Dynamical Structure in Paleoclimate Data

H. Bruce Stewart
Department of Applied Science
Brookhaven National Laboratory
Upton, New York 11973

December 1, 1994

Abstract. Deterministic chaos in dynamical systems offers a new paradigm for understanding irregular fluctuations. The theory of chaotic dynamical systems includes methods which can test whether any given set of time series data, such as paleoclimate proxy data, are consistent with a deterministic interpretation.

Paleoclimate data with annual resolution and absolute dating provide multiple channels of concurrent time series; these multiple time series can be treated as potential phase space coordinates to test whether interannual climate variability is deterministic. Dynamical structure tests which take advantage of such multichannel data are proposed and illustrated by application to a simple synthetic model of chaos, and to two paleoclimate proxy data series.


This work was supported by the CHAMMP initiative of the Office of Health and Environmental Research, U.S. Department of Energy under Contract No. DE-AC02-76CH00016.
The Notion of Dynamical Structure

The emergence of the chaos paradigm of dynamical systems theory raises the possibility that some of the irregular oscillations observed in the natural world may be explainable as the workings of deterministic dynamics. Moreover, there is reason to hope that the deterministic rules governing the behavior of even large and complex natural systems may be expressed in terms of just a handful of active modes of oscillation, which can be fully described using a small number of state variables.

A deterministic dynamical system requires a state space or phase space whose coordinates are the state variables \( x_1, x_2, \ldots, x_n \); these variables describe the state of the system at any instant of time, and the functions \( x_1(t), x_2(t), \ldots, x_n(t) \) describe the system evolution over time. In addition, a dynamical system possesses a dynamical rule which specifies completely and unambiguously for each state \( X = \{x_1, x_2, \ldots, x_n\} \) the immediate future trend of evolution; that is, the rule uses \( X(t) \) at time \( t \) to determine \( X \) a short time interval into the future [Abraham and Shaw, 1992; Thompson and Stewart, 1986].

For many systems, there is a natural discrete unit of time, such as a day or a year. The evolution is described by a sequence \( X_i \) with index \( i \) indicating time, that is, \( X_i = X(t_i) \). The dynamical rule is then conveniently expressed as an iterated function

\[
X_{i+1} = F(X_i) \tag{1}
\]

where \( F \) is a vector function with vector arguments. The evolution begins from an appropriate initial condition

\[
X_0 = \{x_1(t_0), x_2(t_0), \ldots, x_n(t_0)\} \tag{2}
\]

In other systems, it is natural to consider time flowing as a continuum, so that the dynamical rule is a differential equation

\[
dX/dt = \dot{X} = F(X)
\]

that is,
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
The evolution of the system from an initial state

\[x_1 = f_1(x_1, x_2, ..., x_n)\]

\[x_2 = f_2(x_1, x_2, ..., x_n)\]

... \(\vdots\)

\[x_n = f_n(x_1, x_2, ..., x_n)\] \hspace{1cm} (3)

The evolution of the system from an initial state

\[X_0 = \{x_1(t = 0), x_2(t = 0), ..., x_n(t = 0)\}\] \hspace{1cm} (4)

forward in time is the solution \(X(t) = \{x_1(t), x_2(t), ..., x_n(t)\}\) of the initial value problem (3) and (4). The evolution from an initial state – a point in phase space – traces out a smooth trajectory in phase space, provided the functions \(f_1, f_2, ..., f_n\) are continuous functions. That is, the magnitude of the difference \(|F(X_1) - F(X_2)|\) should be small whenever \(|X_1 - X_2|\) is sufficiently small. Here \(| \cdot |\) indicates a distance, for example Euclidean distance, in phase space.

In discrete time systems (1) the trajectory is usually not smooth, but \(F\) should still be continuous in \(X\).

In physical problems, such as mechanical or electrical systems, an appropriate phase space is usually apparent from the form of the laws of motion. For example, mechanical problems require a position and a velocity for each mechanical degree of freedom [Abraham and Shaw, 1992]. However, in other cases it may be a difficult task to choose an economical set of state variables, that is, a reasonably small number of coordinates which still retain the essential property of the dynamical rule: that knowing the instantaneous state \(X = \{x_1, x_2, ..., x_n\}\) is sufficient to specify the immediate future trend of evolution is a completely deterministic way and without ambiguity. In spite of this difficulty, recent successes like the characterization of the historical fluctuations in the level of the Great Salt Lake using a four-dimensional state space [Abarbanel, 1995] show that it is possible. More particularly, it is possible to determine an appropriate state space or phase space by analyzing an observed time series, without recourse to a model based on physical laws.

Low-dimensional long-term behavior can occur in complex natural systems whose phase space would seem to be of very high dimension. Dissipation, which exists in most natural systems, causes volumes of ensembles in phase space to contract as time advances; this is the equivalent of Liouville’s theorem for energy-conserving systems [Thompson and Stewart, 1986,
In many cases, dissipation acts even more strongly, reducing the long-term fluctuations to a subset of dimension much smaller than the number of phase space dimensions suggested by the laws of motion. This is not a theorem, but a commonly observed phenomenon.

If the long-term dynamics of a system has a low-dimensional description, then one may hope that a moderately long observed trajectory will come near to every state possible for the long-term dynamics; that is, it comes near every point in the attractor. If the dynamical rule $F(X)$ is a continuous function, then it will be possible to make good short-term forecasts by identifying dynamical analogs in past observed behavior [Lorenz, 1969].

Most of the methods for detecting dynamical structure begin with the modest assumption that only one time series of a single state variable has been recorded. From a single time series, additional phase space coordinates can be reconstructed using a procedure known as time-delay embedding [Packard et al., 1980]. An excellent review of these methods has recently appeared [Abarbanel et al., 1993]. Here we consider a situation in which two or more concurrent time series of different variables are available for analysis.

The growing body of paleoclimate proxy data with annual resolution makes it possible to consider whether year-to-year climatic fluctuations can be described by a deterministic dynamical rule, in the sense stated above, by examining observed data. Furthermore, the absolute dating of annual resolution data means that the concurrence of two or more independent time series can be established. This concurrence is an essential prerequisite for asking whether such multichannel data represent phase space coordinates for a possible deterministic dynamical rule.

Identifying Dynamical Structure from Time Series

The hallmark of chaos is that evolutions from two nearby states in phase space will gradually diverge from each other as time progresses. In mathematical terms, the system is sensitive to initial conditions. An error or perturbation introduced at any time will grow over time, typically at a geometric rate; this makes long-term forecasting impossible. Given a moderately long time series of all $n$ phase space coordinates, this gradual divergence can be verified by finding good dynamical analogies, that is, pairs of widely separated times in the observed record when the two system states were near each
other in phase space. These good dynamical analogies are manifestations of recurrence.

Let us consider data \( X(t_i) \) from a continuous time evolution sampled discretely at equally spaced times \( t = t_i, i = 1, 2, ..., N \). For each \( t_i \), the best dynamical analogy for \( X(t_i) \) involves its nearest neighbor in phase space. Let us say that this nearest neighbor occurs at time \( t = t_{N(i)} \); in determining \( N(i) \) we exclude times near \( t_i \) so that the analogy belongs to distinct parts of the trajectory and represents a true recurrence.

For each such analogy, the rate of divergence over \( j \) steps forward in time can be measured in terms of the local spreading ratio

\[
S(i,j) = \frac{\|X(t_{N(i)+j}) - X(t_{i+j})\|}{\|X(t_{N(i)}) - X(t_i)\|}.
\]

An important mathematical fact about this spreading ratio is that (roughly speaking) for large values of \( j \) the equivalent rate \( (1/j) \ln S(i,j) \) tends to a limit which is independent of \( i \), and independent of the particular choice of coordinates; in other words, the limit is an invariant quantity. The limiting value is called the largest Lyapunov exponent, denoted \( \lambda_1 \). Sometimes base 2 logarithms are used, so that \( \lambda_1 \) is an inverse doubling time for uncertainties or perturbations. There is in fact a spectrum of limiting rates or Lyapunov exponents; only the largest, \( \lambda_1 \), is manifested in the long-term spreading of two typical nearby trajectories. The criterion for chaos is \( \lambda_1 > 0 \) [Abarbanel et al., 1993].

In the mathematical definition of Lyapunov exponents, it is assumed that the initial separation at \( j = 0 \) is infinitesimal, so that even for large \( j \) the separation is not too large. When dealing with a finite sample of observed data, this is of course not true, so it may not be practical to consider \( j \) large enough to obtain a true invariant quantity. Instead, one may examine the local divergence rate

\[
s(i,j) = (1/j) \ln S(i,j)
\]

and its average

\[
\bar{s}(j) = (1/i) \sum_i s(i,j)
\]

Although these are not invariant quantities, and do depend on the choice of coordinates, it is still possible to obtain from them useful information about
possible dynamical structure.

A Prototype Example

To illustrate how this can be accomplished, we first consider synthetic data generated by numerical solution of a simple system of three first-order ordinary differential equations

\[
\begin{align*}
\dot{x} &= -y - z \\
\dot{y} &= x + 0.36y \\
\dot{z} &= 0.4 + z(x - 4.5)
\end{align*}
\] (8)

This system was devised by Roessler [1976] to give an example of the simplest possible chaotic attractor, the folded band. Any trajectory of this system will, after an initial transient, settle onto a coherent three-dimensional structure. Within this coherent structure, nearby states exhibit gradual divergence over time. A typical trajectory on this attractor is illustrated in Figure 1, which shows the coordinates \(x, y,\) and \(z\) plotted as three time series above, and in phase portraits. The upper left phase portrait shows an orthogonal projection of the three coordinates with the \(z\)-axis tilted at 45 degrees, while the lower left shows a projection along the \(z\)-axis onto the \((x, y)\) plane. On the right is a different trajectory to be discussed below.

On the average, separations are roughly doubled for each circuit around this attractor; sampling at about 60 discrete time steps per circuit, we expect \(\lambda_1(j)\) to be about \((1/60) \ln 2\), or roughly 0.011.

Suppose we are given a three-channel time series \(x(t_i), y(t_i), z(t_i)\); we wish to determine, from the data themselves and without knowing their origin in eqs. (8), whether they were generated by a deterministic rule. One method of diagnosis consists of computing the average local divergence rate \(\bar{s}(j)\), and comparing with the divergence rate computed from a surrogate data set. By surrogate, we mean data which from their appearance as time series could plausibly have come from the same source, but have in fact been arranged or manipulated so that they lack dynamical structure.

Surrogate data sets have been used for this type of diagnosis using time-delay embeddings, where only a single variable has been observed and recorded; see for example [Theiler et al., 1992; Takens, 1993]. Here we are considering multichannel data, so it makes sense to look at surrogate data in which only
Table 1: Average local divergence rate $\bar{a}(j)$ for trial embeddings of the Roessler band attractor in two and three dimensions; 5000-point trajectories sampled 60x per turn.

<table>
<thead>
<tr>
<th>Case</th>
<th>Coordinates</th>
<th>average local divergence rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$j=1$</td>
</tr>
<tr>
<td>1</td>
<td>$x - y - z$</td>
<td>.001</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.010</td>
</tr>
<tr>
<td>3</td>
<td>$x' - y - z$</td>
<td>.796</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>.819</td>
</tr>
<tr>
<td>5</td>
<td>$x - y - z'$</td>
<td>.201</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>.209</td>
</tr>
<tr>
<td>7</td>
<td>$x - y$</td>
<td>.392</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>.363</td>
</tr>
<tr>
<td>9</td>
<td>$x' - y$</td>
<td>2.479</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>2.738</td>
</tr>
</tbody>
</table>

one channel has been replaced by a plausible substitute, with other channels unchanged. We then speak of a surrogate channel of data.

A simple method of generating surrogate channels is to divide the original multichannel data set into two halves. We denote the first half of the data, $i = 1, 2, \ldots, N/2$ by $x(t_i), y(t_i), z(t_i)$, and the second half, for $i = N/2, N/2 + 1, \ldots, N$, is displaced in time to $i = 1, 2, \ldots, N/2$ and denoted by $x'(t_i), y'(t_i), z'(t_i)$. Now $x'$ has dynamical structure when taken with $y'$ and $z'$, but its structure is, time step for time step, unrelated to the structure of $y$ and $z$. Nearest neighbors identified using coordinates $x(t_i), y(t_i), z(t_i)$ will not be true dynamical analogies, and can be expected to diverge rapidly. Thus the value of $\bar{a}(j)$ for the surrogate data $x'(t_i), y(t_i), z(t_i)$ will be much larger than for the original.

On the other hand, if the original data $x(t_i), y(t_i), z(t_i)$ came not from a dynamical system but from random behavior, then the spreading rate should be large for both the original and the surrogate data. Since we do not know a priori what is a large spreading rate, the comparison with surrogate data is essential.

Examples of average local divergence rates $\bar{a}(j)$ for true trajectories of
eqs. (8) and with surrogate channels are given in Table 1. Two disjoint segments each 5000 steps in length were extracted from a longer trajectory. The two segments were spliced into a single six-channel data set to facilitate surrogate substitutions. Various embeddings and surrogate substitutions were tried, with two cases each to give a crude estimate of the variance due to finite sampling of the attractor.

Case 1 is the complete natural embedding of the first segment with coordinates $x(t_i), y(t_i), z(t_i)$, and case 2 is the complete natural embedding of the second segment with coordinates $x'(t_i), y'(t_i), z'(t_i)$. In both cases the average local divergence rate $\bar{\lambda}(j)$ depends on $j$, the number of steps forward. For $j = 1$, the average rate is smaller than the expected long-term value of 0.011, and the variance due to finite sampling is relatively large. For increasing $j$, the variance due to sampling decreases; $\bar{\lambda}(j)$ increases to a maximum near $j = 4$ and then decreases to the expected long-term value at $j = 64$.

Cases 3 and 4 are the results of surrogate substitution in the $x$ coordinate; in case 3, $x'(t_i)$ was substituted for $x(t_i)$ in the first trajectory segment, while in case 4 the opposite was done. Since the trajectories $x'(t_i), y(t_i), z(t_i)$ and $x(t_i), y'(t_i), z'(t_i)$ are of course not trajectories of a dynamical system, it is an abuse of terminology to speak of local Lyapunov exponents. Nevertheless, computing the average local divergence rates as in cases 1 and 2, we find a large increase in magnitude. If we were presented with data of unknown origin, such an effect would be evidence for the significance of the $x$ coordinate in the dynamical structure of the trajectory.

Note that the effect of the surrogate channel upon the local divergence rates is greatest for $j = 1$, and becomes less pronounced as $j$ increases. When trajectory self-crossing inconsistent with dynamical structure occurs, the largest separations occur in the near term. Thus, when using surrogates to test for dynamical structure, the short-term divergence rates provide better diagnosis than the long-term rates, even though the short-term rates are not invariant quantities.

Cases 5 and 6 are likewise obtained by surrogate substitution, this time for the $z$ coordinate. Again the importance of this coordinate in the dynamical structure is confirmed, although it is somewhat less significant than the $x$ coordinate.

Cases 7 through 10 relate to the detection of dynamical structure using an incomplete set of phase space coordinates. Only $x$ and $y$ are used as trial phase space coordinates; we therefore do not expect the divergences, even for
large $j$, to approach the true Lyapunov exponent $\lambda_1$ in the three-dimensional phase space. Nevertheless, when a surrogate channel is substituted (in cases 9 and 10) for the $z$ coordinate in the partial $z, y$ embedding, the effect on local divergence rates is unmistakeable.

This suggests that it may be possible, using surrogate channel substitution, to detect dynamical structure in multichannel time series data, even if there are not enough data channels to fully embed the attractor, that is, there are not enough phase space coordinates to correctly identify good dynamical analogies.

Of course this prototype example of chaos only suggests what may happen with more complicated chaotic attractors in higher-dimensional phase space. The use of surrogate channels and local divergence rates has recently been tested on another model, a very low-order moist general circulation model devised by Lorenz [1984], which is equivalent to a system of 27 ordinary differential equations. Applying multichannel time series analysis to simulations of this model confirmed that surrogate channel substitution and diagnosis from short-term divergence rates can be used to detect dynamical structure in incomplete embeddings [Stewart, 1994].
A Paleoclimate Example

We now describe an example of how such methods might be applied to paleoclimate data. Here we must, at least for the present, abandon thought of any specific model which would describe dynamical structure of climate. Of course there are dynamical models of climate evolution, but the differential equations model the evolution in steps of hours or days, and nearby trajectories diverge in weeks or at most months. Here we wish to pose a different question: Is the evolution of climate governed by dynamical laws, known or unknown, which determine the state one or a few years forward given the present state, expressed in terms of annual or seasonal means.

We emphasize that we know of no persuasive argument that climate has such a dynamical structure (on yearly time scales) at all. In fact, the rate of divergence of weather conditions seems (at least superficially) to make this dynamical structure of climate an unlikely hypothesis. Nevertheless, it is not ruled out. If true, or even partly true, it would be of great consequence: as just one example, it might permit earlier detection of a greenhouse signal manifested as a departure in (or from) dynamical structure. In any case, we are simply posing this as a hypothesis, to be tested using multichannel time series analysis.

For our example, we shall use two paleoclimate proxy data series of very different origin. For a more thorough study, additional data series should of course be included. One series used here is the annual oxygen isotope ratio in cores from the Quelccaya ice cap in Peru published by Thompson and Mosley-Thompson [1989], extending from 1476 to 1984. The other series is a tree ring index developed by Briffa et al. [1992] from Fennoscandian trees, extending from A.D. 500 to 1980, and kindly furnished to the author by Prof. Briffa. These two data series were treated as possible phase space coordinates for dynamical structure, if it exists. The notion underlying this choice is that these two proxies would represent different and complementary modes or degrees of freedom of a hypothesized dynamical structure. Of course any climatological insight which would bear on the suitability of these or other data series as reflecting the state of the climate should be considered; see for example Cole et al. [1993]. For present illustrational purposes, we take these two series to represent the state of the art.

These two data series are analyzed on two different time scales. First, we test for dynamical structure on the yearly time scale. Both data series
exhibit large fluctuations from year to year; for this reason, a smooth evolution on yearly time scales as with differential equations is not an appropriate hypothesis, and instead we must suppose a system in which time passes in discrete units as in the iterated function (1). Second, we test for dynamical structure on multiyear time scales, taking running means to smooth the data; this gives somewhat less ragged trajectories which might correspond to a differential equation model. The numerical evaluation of local divergence rates is essentially the same in both cases, with one small difference: with the smoothed data we determine nearest neighbors by interpolating between successive data points, while with the raw yearly data, there is no interpolation.

In the previous example of synthetic data from a simulation, there was little harm in generating surrogates by cutting a long trajectory into two disjoint segments: small sample size was not a problem, as the simulation can always be extended. However, with limited and precious paleoclimate data, cutting the time series in half would seem extravagant: one would prefer to test the full length of available data against surrogates of equal length.

Another means of generating surrogate channels is by randomization of the given data. Substituting data from a random number generator is too crude: we want our surrogates to lack dynamical structure, but be able to pass superficially for the real data. One algorithm for achieving this is the following: take the complex Fourier transform of the data, then randomize the phases of the complex coefficients $a(\omega)$ in the frequency domain. In order that the inverse transform yield a real time series, the phases must be randomized under the constraint that $a(\omega) = a^*(-\omega)$. The inverse transform of data randomized in this manner will be a surrogate having the same power spectrum as the original data, but with any dynamical structure removed.

This is the procedure we adopt for generating surrogates for both the ice core and the tree ring data. It is not the only, or necessarily the best procedure; see Theiler et al. [1992], Takens [1993], and Kennel and Isabelle [1992] for further discussion.

Figure 2 shows local divergence rates $\bar{s}(j)$ for a two-dimensional trajectory whose coordinates are the Quelccaya oxygen isotope ratio and the Fennoscandian tree ring index; the values of $\bar{s}(j)$ are indicated by dots and connected with solid lines. Also shown are the divergence rates for three different surrogate substitutes for each coordinate, plotted as broken lines. In the upper plot, surrogates are substituted for the tree ring data. The values of $\bar{s}(j)$ for
surrogates fall very slightly above the values for the data at \( j = 1 \) and \( j = 2 \); we can easily imagine that with more than three surrogate surrogate substitutions, their range would include the \( \tilde{S}(j) \) of the original data. For larger \( j \), the \( \tilde{S}(j) \) for the data lie within the range for the three surrogates. In the lower plot, surrogates are substituted for the ice core data; all the \( \tilde{S}(j) \) for the original data are within the ranges for the surrogates. We conclude that these tests show no evidence of dynamical structure in the data on yearly time scales.

The same two time series were then smoothed and re-tested for local divergence on multiyear time scales. Seven-year running means were used, based on the presence of a strong 14-year component in spectrum of the ice core data. We also used eleven- and thirteen-year running means. When generating surrogates for smoothed data, we randomize phases of the raw data and then smooth, rather than randomizing smoothed data; this is in accord with the recommendation of Prichard [1994] to randomize before filtering. Typical results for the seven-year running means are shown in Figure 3, again using three surrogate substitutes for each channel. The divergence rates for the data again fall within or are very close to the range for the surrogates. We conclude that there is also no evidence for dynamic structure in the data on multiyear time scales.

The development and application of dynamical systems approaches to time series analysis is still in its infancy. In due course, with deeper insight into climate and perhaps some serendipity, these multichannel methods should achieve successes with climate data comparable to the recent success analyzing single time series [Abarbanel, 1995].

Acknowledgement. The author is indebted to Henry Abarbanel, Keith Briffa, and Caroline Isaacs for stimulating discussions. This work was supported by the CHAMMP initiative of the Office of Health and Environmental Research, U.S. Department of Energy under Contract No. DE-AC02-76CH00016.


LIST OF FIGURES

Figure 1. Time series and phase space projections of a trajectory of Roessler's equations (8), and phase projections of a trajectory obtained by substituting surrogate channel data $x'$ for $x$.

Figure 2. Average local divergence rate $\bar{s}(j)$ of a paleoclimate proxy data trajectory from 1476-1980; coordinates are oxygen isotope ratio from Quelccaya ice core, and Fennoscandian tree ring index. Above, paleoclimate trajectory and three trajectories with surrogate tree ring data; below, paleoclimate trajectory and three trajectories with surrogate ice core data.

Figure 3. Average local divergence rate $\bar{s}(j)$ of a paleoclimate proxy data trajectory from 1476-1980, smoothed by taking seven-year running means of oxygen isotope ratio and Fennoscandian tree ring index. Above, paleoclimate trajectory and three trajectories with smoothed surrogate tree ring data; below, paleoclimate trajectory and three trajectories with smoothed surrogate ice core data.
Average local divergence rates for paleoclimate data and surrogates
Average local divergence rates for smoothed paleoclimate data and surrogates