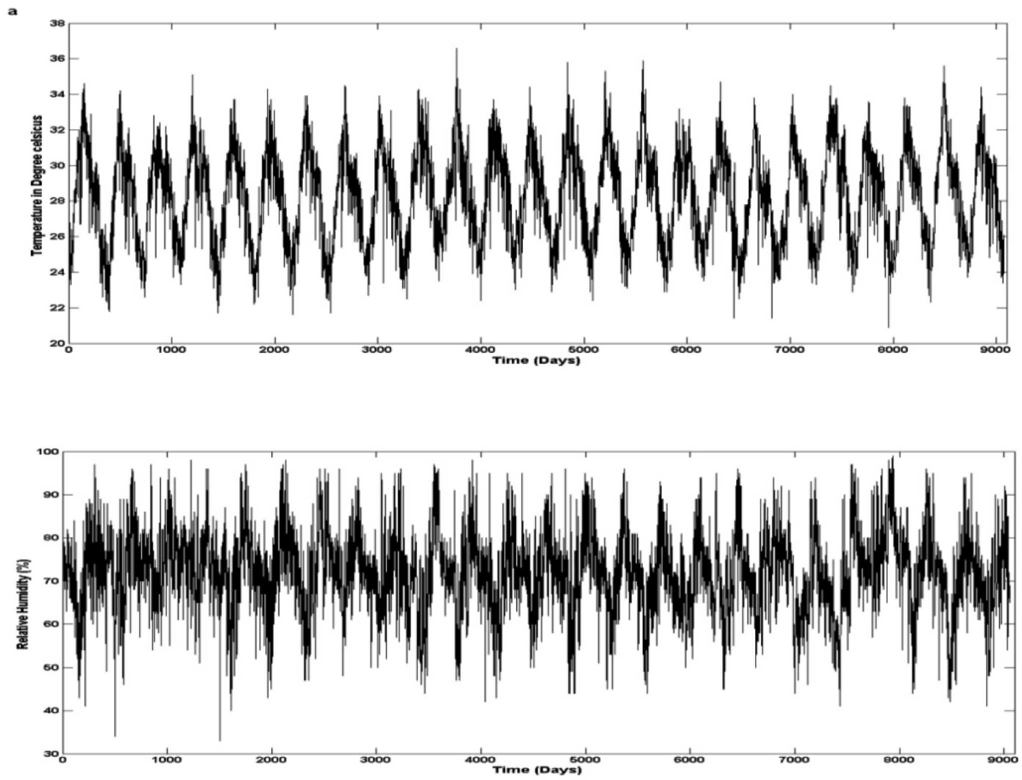

Figure 2(a, b): Time series plot of mean temperature and humidity.

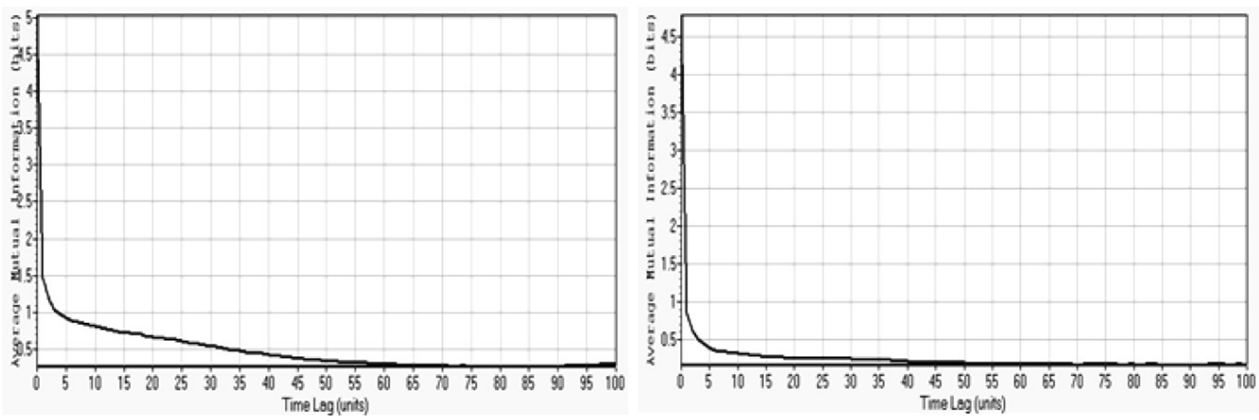

Figure 3. AMI bits Vs Time Lag for daily mean Air temperature and humidity.

Analysis and results

The time series of the daily mean temperature and humidity, collected over a period of about 26 years at chennai, India is shown in the Figure 2 (a,b).

Determination of reconstruction parameters

In order to reconstruct the original phase space, we need to calculate approximate reconstruction parameters, the delay time (τ) and embedding dimension (m). For both daily temperature and humidity time series, first minima time lag is calculated by AMI. The results (Figure 3) expose first minima at time lag 50 days and 7 days and for temperature and humidity time series respectively.

Calculating the percentage of false nearest-neighbors for the time series is the method used for the determination of the sufficient embedding dimension. The method shows that the estimation value of embedding dimension is 20 for both temperature and humidity data which is illustrated in the Figure 4. Both AMI and embedding dimension are estimated using Visual Recurrence Analysis (VRA) software.

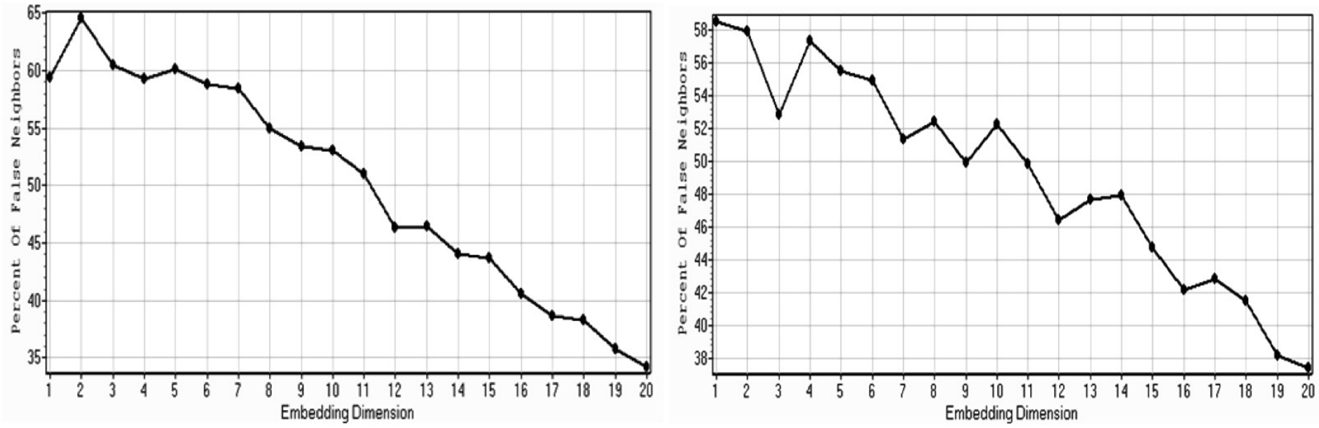

Figure 4. Embedding dimension Vs False neighbors for daily mean Air temperature and humidity.

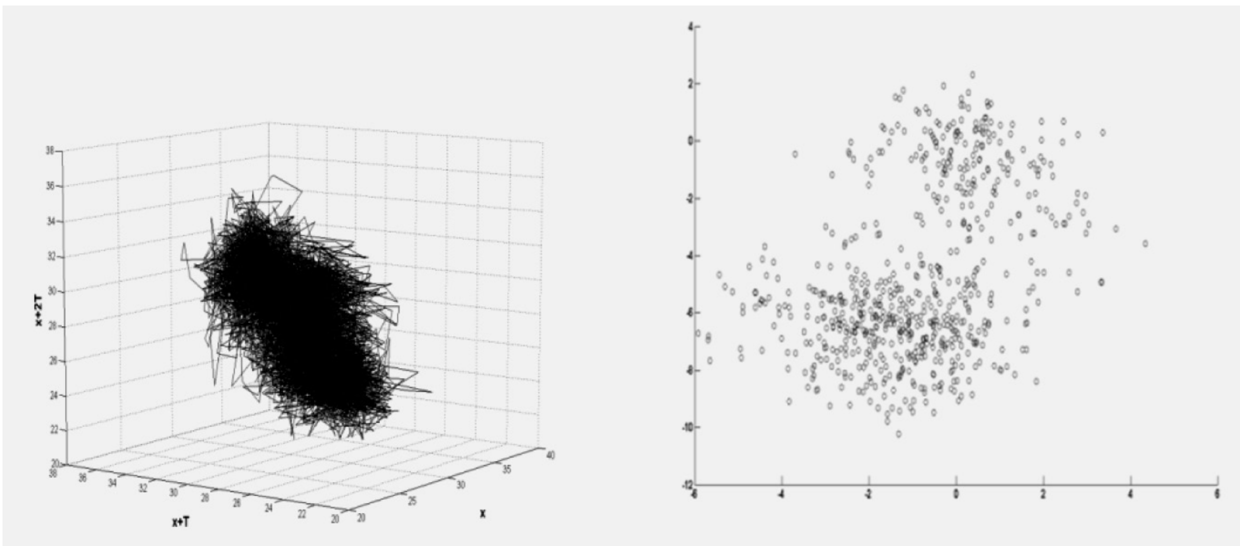

Figure 5 (a). Shows three- dimensional phase portraits and Poincaré map of the daily mean air temperature time series.

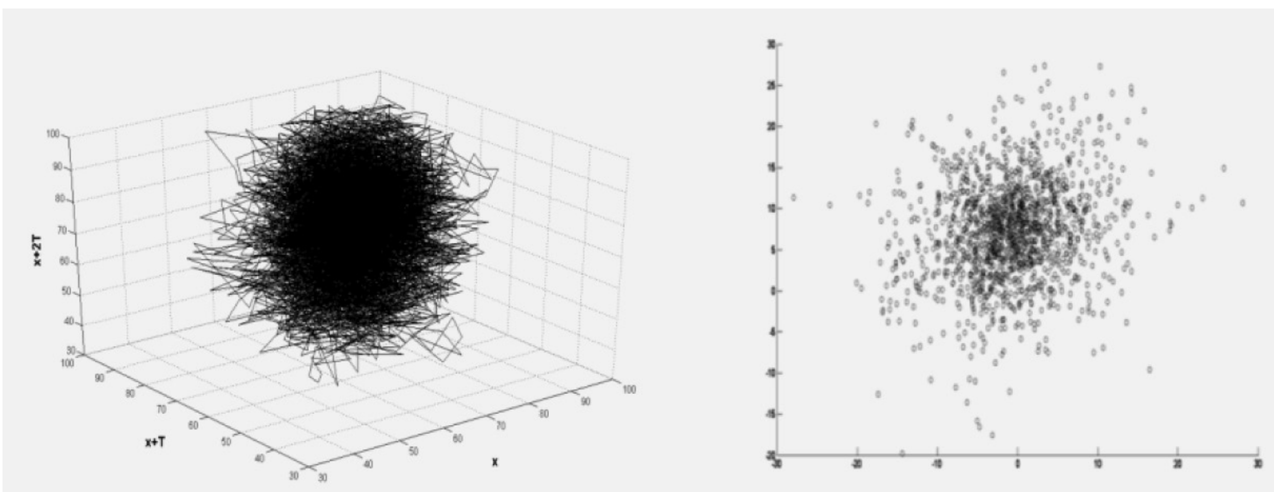

Figure 5 (b). Shows three- dimensional phase portraits and Poincaré map of the daily mean relative humidity time series.

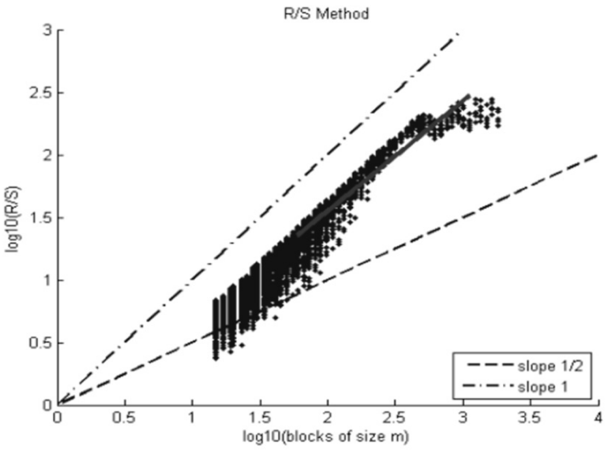

Figure 6(a). Hurst exponent plot for temperature data.

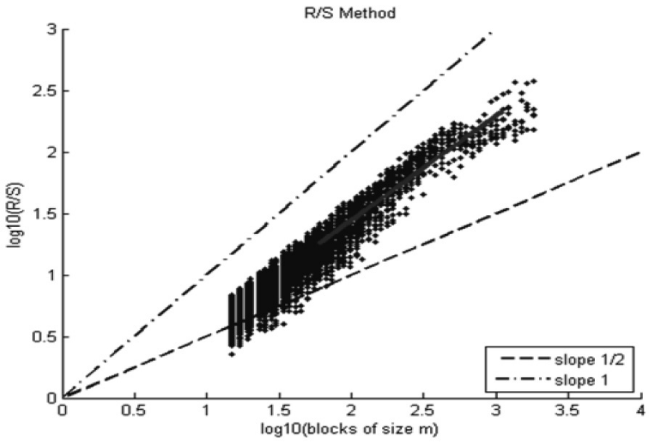

Figure 6(b). Hurst exponent plot for humidity data.

Table 1. complete analysis of time series data.

Period	Data type	Time delay (τ)	Embedding Dimension (m)	Hurst exponent (H)	Lyapunov exponent (λ)	Correlation coefficient between temperature and humidity data
1988-1993	Temperature	13	17	0.9826	0.084	-0.54
	Humidity	8	17	0.8953	0.528	
1994-1999	Temperature	13	17	0.9920	0.185	-0.65
	Humidity	11	17	0.9431	0.485	
2000-2005	Temperature	7	20	0.9775	0.266	-0.70
	Humidity	7	18	0.9617	0.335	
2007-2013	Temperature	16	20	0.9741	0.383	-0.65
	Humidity	10	19	0.9699	0.358	

Phase Space Representation

Phase space is a representer of a dynamical system where each point on that phase space represents a particular state of the system at a particular time. Phase space representation is versatile tool in time series analysis. Due to the fact that, phase space determines all the states of a dynamical system, analysis of that system can be achieved in both identifying the system and predicting the future states via Phase space representation. Figure 5(a, b): Shows three dimensional phase portraits of the reconstructed attractor for the daily mean air temperature time series reconstructed for $\tau=50$ and reconstructed at $\tau=7$ for relative humidity time series. Poincaré map is geometrically fractal structure, which confirms the chaos in the time series.

Estimation of Hurst exponent

Hurst exponent is derived using R/S analysis (MATLAB). Hurst exponent values calculated from Figure 6a and Figure 6b for temperature and humidity time series data are 0.8762 and 0.8495 respectively. The results show that both the time series data exhibit persistence nature.

Forecasting daily surface air temperature by Phase Space Reconstruction approach

Phase-space is a useful tool for characterizing dynamical systems. Firstly, in 1980, Packard, Crutchfiled, Farmer and Shaw suggested the theory of generating a reconstruction space from a single time series to characterize nonlinear dynamical system, and the theory was completed by F.Takens in 1981. Reconstruction of a single dimensional time series in a multi-dimensional phase-space explores the underlying phenomena. In that embedded phase space, the phase space analysis can be applied, because it owns the geometric properties as the state space. This fact arises from the fact that the attractor in reconstructed phase space is one to one image of the attractor in state space. Takens phase space reconstruction is the most popular method, using the past history and an appropriate time delay.

Math code for the phase-space reconstruction (PSR) approach is written, and forecasting is carried out in MATLAB.

The prediction for the daily mean temperature time series in Chennai is shown in Figure 7. The difference between the actual value and predicted data is negligible,

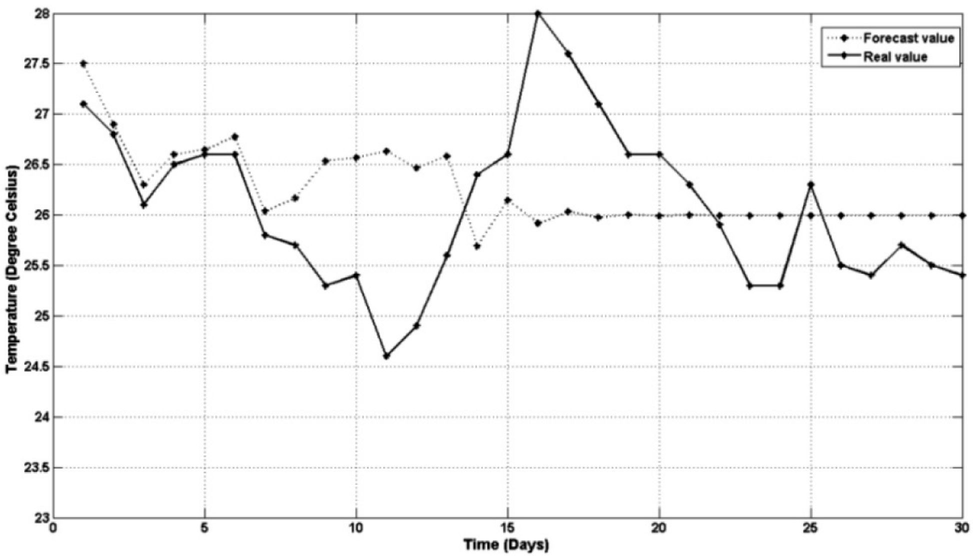

Figure 7. Comparison of observed and predicted daily mean air temperature in Chennai.

for 7 days, and it doesn't hold well beyond that. In fact, the largest Lyapunov exponent, from Eq. (5), suggests that the maximum window for accurate predictions of daily mean temperature data is about 2-3 days. The sensitivity to initial condition limits the predictive ability of a chaotic dynamical system. As Sugihara suggested, it is expected that chaos decreases the correlation between observed and predicted values as prediction time stretches (Sugihara and May, 1990). Also, the prediction results firmly develop the fact that information is contained in the data, so it is very effective for short term predictions only.

CONCLUSION

In this paper an attempt was made to study on the probable use of the concept of Phase-space reconstruction for understanding nonlinear dynamics behavior and to predict meteorological parameters like Air temperature and humidity data. Series of techniques were used to investigate chaotic behaviors in the temperature and humidity time series data. We have analyzed the time series observed over 26 years (January 1988 – December 2013) in Chennai, using Phase space reconstruction techniques. The results have shown that chaotic features evidently exist in the temperature and humidity data from the positive largest Lyapunov exponent and from the Poincaré map. The Largest Lyapunov tells us the maximum length of an accurate prediction is 3 days. These techniques can be further applied to other meteorological chaotic time series data.

ACKNOWLEDGEMENTS

Authors are thankful to Dr. V. Jothiprakash for constructive review and suggestions. They also thank Prof. B. V. S. Murty for editing the manuscript. We thank the Chief Editor for the support and guidance.

Compliance with ethical Standards

The author declares that they have no conflict of interest and adhere to copyright norms.

REFERENCES

Abarbanel, H.D.I., and Lali, U., 1996. Nonlinear dynamics of the Great Salt Lake: system identification and perdition. Climate Dyn., v.12, pp: 287-297.

Abarbanel, Henry, D.I., Brown, R., and Kartke, J.B., 1990. Prediction in chaotic nonlinear systems: methods for time series with broadband Fourier spectra. Phys. Rev., A., v.41, pp: 1782–1807.

Andrew M. Fraser., and Harry L. Swinney., 1986. Independent coordinates for strange attractors from mutual information. Phys. Rev. A., v.33, pp: 1134.

Berndtsson, R., Jinno, K., Kawamura, A., Olsson, J., and Xu, S., 1994. Dynamic system theory applied to long-term temperature and precipitation time series. Trends Hydrol., v.1, pp: 291-297.

Basu, A., 2007. Project in School of Electric and computer Engineering Georgia Institute of Technology Atlanta, GA.

Cao, L., 1997. Practical method for determining the minimum embedding dimension of a scalar time series. Physica, D., v.110, pp: 43-50.

Chen, P., 1998. Empirical and theoretical evidence of economic chaos. Syst Dynam Rev., v.4, no.1-2, pp: 11-38.

Das, A., and Das, P., 2007. Chaotic analysis of the foreign exchange rates, Applied Mathematics and Computation, v.185, pp: 388-396.

Drazin, P.G., 1994. Nonlinear systems. New York: Cambridge University Press.

Edman, A.N., 1996. Time Series Prediction Using Supervised Learning and Tools from Chaos Theory. PhD Thesis, University of Luton.

Elshorbagy, A., Simonovic, S.P., and Panu, U.S., 2002. Estimation of missing stream flow data using principles of chaos theory. J Hydrol., v.255, pp: 123–133.

Farmer, J.D., and Sidorowich, J.J., 1987. Predicting chaotic time series. Phys. Rev., Lett., v.59, no.8, pp: 845-848.

Fraser, A.M., and Swinney H.L., 1986. Independent coordinates for strange attractors from mutual information. Phys. Rev., A., v.87, pp: 1134-1140.

Frede, V., and Mazzega, P., 1999. Detectibility of deterministic nonlinear processes in earth rotation time series I. Embedding. Geophys. J. Int., v.137, pp: 551-564.

Froyland, J., 1992. Chaos and Coherence, Institute of Physics Publications.

Harris, C.M., Todd, R.W., Bungard, S.J., Lovitt, R.W., Morris, J.G., and Kell, D.B., 1987. The dielectric permittivity of microbial suspensions at radio-frequencies: a novel method for the estimation of microbial biomass. Enzyme Microbial Technology, v.9, no.3, pp: 181-186.

Hurst, H.E., 1951. Long term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers., v.116, pp: 770-799.

Li, T.Y., and Liu, Z.F., 2000. The chaotic property of power load and its forecasting. Proc CSEE v.20, no.11, pp: 36-40.

Liebert, W., and Schuster, H.G., 1989. Proper choice of the time delay for the analysis of chaotic time series, Phys. lett. A., v.142, no.2-3, pp: 107-111.

Lorenz, E.N., 1993. The Essence of Chaos, University of Washington Press, Seattle.

Nair, A.S., Liu, J.C., Rilett, L., and Gupta, S., 2001. Non-linear analysis of traffic flow, In: Proceedings of the IEEE Intelligent Transportation Systems Conference, Oakland, pp: 25-29.

Packard, N.H., Crutchfield, J.P., Farmer, J.D., and Shaw, R.S., 1980. Geometry from a time series. Physical Review Letters, v.45, no.9, pp: 712-716.

Park, Y.R., Murray, T.J., and Chen, C., 1996. Predicting sun spots using alayered perceptron neural network, IEEE Transactions on Neural Networks1(2)., pp: 501-505.

Ramasubramanian, K., and Sriram, M.S., 2002. A comparative study of computation of Lyapunov spectra with different algorithms.

Rodriguez-Iturbe, I., De Power, F.B., Sharifi, M.B., and Geograkakos, K.P., 1989. Chaos in rainfall. Water Resour. Res., v.25, no.7, pp: 1667-1675.

Selvam, A.M., 2012. Nonlinear Dynamics and Chaos: Applications in Atmospheric Sciences Journal of Advanced Mathematics and Applications, v.1, no.2, pp: 181-205 (25).

Shaw, R.S., 1985. The Dripping Faucet as a model chaotic system (Aerial Press, Santa Cruz, CA.

Shaw, R., 1981. Strange Attractors, Chaotic Behaviour, and Information Flow, Z. Naturforsch. 36A 80.

Sivakumar, B., 2004. Chaos theory in geophysics: past, present and future. Chaos, Solitons & Fractals., v.19, no.2, pp: 441-462.

Sivakumar, B., 2000. Chaos theory in hydrology: important issues and interpretations. J Hydrol; v.227, pp: 1-20.

Sivakumar, B., Berndtsson, R., Olsson, J., and Jinno, K., 2001b. Evidence of chaos in the rainfall- runoff process. Hydrol. Sci.J., v.46, no.1, pp: 131-145.

Sugihara, G., and May, R.M., 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature, v.344, pp: 734-741.

Takens, F., 1981. Detecting strange attractors in turbulence, in dynamic systems and turbulence, Lecture Notes in Mathematics, Springer Verlag, Berlin, pp: 366-381.

Whitney, H., 1936. Differentiable manifolds. Annals of Mathematics, v.37, pp: 645-680.

Zhang, J., and Man, K.F., 1998. Time series prediction using recurrent neural network in multi-dimension embedding phase space, IEEE International Conference on Systems, Man, and Cybernetics2, pp: 11-14.

Received on: 7.9.16; Revised on: 10.2.17; Accepted on: 28.2.17

"Even as global warming increases the frequency of El Nino and the Atlantic event, their effects are being amplified by the annual loss of an area of rain forest the size of New Jersey. Less rain falls, and the water runs into the rivers instead of being sucked up by the fungus filaments and tree roots." Alex Shoumatoff (1946--) is an American writer known for literary journalism, nature and environmental writing.