Chaos, Dynamical Structure and Climate Variability

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

Abstract. Deterministic chaos in dynamical systems offers a new paradigm for understanding irregular fluctuations. Techniques for identifying deterministic chaos from observed data, without recourse to mathematical models, are being developed. Powerful methods exist for reconstructing multidimensional phase space from an observed time series of a single scalar variable; these methods are invaluable when only a single scalar record of the dynamics is available. However in some applications multiple concurrent time series may be available for consideration as phase space coordinates.

Here we propose some basic analytical tools for such multichannel time series data, and illustrate them by applications to a simple synthetic model of chaos, to a low-order model of atmospheric circulation, and to two high-resolution paleoclimate proxy data series.

The atmospheric circulation model, originally proposed by Lorenz, has 27 principal unknowns; we establish that the chaotic attractor can be embedded in a subspace of eight dimensions by exhibiting a specific subset of eight unknowns which pass multichannel tests for false nearest neighbors. We also show that one of the principal unknowns in the 27-variable model – the global mean sea surface temperature – is of no discernible usefulness in making short-term forecasts.

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 chaos paradigm of dynamical systems theory raises the possibility that some of the irregular oscillations observed in the laboratory and in the natural world may be explainable as the workings of deterministic dynamics. Indeed, the deterministic rules governing the behavior of even large and complex natural systems may in some instances 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 comprises two essential ingredients: first, a state space or phase space whose coordinates are the state variables $x_1, x_2, ..., x_n$; these variables describe the state of the system at any instant of time, and the functions $x_1(t), x_2(t), ..., x_n(t)$ describe the system evolution over time. Second, a dynamical system possesses a dynamical rule which specifies completely and unambiguously for each state $X = \{x_1, x_2, ..., x_n\}$ 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 [1, 2].

In 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

$$X_{i+1} = F(X_i)$$

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

$$X_0 = \{x_1(t_0), x_2(t_0), ..., x_n(t_0)\}$$

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

$$\frac{dX}{dt} = \dot{X} = F(X)$$

that is,

$$\dot{x}_1 = f_1(x_1, x_2, ..., x_n)$$
$$\dot{x}_2 = f_2(x_1, x_2, ..., x_n)$$
$$\vdots$$
$$\dot{x}_n = f_n(x_1, x_2, ..., x_n)$$

The evolution of the system from an initial state

$$X_0 = \{x_1(t = 0), x_2(t = 0), ..., x_n(t = 0)\}$$
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
forward in time is the solution \( X(t) = \{x_1(t), x_2(t), \ldots, x_n(t)\} \) 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 \( f_1, f_2, \ldots, f_n \) 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 [1]. 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 \( X = \{x_1, x_2, \ldots, x_n\} \) 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 [3, 4] 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 [2, p. 221]. In many cases, dissipation acts even more strongly, reducing the long-term fluctuations to a subset of dimension much smaller that 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 [5].

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 \[6, 7\]. An excellent review of these methods has recently appeared \[8\]. Here we consider situations in which two or more concurrent time series of different variables are available for analysis. After describing the basic diagnostic of trajectory divergence, we discuss three applications: a prototype of chaos in a system with three-dimensional phase space; a low-order model of atmospheric circulation with 27 phase space coordinates; and an example involving high-resolution paleoclimate proxy data.

**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

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

(5)

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 \( \lambda_1 > 0 \) \[8\].
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

\[
s(i, j) = \frac{1}{j} \ln S(i, j)
\]

and its average

\[
\bar{s}(j) = \frac{1}{i} \sum_{i} s(i, j)
\]

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

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

This system was devised by Roessler [9] 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 \( \frac{1}{60} \ln 2 \), or roughly 0.011.
Figure 1: Time series and phase space projections of a trajectory of Roessler’s equations, and phase projections of a trajectory obtained by substituting surrogate channel data $x'$ for $x$. 
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 already been used for this type of diagnosis with time-delay embeddings, where only a single variable has been observed and recorded; see for example [10, 11]. Here we are considering multichannel data, so it makes sense to look at surrogate data in which only 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, $i = 1, 2, ..., N/2$ by $x(t_i), y(t_i), z(t_i)$, and the second half, for $i = N/2 + 1, ..., N$, is displaced in time to $i = 1, 2, ..., N/2$ 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 $a(j)$ 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 $\bar{s}(j)$ for true trajectories of eqs.(8) and with surrogate channels are plotted in Figure 2. 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.

The complete natural embedding of the first segment with coordinates $x(t_i), y(t_i), z(t_i)$, and the complete natural embedding of the second segment with coordinates $x'(t_i), y'(t_i), z'(t_i)$ are represented in Fig. 2(a) by the solid curves which give the average local divergence rate $\bar{s}(j)$ as a function of $j$, the number of steps ahead. These average rates range up to 0.020, which is of
Figure 2: Average local divergence rate $\bar{\sigma}(j)$ for trial embeddings of the Roessler band attractor in two and three dimensions; 5000-point trajectories sampled 60x per turn.
the order of the expected long-term value of 0.011. The difference between
the two samples is just visible for \( j = 1 \). For increasing \( j \), this difference
decreases. In both cases, \( \delta(j) \) increases to a maximum near \( j = 4 \) and then
decreases to the expected long-term value by \( j = 64 \).

The results of surrogate substitution in the \( x \) coordinate are shown in
Fig. 2(a) by the evenly broken lines. In one case, \( x'(t_i) \) was substituted
for \( x(t_i) \) in the first trajectory segment, while in the other case 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 before, 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.

The remaining two cases in Fig. 2(a), shown as unevenly broken lines,
are again obtained by surrogate substitution, this time for the \( z \) coordinate.
Again the importance of this coordinate in the dynamical structure is con-
firmed, although its significance is not so strong as that of the \( x \) coordinate.

Figure 2(b) relates to the detection of dynamical structure using an in-
complete 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 (broken curves)
for the \( x \) coordinate in the partial \( x,y \) embedding, the effect on local diver-
gence rates is unmistakable.

This suggests that it may be possible, using surrogate channel substitu-
tion, 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.
We now turn to a more challenging example.
We next apply similar techniques to a system of 27 ordinary differential equations introduced by Lorenz as a low-order model of atmospheric circulation [12]. The model is based on a two-layer quasi-geostrophic beta-plane approximation of circulation at middle latitudes, with moisture in the atmosphere and a shallow, non-circulating ocean. The principal unknowns are the atmospheric temperature $T$, stream function $\psi$, moisture content $W$, and sea surface temperature $S$; each field is represented horizontally by an expansion in seven orthogonal spatial modes numbered 0 through 6. A global constraint eliminates the zeroth mode $\psi_0$ of the stream function, leaving $4 \times 7 - 1 = 27$ unknowns. Lorenz found plausible values of the parameters for which the model exhibits sustained chaotic dynamics with nearby trajectories separating exponentially in time. Similar solutions exist over a range of parameter values [13].

This model has attracted interest because its long-term behavior is suggestive when interpreted as a toy model of natural climate variability. With no daily or yearly variation in solar forcing, and numerical solution advancing four time steps per nominal day, the model has oscillations with an average period of about two weeks. Figure 3 shows atmospheric temperature at a fixed location on this time scale. A second trajectory perturbed from the first by a very small amount at time zero diverges from the first on a similar time scale. And yet the global mean atmospheric temperature — the zeroth mode $T_0$ — fluctuates substantially on very much longer time scales. Figure 4 shows $T_0$ averaged yearly (i.e. over 360-day periods) for a thousand years.

The question has therefore been raised whether the long-term fluctuations in Fig. 4 correspond to an attractor, that is, whether in some appropriate phase space there is a vector field which determines the long-term changes of mean temperature. Any such vector field would not be the one given directly by the 27 ordinary differential equations, but would presumably emerge from them in some unspecified manner. This hypothetical dynamical structure on long time scales need not exist at all, but there is some evidence for its existence, including a red spectrum, dimension estimates [14], and an argument based on rescaled range analysis and the Hurst exponent [15].

Even on the faster time scale where deterministic dynamical structure certainly governs the dynamics, the Lorenz 27-variable model has considerable interest as a dynamical system in its own right, and poses considerable challenges to the analyst. The form of the ordinary differential equations is complicated, including algebraic constraints and highly nonlinear thermodynamic relations, and the numerical solution procedure combines the modal representation to resolve some terms and conversion to a finite spatial grid to resolve other terms. These complications make it unlikely that much can
Figure 3: Two solutions of the Lorenz 27-variable moist atmospheric circulation model diverging after a small perturbation at day zero.

Figure 4: Long-term fluctuations of the yearly averaged global mean atmospheric temperature in the Lorenz model.
be learned by analysis of the equations. For example, at least one unstable
equilibrium solution probably exists in the chaotic regime, but there is no
obvious way to locate it, let alone perform a local linear stability analysis.
In short, the only available tools of analysis are those which treat the numeri-
cally generated solutions as if they were produced by a black box. On the
brighter side, any tools which work for this system may also be useful for the
analysis of observed experimental data, even when a mathematical model of
the system is unavailable.

We shall undertake two different analyses of the Lorenz 27-variable model,
treating the model output as time series data. First we attempt to describe
the chaotic attractor on the fast time scale, interpreted as a toy model of
weather. We know there is dynamical structure, and aim to describe it
efficiently by making use of multiple channels of time series data, the 27
modal coefficient time series. We use estimates of local divergence rates and
comparison with surrogate channel data to find a subset of the 27 channels
which gives a faithful embedding of the chaotic attractor. The goal is to
identify dynamical structure in a challenging test case. In the process we
will obtain an unexpected bonus of useful information.

Second we apply similar methods to yearly averaged output of the 27-
variable model, interpreted as a toy model of climate. Here we seek evidence
to accept or reject the hypothesis of deterministic dynamical structure on
decaladal time scales in the model.

Behavior of the Atmospheric Circulation Model

Before embarking on the time series analysis, a survey of qualitative be-
havior was made by running the model a number of times with different
values of key parameters. Specifically, we varied the global mean solar forc-
ing $T_0^*$ and the vertical lapse rate $\lambda$. The purpose of this survey was twofold:
to ascertain the robustness of chaotic behavior, and to locate any transitions
from regular to chaotic dynamics which might yield some insight into the
chaotic regime.

With $T_0^*$ reduced to produce winter-like temperatures, the model was
found to have quasi-periodic dynamics, with a fast period of several days
and a slow period of more than a year. The fast dynamics correspond to
the circulation of a low pressure region in the atmosphere which is most
pronounced at higher latitudes. The slow dynamics correspond to a warmer-
than-average region in the sea gradually drifting around the globe. The fast
dynamics are most clearly seen in modes 4 and 5, particularly $T_4$ and $T_5$,
while the slow dynamics are clear in $S_1$ and $S_2$.

As the mean solar forcing $T_0^*$ is increased, the dynamics change first to
a mild chaos with very small fluctuations of $T_0$. Then a dangerous bifur-
cation [16] occurs as a threshold value of $T_0$ is passed: the system makes
a transient jump to a different chaotic attractor. While temperatures are still winter-like, the fluctuations in $T_0$ are larger, being now comparable in magnitude to the fluctuations observed at Lorenz's parameter values (which produce spring-like temperatures). Although this transition is a dangerous bifurcation, there is also evidence of a nearby transition to the spring-like chaos by type 3 Pomeau-Manneville intermittency [17, 16]. This suggests that the warmer, spring-like chaos may have attractor dimension much lower than 27. Further increasing $T_0^*$ to the spring-like conditions of Lorenz reveals no further obvious bifurcations.

In short, parameter variations showed that chaos with substantial long-term fluctuations of $T_0$ is reasonably robust, and yet not far from a transition to regular, quasiperiodic dynamics with fast and slow periods.

**Building an Embedding**

Henceforth we consider only the reference case with $T_0^* = 273K$, as proposed by Lorenz. A previous study [14] found this chaotic attractor to be characterized by just two positive Lyapunov exponents. This could be realized by an attractor in a phase space which is locally four-dimensional. Together with the transition scenarios mentioned above, this suggests that the chaotic attractor occupies a subset of phase space of dimension much smaller than 27. Our aim is then to find a reasonably small subset of the 27 modal coefficients which faithfully represent the chaotic attractor, that is, an attractor embedding in fewer than 27 dimensions. Because the model appears to have two widely different time scales, it seems likely that such a multichannel embedding may be more efficient than the popular time-delay reconstruction from a single channel of data, using a fixed size time delay for each added coordinate. Although reconstruction of phase space from a single time series is an invaluable tool when only one time series is available, it may not be preferred when multichannel data are available. We want to see what advantage can be gained from more complete multichannel data. In the process, we shall discover an unexpected property of the most interesting modal coefficients, the global means $T_0$ and $S_0$.

Shown in Figure 5 is the result of a typical 20-year simulation in the form of time series for six of the modal coefficients, modes 0, 1, and 4 of the atmospheric temperature $T$ and the sea surface temperature $S$. Mode $T_4$ consists predominantly of fast oscillations; the individual cycles of about two weeks duration cannot be distinguished on this time scale. Modes $S_0$ and $S_1$ show predominantly slow dynamics; note however that there is no evident correlation between $S_0$ and $S_1$. Atmospheric modes $T_0$ and $T_1$ parallel the corresponding sea surface modes, with additional fast oscillations superimposed.
Figure 5: Selected modal coefficients of a solution of the Lorenz moist circulation model over 15 simulated years.
Figure 6: A solution of the Lorenz moist circulation model projected onto the planes of modes 4 and 5.
Figure 7: A solution of the Lorenz moist circulation model projected onto the planes of modes 1 and 2.
From extensive visualization of solutions in both time and space-time (using purpose-built visualization software), it is apparent that the underlying structure of the dynamics in the chaotic regime is a fast circulation most evident in modes 4 and 5, and a slower circulation most evident in modes 1 and 2. Thus any choice of modal coefficients to capture the attractor dynamics should be capable of clearly representing a two-torus, topologically the product of two circles. None of the 27 modal coefficients is by itself an angular variable; that is, any one coordinate will take the same numerical value during the ascending and the descending phase of its cycle. To distinguish the ascending and descending phases of mode 1, the complementary mode 2 must be included. Thus the representation of each circular motion (=circulation) will require a pair of coordinates.

Figure 6 shows projections of a 20-year trajectory onto the planes of modes 4 and 5 for $\psi, T, W, S$. One of these pairs of coordinates should be suited to describe the fast circulation. A roughly circular motion is particularly evident in $T$ and $W$. Projections of the same trajectory onto the planes of modes 1 and 2 are shown in Figure 7. Here the rough circularity of the slow circulation is obscured by the superimposed fast wiggles, but it becomes more evident if one watches the trajectory being drawn very rapidly.

It is of course possible to describe a circular motion with one angular variable, for example an angle in the $(T_4, T_5)$ plane. We have decided against this tactic for two reasons. First, we limit our analysis to picking coordinates from the 27 given, without combining in any way; while new coordinates combining the 27 might be very useful, they are beyond the scope of this study, which is our first effort at mutlichannel embedding. Second, we noted after careful consideration that an algorithm central to our analysis -- the k-d tree search for nearest neighbor in phase space [18, 8] -- is awkward to implement for angular variables. Considering the extra program logic required to identify $\theta = 2\pi$ with $\theta = 0$ for an angular variable in the k-d tree search, it was judged a more reliable and no less efficient strategy to require that any angular variable $\theta$ be represented by two displacements, analogous to $\sin \theta$ and $\cos \theta$.

Therefore we begin building an embedding of the attractor by considering all four mode 1,2 pairs and all four mode 4,5 pairs. We may be paying a penalty for this, using pairs of coordinates where perhaps one (angular) coordinate would suffice. According to topological embedding theorems [7, 19] any method of reconstruction in Euclidean space (including the time-delay method) must expect to pay a similar penalty.

The average local divergence rates of a 20-year trajectory using all eight complementary pairs of coefficients are shown in Figure 8(a). The pair $(S_1, S_2)$ has the lowest divergence rates, and this would seem to be the best
Figure 8: Average local divergence rate $\bar{s}(j)$ for trial embeddings of the Lorenz 27-variable model in two dimensions: (a) trial embeddings; (b) comparisons with surrogate channels.
choice to begin recovering the dynamical structure and near-term predictability of the attractor from data. However, the divergence rate of a subset of coordinates does not indicate the forecastability of all coordinates, but only the forecastability of the coordinates in the subset. Good forecasts for the pair \((S_1, S_2)\) are easier because \(S_1\) and \(S_2\) change more slowly. One might imagine that a better diagnostic would measure divergence in 27 dimensions from nearest neighbors in two dimensions; but this strategy has problems, as we shall see below. Instead we propose the following: include a coordinate in the embedding if it yields low divergence rates, and yields separation rates lower than those with surrogate channel data substituted.

In Fig. 8(b) divergence rates for \((S_1, S_2)\) are compared with pairs \((S_i', S_j)\) having surrogate data substituted for \(S_1\). The surrogate data were obtained from the same coefficient of another independent simulation using different starting conditions. Two different series of surrogate channel data were tested; one could imagine statistical tests on larger samples, but in this case the evidence is clear: both surrogates yield lower divergence rates than the original pair \((S_1, S_2)\). Thus the apparent advantage of \((S_1, S_2)\) in Fig. 8(a) is due to the persistence of \(S_1\) and \(S_2\), and not to the dynamical information they contain.

The second candidate from Fig. 8(a) is the pair \((S_4, S_5)\); this pair is also compared with surrogates in Fig. 8(b). Although \((S_4, S_5)\) diverge slightly less than with a surrogate \(S_i'\), the difference is not greater than the variation among the two surrogate samples. Referring back to the time series in Fig. 5, we see that \(S_4\) closely parallels \(S_1\) with some superimposed fluctuations; the favorable divergence of \((S_4, S_5)\) is again due more to persistence than to dynamical information. We therefore pass over \((S_4, S_5)\) and consider the third candidate in Fig. 8(a), which is \((T_4, T_5)\). In this case, surrogates are clearly worse, as seen in Fig. 8(b). We therefore accept \((T_4, T_5)\) as the first pair of coordinates for embedding the attractor. Referring back to Fig. 6, we see that among the mode 4,5 pairs, \(T\) has the greatest visual coherence, that is, its self-intersections are on average more oblique than those of other pairs.

Next we consider \((T_4, T_5)\) in a four-dimensional embedding with each of the remaining pairs of modes 4,5 or modes 1,2. As seen in Figure 9(a), adding the pair \((S_1, S_2)\) yields the lowest divergence rates. As before we check this by substituting surrogate channel data \(S_i'\) this time joined with \(S_2, T_4, T_5\). With the fast oscillations included, the dynamical information in \(S_1\) is now apparent. The added pair \((S_1, S_2)\) as third and fourth coordinates is now better than the four surrogates shown in Fig. 9(b). We accept \((S_1, S_2)\) into our embedding with \((T_4, T_5)\).

If we had omitted the surrogate channel test, and had accepted \((S_1, S_2)\)
Figure 9: Average local divergence rate $\bar{s}(j)$ for trial embeddings of the Lorenz 27-variable model in four dimensions: (a) trial embeddings; (b) comparison with surrogate channels.
as the first coordinate pair, it turns out that \((T_4, T_5)\) would have been chosen as the second pair. So in this instance the surrogate test was unnecessary. Nevertheless, we believe that the only safe strategy is to include the surrogate channel test, even when all channels of data are known to come from a single deterministic dynamical system. Support for this claim will appear in the next stage of coordinate selection.

Also note that \((S_1, S_2)\) failed the surrogate test standing alone, although they passed when combined with \((T_4, T_5)\). This differs from the results with the simple Roessler attractor, where every possible pair of coordinates passes the surrogate test even though the third coordinate is not used. Thus in at least some cases, a partial embedding or projection of a dynamical trajectory may fail to show signs of deterministic dynamical structure when diagnosed with average local divergence rates and tested against surrogate channel substitutions. It may be that this failure is a consequence of the two very different time scales in the Lorenz 27-variable model.

We next consider each of the remaining pairs of modes 1,2 or modes 4,5 with \((T_4, T_5, S_1, S_2)\) in six-channel embeddings. Note that since we have already accepted two pairs of coordinates which should capture an underlying two-torus of fast and slow circulations, it is no longer strictly necessary to test channels in pairs. Our tactic here is to try one more pair, and then consider channels singly. As seen in Figure 10(a), the best pair to add is now \((T_1, T_2)\); this pair passed a surrogate test (not illustrated). We then substituted coordinates singly into the six-channel set \((T_4, T_5, S_1, S_2, T_1, T_2)\). These tests were not exhaustive, but none of the single channel substitutions tested gave better divergence rates. In one case, divergence rates almost as low were obtained, by substituting the global mean sea surface temperature \(S_0\) for \(S_1\). However, surrogate tests in Fig. 10(b) showed that although \(S_1\) is better that \(S'_1\), \(S_0\) is not significantly better than \(S'_0\).

Referring again to Fig. 5, we note that \(S_0\) varies slowly, on a time scale like that of \(S_1\). But \(S_1\) passes the surrogate channel test in four- and six-channel partial embeddings. So the failure of the surrogate test for \(S_0\) in Fig. 10(b) has a different meaning from the failure of the surrogate test for \(S_1\) in Fig. 8(a). We interpret Fig. 10(b) as indicating that the level of the sea surface temperature \(S_0\) is not dynamical information usable for (short-term) forecasting. Of course \(S_0\) is generated by solving a deterministic model, so in some sense it must contain dynamical information. Indeed, when the time derivative \(dS_0/dt\) is used as a coordinate, it does pass the surrogate channel test (although its divergence rates are worse than those of \(S_0\)). In other words, these simple surrogate channel tests suggest that while changes in \(S_0\) contain dynamical information usable for short-term forecasting, \(S_0\) itself does not.
Figure 10: Average local divergence rate $\bar{\delta}(j)$ for trial embeddings of the Lorenz 27-variable model in six dimensions: (a) trial embeddings; (b) comparisons with surrogate channels.
This becomes understandable if one thinks of the sea in the Lorenz model as a thermal buffer, each day changing its temperature slightly by interaction with the atmosphere. Each daily change is small and deterministic, but the net level of $S_0$ depends on many daily changes, accumulated over times longer than the forecast horizon implied by the size of the largest Lyapunov exponent. In the process of accumulating these changes over long times, the deterministic origin of each daily change is lost, so the level of $S_0$ is no help in forecasting.

The apparent lack of dynamical information in $S_0$ has several important consequences. If one wished to reconstruct this chaotic attractor from a single time series, it would seem unwise to choose $S_0$. Furthermore, since $T_0$ parallels $S_0$ (with superimposed fast oscillations), $T_0$ also seems a poor choice; using the difference $T_0 - S_0$ rather than $T_0$ would probably enhance the dynamical information content for the purpose of short-term forecasting.

Moreover, even when using multichannel data as phase space coordinates to study this attractor (for example, to estimate Lyapunov exponents), it would seem that $S_0$ should not be used as a coordinate, and $T_0$ should be replaced by $T_0 - S_0$.

Finally, because the yearly average of $T_0$ is nearly identical with the yearly average of $S_0$, it becomes very difficult to imagine how the long-term fluctuations of $T_0$ could possibly possess any dynamical structure usable for forecasting. Nonetheless, we shall not abandon our plan of testing this hypothesis directly. But first we shall complete the multichannel embedding of the short-term dynamics.

Completing the Embedding

The best six-dimensional embedding achieves local divergence rates of about 0.025 per time step, corresponding to doubling in about 28 time steps or one nominal week. This is comparable to the time scale for the growth of perturbations, as illustrated in Figure 3. So far we have considered how to choose successive additions to the embedding coordinates; we now need a means of determining when the embedding is complete.

One popular approach is to look for saturation of an invariant: when an added coordinate no longer changes the value of an invariant quantity such as the largest Lyapunov exponent. Our crude estimates of local divergence rate are meant to approximate the largest Lyapunov exponent for sufficiently large $j$, so we could look for $\tilde{s}(j)$ to stop changing. However this approach is not practical because extraordinarily long time series would be required to make $\tilde{s}(j)$ with fixed $j$ behave as an invariant.

In any case, a more direct determination of embedding completeness can be calculated using false nearest neighbors [8]. The idea is to de-
termine whether points which are close neighbors in an embedding of dimension \( d \) remain close in a higher dimension, or are separated by the additional coordinate. In the latter case, the \( d \)-dimensional close neighbors only appear close due to projection to a dimension which is too low. The algorithm for this is as follows. Consider a \( d \)-dimensional embedding with coordinates \( Y^d = (y_1, y_2, \ldots, y_d) \). For each observation \( Y^d(t_i) \), find the nearest neighbor in phase space \( Y^d(t_{N(i)}) \). Then augment the embedding coordinates to \( Y^{d+1} = (y_1, y_2, \ldots, y_d, y_{d+1}) \), and compare the new distance \( |Y^{d+1}(t_{N(i)}) - Y^{d+1}(t_i)| \) with the former distance \( |Y^d(t_{N(i)}) - Y^d(t_i)| \). Note that the same nearest neighbor \( N(i) \), determined in \( d \) dimensions, is used in both cases; only the distance formula is affected by the added coordinate. If the ratio

\[ \frac{|Y^{d+1}(t_{N(i)}) - Y^{d+1}(t_i)|}{|Y^d(t_{N(i)}) - Y^d(t_i)|} \]  

(9)

of new to old distance is large, say greater than 10, then the added coordinate has separated false nearest neighbors; otherwise the nearest neighbor is considered a true near neighbor. If the nearest neighbor is not false for any \( t_i \), then the added coordinate is redundant.

There is also a second criterion for declaring a nearest neighbor to be false; this criterion is needed when the deterministic nature of the time series is in question. We shall discuss this second criterion below when we analyze the long-term dynamics; for the present purpose, only the criterion based on the distance ratio (9) is required. (In fact we evaluated the second criterion and confirmed that it is never satisfied in analyzing the short-term dynamics.)

The computation of false nearest neighbors has been used successfully with time delay embeddings; note that in this case coordinate \( d + 1 \) should be the one with the least delay. The extension of the false nearest neighbor algorithm to multichannel embeddings is immediate. The only difference is that for a given \( d \)-dimensional embedding, every remaining channel should be tested as a possible \((d+1)\)st coordinate. If each remaining channel yields no false nearest neighbors, then the embedding is complete.

Figure 11 shows the results of various false nearest neighbor computations using multichannel embeddings, and for the sake of comparison, a time delay embedding using only the \( T_4 \) time series. The numbers are expressed as percentages; for 60-year simulations the number of observations is 86400. Based on experience which shows that true nearest neighbors are occasionally identified spuriously as false, we treat fewer than 0.01% false nearest neighbors (which in this case means a count of 8 or fewer) as essentially zero.

For dimensions 1 through 5, Fig. 11 shows the effect of adding (or removing) each channel in succession from our best six-channel embedding. Thus for example the number of false nearest neighbors in the multichannel
case for dimension 5 represents the result of adding $T_2$ to the five channels $T_4, T_5, S_1, S_2, T_1$. For dimensions 1 through 5, we have not re-evaluated alternative embeddings by comparing false nearest neighbors, although this could certainly be done. Here we accept the results of the comparisons based on local divergence; the false nearest neighbor percentages for dimensions 1 through 5 are shown for the sake of completeness. Our analytical use of false nearest neighbors is in choosing coordinates 7 and 8, corresponding to dimensions 6 and 7 in Fig. 11.

When a coordinate is added to the six-channel embedding, a number of choices yield significant false nearest neighbors. The greatest number (about 0.4%) is produced by adding $W_6$. If we select this as coordinate 7, and then check for false nearest neighbors produced by an eighth coordinate, only $W_3$ yields a significant number of false nearest neighbors (about 0.1%). From this it follows that if we now add any remaining coordinate to the eight-dimensional embedding $T_4, T_5, S_1, S_2, T_1, T_2, W_6, W_3$, only an insignificant number of false nearest neighbors will be generated. We therefore consider this eight-dimensional embedding to faithfully represent the dynamics on the attractor. According to the embedding theorems, eight is the number we may expect for an attractor which has two positive Lyapunov exponents and is therefore locally four-dimensional (and assuming the global dimension
For comparison, Fig. 11 also shows false nearest neighbors for time delay embeddings of coordinates $T_4$. A basic delay interval of 16 steps was chosen, which is $1/4$ of the approximate period of fast oscillations. Note that even at dimension 9 the number of false nearest neighbors is still significant. This would seem to suggest that the attractor has more than four dimensions. However, the embedding theorems states that eight should suffice in principle for a four-dimensional attractor: it may nevertheless fail in practice. We suspect that in this instance, time delay embedding does indeed fail to embed a four-dimensional attractor in eight dimensions. The reason for this can be found in the two widely differing time scales in the model: the fast oscillation is well sampled with a delay of 16 steps, but even $10 \times 16$ does not approach the characteristic time scale of the slow oscillation.

In sum, we have used local divergence rates of nearest neighbors, comparisons with surrogate channel substitutions, and false nearest neighbors to select a subset of 8 of the 27 time series which faithfully represents the chaotic attractor, and can be used to make short-term forecasts. As a significant by-product of this analysis, we learned that one channel, the global mean sea surface temperature $S_0$, is not directly useful for identifying dynamical analogs and making short-term forecasts; if $S_0$ is to be used for this purpose, one must first differentiate it with respect to time.

Long-term Dynamical Structure in the Model

We now turn to the question of whether yearly averages of the Lorenz 27-variable model have dynamical structure on annual or decadal time scales. This question has already been discussed in the literature. Nese and Dutton [14] computed the correlation dimension from time-delay embedding using the times series $T_0$ and a fundamental delay of 12 years, determined from a mutual information algorithm [20, 8]. The computed dimension was about 5.8. However, as noted by Ruelle [21], the correlation dimension will saturate and give an incorrect low dimension for an insufficiently long time series, even if the data are truly random. Ruelle estimates the saturation value to be $2 \log_{10}(N)$; Nese and Dutton used 10,000 years of data with a delay of 12 years, so $N = 10,000/12$, and the saturation value is about $-5.8$! No other evidence of dynamical structure on the long time scale was presented in [14].

McDonald [15] has analyzed annual averages of $T_0$ from the Lorenz model using rescaled range analysis to obtain a persistence index for various time scales. He concluded that for periods less than 20 years, the Lorenz model can be modeled as a Brownian motion with independent increments. For periods from 20 to 400 years the persistence index showed partial coherence. McDonald conclude with appropriate caution that predictions on these time
Figure 12: Percentages of false nearest neighbors for time delay embedding of the yearly averaged global mean atmospheric temperature $T_0$ in the Lorenz model.

scales might be possible; that is, the evidence is consistent with dynamical structure on these time scales.

False nearest neighbors can also be used to test for dynamical structure. As noted above, this requires a second criterion to identify false nearest neighbors [8]. The reason for this is that when a finite length time series of random data is embedded in successively higher dimensions, eventually the nearest neighbor of any point will be far enough distant that it becomes virtually impossible to increase the distance by a factor of ten. Thus the second criterion for declaring a nearest neighbor false is that if its distance is larger than some preset value, we don’t care what the added dimension does, the nearest neighbor is automatically counted as false. With proper normalization, this will never happen for deterministic data. We rescale each channel to the interval $[0,1]$ and take the threshold for this second criterion to be $0.02d^{1/2}$.

Although we would like to use multichannel analysis, the annual averaging of coefficients in the Lorenz model reduces the variance of almost all coefficients to a tiny fraction of their original variance. Only the global means $T_0$, $W_0$, and $S_0$ retain substantial variance after taking yearly averages, and all three of these are essentially the same time series. Therefore we have little
choice but to consider time delay embeddings. We have tried a delay of 10 years, as suggested by the mutual information calculation of [14], and also 20 years as indicated by the rescaled range analysis of [15]. The time series used was $T_0$, as in both previous studies.

The results of these false nearest neighbor calculations are shown in Fig. 12. The increase following the minimum at dimension 7 is the typical indication of data without deterministic dynamical structure. The false nearest neighbors counted for dimensions 1 through 6 are all generated by the first criterion for falseness; thereafter, most false nearest neighbors come from the second criterion. We conclude that the evidence from rescaled range analysis, while tantalizing, is not borne out by the more direct test for dynamical structure using false nearest neighbors.

A PALEOCLIMATE EXAMPLE

Finally we show by example 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 elaborate dynamical models of climate evolution, but when atmospheric dynamics are included, the differential equations model the evolution in steps of hours or days, and nearby trajectories diverge in weeks or at most months, as in the low-order Lorenz model. Here we wish to pose a different question, analogous to the hypothesis of long-term dynamical structure in the Lorenz model: 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 positing this as a hypothesis, to be tested using multichannel time series 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.

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 [22], extending from 1476 to 1984. The other series is a tree ring index developed by Briffa et al. [23] 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 [24]. For present illustrational purposes, we take these two series to represent the state of the art.

These two data series are analyzed on two slightly 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 examples of synthetic data from simulations, 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
Figure 13: Average local divergence rate $\bar{\alpha}(j)$ of a paleoclimate proxy data trajectory from 1476-1980. Phase space coordinates are oxygen isotope ratio from Quelccaya ice core, and Fennoscandian tree ring index. Above, $\bar{\alpha}(j)$ vs. $j$ for the paleoclimate trajectory and three trajectories with surrogate tree ring data; below, the paleoclimate trajectory and three trajectories with surrogate ice core data.
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 [10, 11, 25] for further discussion.

Figure 13 shows local divergence rates $\bar{z}(j)$ for a two-dimensional trajectory whose coordinates are the Quelccaya oxygen isotope ratio and the Fennoscandian tree ring index; the values of $\bar{z}(j)$ 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 $\bar{z}(j)$ 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 $\bar{z}(j)$ of the original data. For larger $j$, the $\bar{z}(j)$ 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 $\bar{z}(j)$ 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 [26] to randomize before filtering. Results for the running means were similar to the unsmoothed results. 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 [3].

Acknowledgement. The author is indebted to Henry Abarbanel, Keith Briffa, Caroline Isaacs and Ed Lorenz for stimulating discussions. This work was supported by the CHAMMP initiative of the Office of Health and En-
References


