INVESTIGATION OF CHAOS EXISTENCE IN THE TIME SERIES OF LAHORE PRECIPITATION

*M.S. KHAN and M.J. IQBAL

Department of Mathematics, Government National College, Karachi, Pakistan

1Institute of Space & Planetary Astrophysics, University of Karachi, Karachi, Pakistan

(Received February 08, 2012 and accepted in revised form March 13, 2012)

The science of chaos is a burgeoning field, and the available methods to investigate the existence of chaos in a time series are still being developed. Chaos is also characterized by a positive Lyapunov exponent, which can be thought of as a measure of the long-term unpredictability of the system; equilibrium and periodic attractors have a negative exponent, whereas a quasiperiodic attractor has an exponent of zero. During the last few decades there have emerged several attempts to use the paradigm of ‘chaos’ for a description and forecasting of climatic processes. The predictability of daily rainfall is the most difficult task due to the nonlinear complex climate’s dynamical system. This paper aims to investigate the existence of chaos in the time series of Lahore precipitation.

Keywords: Quasiperiodic, Deterministic chaos, Delay-embedding, Phase space, Lyapunov exponent

1. Introduction

Any influence in nonlinear system may raise the complex behavior called chaos. The mere fact that simple deterministic systems generically exhibit complicated temporal behavior in the presence of nonlinearity. Time evolution as a system property can be measured by recording the time series. Thus, nonlinear time series methods will be the key to study nonlinear system and to find its physical interpretation. The nonlinear time series methods are the correlation dimension method [1, 2] the nonlinear prediction method [3-5] including deterministic versus stochastic diagram the Lyapunov exponent method [6], the Kolmogorov entropy method [7] the surrogate data method [8] and the linear and nonlinear redundancies [9, 10] Among these the correlation dimension method has been the most widely used one for the investigation of deterministic chaos in hydrological phenomena [11-21] This paper employs Kantz’s algorithm [22] to estimate the largest Lyapunov exponent, on the summer monsoon daily precipitation time series at Lahore. Lyapunov exponents are the average exponential rates of divergence or convergence of nearby orbits in phase space.

2. Data Analysis

This paper analyses daily monsoon precipitation data at Lahore PBO station (31° 33 N, 74° 31 E) for the summer season from June to September for the period from 1986 to 2005 inclusive. We have obtained this data from Pakistan Meteorological Department. The average of summer monsoon rainfall over Lahore is 4.05 mm per day. A time trace of precipitation data is plotted in Figure 1. It is erratic and distributed randomly. There is a small negative linear trend; this small negative trend could be due to urbanization. We remove trend by subtracting least square fit trend line from the data and obtained residuals for further analysis particularly spectral analysis [23]. Figure 1 shows excessively nonlinear patterns and these patterns could be generated by the deterministic nonlinear interaction of a few degrees of freedom [24-26] and which lead to the possibility of deterministic quasi-periodicity or "chaos" [27].

![Figure 1. Precipitation series of Lahore’s Monsoon season from 1986 to 2005.](image-url)
2.1. Power Spectrum

Spectral analysis displays the intensity or variability of the phenomenon versus period or frequency. The power spectrum represents the amount of energy associated with a specific frequency component. It can display for a system that is periodic or quasiperiodic by dominating frequencies and sub-harmonics. Chaotic and stochastic systems are easily distinguishable from periodic or quasiperiodic systems. The power spectrum proves useful in displaying the serial dependence present, in discovering periodic phenomena and in diagnosing possible models for a series.

It was developed by Blackman and Tukey [28] and is based on the Wiener-Khinchin theorem, which states that the Fourier transform of autocorrelation of a series yields power spectrum. In Blackman and Tukey's approach, power spectrum \( P(f) \) is estimated by

\[
P(f) = \frac{1}{\tau} \sum_{j=0}^{\tau-1} \rho_j W_j e^{-2\pi j f}
\]

Where \( \rho_j \) is the auto correlation function, \( \tau \) is the maximum lag considered and window length and \( W_j \) is the Hamming windowing function. Fig. 2 shows the power spectrum of trend-removed data. It shows a broadband and a periodic power spectra, whose sharp peaks indicate the presence of randomness and noisiness in the time series. Fig. 2 also shows a stochastic process which can be governed by an autoregressive moving average model or a low dimensional deterministic chaotic process [29].

2.2. Phase Space Reconstruction

Phase space and its parameters help to investigate nonlinear behavior of a time series. Any time series generated by a nonlinear process can be considered as the projection on the real axis of a higher-dimensional geometrical object that describes the behavior of the system under study [30]. A point in such a space is defined by a set of \( m \) dynamical variables. Delay Embedding Theorem [31, 32] states that a series of scalar measurements \( x(t) \) can be used in order to define the orbits describing the evolution of the states of the system in an \( m \)-dimensional Euclidean space. The orbits will then consist of points \( X(t) \) with coordinates.

\[
X(t) = [x(t), x(t+\tau), \ldots x(t+(m-1)\tau)]
\]

Where \( \tau \) is the delay time, and the dimension \( m \) of the vector \( X(t) \) is known as the embedding dimension. A new time series of the state space vector \( X(1), X(2), \ldots X(t) \) is generated from Eq. (2). Each vector \( X(t) \) describes a point in an \( m \) dimensional phase space. Thus, the sequence of these vectors defines a trajectory in time as shown in Fig. 3. Geometrically, the entire set of these points forms a pattern, termed an attractor, in the phase space. According to [31] theorem, if \( d \) is the dimension of the original attractor, it is sufficient that the embedding phase space dimension \( m \) must be greater than or equal to \( 2d+1 \). However, in reconstructing an attractor from a time series of unknown dynamics, the dimensionality of the attractor is unknown. It is important that the reconstruction must be embedded in a space of sufficiently large dimension to represent the dynamics completely.

Figure 2. Power spectrum of trend-removed data against frequency.

Figure 3. The diagram of evolution in \( m \)-dimensional phase space.
2.3. Determination of Delay Time

Fraser and Swinney, [33] described a method called Average Mutual Information (AMI) for estimating delay time. They argue that a better value for $\tau$ is the value that corresponds to the first local minimum mutual information between two time series $X(t)$ and $X(t+\tau)$. The average mutual information is given mathematically as

$$I(\tau) = \sum_{X(t)X(t+\tau)} P(X(t)X(t+\tau)) \log \left[ \frac{P(X(t)X(t+\tau))}{P(X(t))P(X(t+\tau))} \right] \tag{3}$$

$P(X(t))$ and $P(X(t+\tau))$ are individual probabilities for the measurements of $X(t)$ and $X(t+\tau)$. $P(X(t),X(t+\tau))$ is the joint probability density for measurements $P(X(t),X(t+\tau))$. The appropriate time delay $\tau$ is defined as the first minimum of the average mutual information $I(\tau)$. Then the values of $X(t)$ and $X(t+\tau)$ are independent enough of each other to be useful as coordinates in a time delay vector but not so independent as to have no connection with each other at all.

We plot AMI of the daily summer rainfall of Lahore after removing trend in Fig. 4. As the first minima occurs at $\tau = 4$, we infer that delay time for our climate data is 4 days.

![Figure 4. Average mutual information graph of Trend removed monsoon data.](image)

2.4. Determination of Embedding Dimension

A useful technique to estimate an optimal value of $m$ is to look for the closed false nearest neighbors (FNN) in the phase space at a given value of $m$ [34] developed an algorithm, called false nearest neighbor algorithm that estimates the sufficient dimension for phase space reconstruction.

The false nearest neighbor algorithm identifies points within a nonlinear time series that seems to correlate, or relate, at a certain point in time. This is the same goal as the correlation integral; however, the false nearest neighbor could more accurately determine a chaotic system since it graphs the data on an n-dimensional scale represented by n-dimensional vectors. By increasing the dimension, it is possible to detect "false neighbors" within the vectors because once the attractors unfold; the vectors near in this dimension move a significant distance apart in the next state. This would indicate that the attractor of the system has not been accurately identified. Then the dimension is increased by one.

The algorithm is that for each point $X(i, m)$ in the time series looks for its nearest neighbor $X(j, m)$ in an m-dimensional space. Calculate first the distance $\|X(i, m) - X(j, m)\|$. Then, iterate both points and compute the ratio $R(i, m)$.

$$R(i, m) = \frac{\|X(i, m+1) - X(j, m+1)\|}{\|X(i, m) - X(j, m)\|} \tag{4}$$

If $R(i, m)$ exceeds a given heuristic threshold, say $R(t)$, this point is marked as having a false nearest neighbor. The criterion that the embedding dimension is high enough that the fraction of points for which $R(i, m) > R(t)$ is zero, or at least sufficiently small. In order to apply the method to the present time series a suitable value for the Theiler window has to be selected.

We use TISEAN software [35] to calculate embedding dimension of our climate data. Setting the value of $\tau = 4$, ratio factor = 10 and initially Theiler window = 0. Fig. 5 is the plot of the false neighbors against the embedding dimensions of the trend removed climate data. Figure 5 shows, a slow convergence towards FNN from $m = 8$ to 27, we should analysis for noise in the data.

![Figure 5. Percentage of false neighbors of trend removed climate data come to zero at $m = 27$.](image)
2.6. Noise Reduction

There is high probability for random noise in the data, which is responsible to spread the time series. Therefore, it needs to be filtered. As a stochastic process, noisy data exhibits large number of degrees of freedom and therefore, it should show no tendency to unfold at any specific dimension. Thus we able to eliminate events that show high embedding dimension. Moving average and low-pass filter methods are commonly used for noise reduction. In the present study, however, we use a nonlinear locally projective noise reduction scheme specifically developed for chaotic data as proposed by [36].

The hypotheses of nonlinear locally projective noise reduced algorithm are that the measured data is composed of the output of a low-dimensional dynamical system and of random or high-dimensional noise. This means that in an arbitrarily high-dimensional embedding space the deterministic part of the data would lie on a low-dimensional subspace, while the effect of the noise is to spread the data off this subspace. The idea of the projective nonlinear noise reduction scheme is to identify the subspace and to project the data onto it [35].

Suppose the dynamical system forms a \( v \)-dimensional subspace \( V \) containing the trajectory. All embedding vectors \( X(i) \) would lie inside another subspace \( \tilde{V} \) in the embedding space. For each \( X(i) \) there exists a correction \( \Delta X(i) \), with \( \| \Delta X(i) \| \) small, in such a way that \( X(i) = \Delta X(i) \) and that \( \Delta X(i) \) are orthogonal on \( \tilde{V} \). So, vectors have to be over embedded in \( m \)-dimensional spaces with \( m > v \). This idea is realized through the TISEAN Software [35].

Fig. 5 shows a possible value of \( m \) between 5 and 12, taking \( \tau = 4 \), \( m = 27 \) and minimum false neighbours = 10 with \( v = 3 \) in safe for locally projective noise reduced data in the TISEAN software; we obtain a locally projective noise reduced data. Hence, we calculate the time delay and embedding dimension again from the locally projective noise reduced data. AMI plot in Fig. 6 shows time delay \( \tau = 2 \). With new time delay, we calculate new embedding dimension by false neighbors method of locally projective noise reduced data with Thiler window = 10 and \( \tau = 2 \). Fig. 7 shows a considerable changing in embedding dimension and the new value of embedding dimension is 8.

![Figure 6. Average mutual information’s graph of locally projective noise reduced data shows \( \tau=2 \).](image)

![Figure 7. Embedding dimension of locally projective noise reduced data from false neighbors method. Embedding dimension = 8 for \( \tau = 2 \).](image)

2.7. Phase Space Trajectories

If a time series contains chaotic properties, the state vector \( X(t) \) will be attracted to a particular region in the phase space known as the strange attractor [14, 37]. The attractor may, however, be completely concealed if the time series contain noise.

Figures 8 (a,b) and 9 show respectively 2 and 3-dimensional phase space trajectories of trend removed summer monsoon’s data for \( \tau = 4 \). In Fig.8 (a, b) and Fig. 9 show 3-dimensional phase space trajectories of noise reduced data for \( \tau = 2 \).
2.8. Recurrence Plots

Recurrence is a fundamental property of dissipative dynamical systems. Although small disturbance of such systems cause exponentially divergence of its state, after some time the system will come back to a state that is arbitrary close to a former state and pass through a similar evolution [38]. Recurrence plot is a 2-dimensional N x N pattern of points where N is the number of embedding vectors X(t) obtained from the delay coordinates of the input signal. A point (i, j) in this plot is set of

$$ R_{ij} = \Theta(r - |X(i) - X(j)|), \quad i, j = 1, 2, ..., N, $$

where $r$ is a predefined threshold and $\Theta$ is the Heaviside step function. The norm is arbitrary, although we use the maximum norm. Recurrence plots give information about the temporal correlation of phase space points. From the occurrence of lines parallel to the diagonal in the recurrence plot it can be seen how fast neighboured trajectories diverge in phase space and also show periodicity. However, these lines might not be so clear and it could contain subtle patterns which can not be visualized easily, in that case [39] propose recurrence quantification analysis (RQA) to quantify recurrence behaviour. In this paper we employ only recurrence plot analysis.

Recurrence plots help revealing phase transitions and instationarities [40]. Visible rectangular block structures with a higher density of points in the recurrence plot indicate phase transition within the signal. If the texture of the pattern within such a block is homogeneous, stationarity can be assumed for the given signal within the corresponding period of time, i.e. the points should cover the plane uniformly on average, whereas nonstationarity expresses itself by an overall tendency of the dots to be close to the diagonal. The contrast of the resulting images can be selected by the distance $r$ and the percentage of points that should be actually evident by a black region far away from the diagonal. The recurrence plot in the gray scale of the trend removed summer monsoon precipitation time series of Lahore is obtained and displayed in Fig. 10, which provides evidence of the stationary state in the data.
The Nucleus 49, No. 1 (2012)

3. Chaotic Analysis of Precipitation Data

The most striking feature of chaos is the unpredictability of the future despite a deterministic time evolution. The average error made when forecasting the outcome of a future measurement increases very rapidly with time. This unpredictability is a consequence of the inherent instability of the solutions, reflected by what is called sensitive dependence on initial conditions.

A more careful investigation of this instability leads to two different, although related, concepts. One aspect, which we do not want to elaborate in this paper, is the loss of information related to unpredictability. This is quantified by the Kolmogorov-Sinai entropy [7]. The other aspect is a simple geometric one, namely, that nearby trajectories separate very fast, or more precisely, exponentially fast over time. We analyse the precipitation data from the second approach.

3.1. Kantz’s Algorithm for Lyapunov Exponent

Exponential divergence of nearby orbits in phase space identifies chaotic behavior [41]. The properly averaged exponent of this increase is characteristic for the system underlying the data and quantifies the strength of chaos. It is called the Lyapunov exponent [30].

Lyapunov exponents measure the rate of divergence of initially close trajectories. A positive but finite Lyapunov exponent is, therefore, a sharp criterion for the existence of deterministic chaos [42] as well as [43] introduced locally linear fits to the dynamics in order to follow the evolution in tangent space. The algorithm by [6] follows several nearby trajectories to measure the average increase of local phase space volume.

Similar algorithms for this purpose have been proposed independently by [22, 44]. We follow the Kantz’s algorithm to investigate chaos in this paper. His algorithm works as, if we take two points in the phase space initially $X_0 (i)$ and $X_0 (j)$ and indicate their distance as

$$| X_0 (i) - X_0 (j) | = \delta_0$$

Then after time $t$ it is expected that the new distance $\delta_t$ between same two points will be

$$\delta_t = \delta_0 e^{\lambda t}$$

$$\delta_t = \delta_0 , \text{ if } \lambda = 0 , \text{ that is the case of a cyclical series or a steady state.}$$

$$\delta_t < \delta_0 \text{ when } \lambda < 0 , \text{ that is the case of convergent series towards a steady.}$$

M.S. Khan and M.J. Iqbal
The Nucleus 49, No. 1 (2012)

Fig. 12. Estimation of largest Lyapunov exponent by Kantz’s algorithm for trend removed data (noisy data) and (b) noise reduced data.

\( \delta_i > \delta_0 \) when \( \lambda > 0 \), that is the case of divergent series.

Where \( \lambda \) is called the Lyapunov exponent. Kantz calculated an unbiased stretching factor \( S(r,m,t) \) for maximal Lyapunov exponent,

\[
S(r,m,t) = \frac{1}{N} \sum_{i=0}^{N-1} \ln \left( \frac{1}{|U_i|} \sum_{j \in U_i} |X(i+t) - X(j+t)| \right) 
\]

(8)

Where \( |U_i| \) is the number of neighbors found around point \( X(i+t) \).

For a low-dimensional deterministic process the Lyapunov exponent should be a positive finite
number, for a linear process it should be zero and for a stochastic process it should be infinite. In general, for a Lyapunov exponent describes an m-dimensional phase space the rate of expansion or contraction of orbits for each direction, resulting in m different Lyapunov exponents of whom some are zero or negative. However, the main interest is focused on the largest of these exponents since it can be calculated relatively easy and it yields evidence for the presence of deterministic chaos in the observed data.

If S (r, m, t) exhibits a linear increase with identical slope for all m larger than some m0 and for a reasonable range of initial neighborhood size r, then this slope can be taken as an estimate of the maximal exponent λ. The smaller r, the large the linear range of S, if there is one. Obviously, noise and the finite number of data points limit r from below. It is not always necessary to extend the average in Eq. (8) over the whole available data; reasonable averages can be obtained already with a few hundred reference points. If some of the reference points have very few neighbors, the corresponding inner sum in Eq. (8) is dominated by fluctuations. Therefore, one may choose to exclude those reference points, which have less than, say, ten neighbors. However, discretion has to be applied with this parameter since it may introduce a bias against sparsely populated regions. This could in theory affect the estimated exponents due to multifractality. Like other quantities, Lyapunov exponent may be affected by serial correlations between reference points and neighbors. Therefore, a minimum time for ∆t =|t−t′| is considered.

Now we calculate the Lyapunov exponents for both noisy data and noise-removed data. Using Thierler window |t−t′|= 10 and other embedding values such as τ = 4, m = 20 to 27, minimum length scale to search neighbours, r = 0.005 and maximum length scale to search neighbours, R = 0.05 for noisy data. We also calculate maximal Lyapunov exponents, setting τ = 2, m = 3 to 8, minimum length scale to search neighbours, r = 0.005 and maximum length scale to search neighbours, R = 0.05 for noise reduced data using TISEAN software. Results are summarized in Table 1. The Lyapunov exponents for both data are 0.0425 and 0.0103 respectively. Local Lyapunov exponent for both data are very small and positive. This indicates that the time series of summer monsoon rainfall can exhibit chaotic behaviour after long time.

**Table 1. Largest Lyapunov exponents for trend removed and noise reduced data.**

<table>
<thead>
<tr>
<th></th>
<th>Time Delay</th>
<th>Embedding Dim.</th>
<th>Max. Lyapunov Exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trend Removed Data</td>
<td>4</td>
<td>26</td>
<td>0.0425</td>
</tr>
<tr>
<td>Noise Reduced Data</td>
<td>2</td>
<td>8</td>
<td>0.0103</td>
</tr>
</tbody>
</table>

4. **Conclusions**

In this paper we outline a methodology to investigate chaos and its characteristics of the summer monsoon precipitation data of Lahore's climate by using ideas from dynamical systems theory. The power spectrum of trend removed data shows no proper frequency or periodicity in the date. Phase space and its parameters indicate that the data is high dimension having highly nonlinear trajectories. Recurrence plots of trend removed and noise reduced data show stationarity, nonperiodic in the data but vertical and horizontal lines structure contains information that the system changes its stage slowly and indicates deterministic system or a laminar state. Finally, the largest local Lyapunov exponents for trend removed and noise reduced data are computed from local slopes of trend line fit and these are found to be 0.0425 & 0.0103 respectively. Both results are positive but strongly not recommended for the redirection of chaos but beginning of chaos. This proves the chaotic dynamic of the system.

5. **Acknowledgements**

We thank Mr. Nadeem of Pakistan Metrological Department for providing Lahore's precipitation data and University of Karachi for awarding fellowship. We are also thankful to the Institute of Space and Planetary Astrophysics, University of Karachi for the lab. facility provided to us.

**References**

Investigation of chaos existence in the time series