PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Search for a Low-Dimensional Characterization of a Local Climate System

Anupam Sahay and K. R. Sreenivasan

Phil. Trans. R. Soc. Lond. A 1996 **354**, 1715-1750
doi: 10.1098/rsta.1996.0076

The search for a low-dimensional characterization of a local climate system

By Anupam Sahay and K. R. Sreenivasan

Mason Laboratory, Yale University, New Haven, CT 06520-8286, USA

Contents

Along with the computation of attractor dimension via the Grassberger–Procaccia method and the nearest neighbour algorithm, a variety of phase space tests are used to search for low-dimensional characterization of daily maximum and minimum atmospheric temperature data (*ca.* 25 000 points each, spanning about a 70-year period). These tests include global and local singular value decompositions, as well as others for uncovering nonlinear correlations among amplitudes of the global singular vectors and for recognizing determinism in a time series. The results show that a low-dimensional characterization of the temperature data is *unlikely*.

1. Introduction

Understanding the dynamics of the weather and the climate, with the ultimate application to prediction, is a major research challenge in geophysical fluid dynamics. Since these systems involve complex interactions of numerous subsystems (such as

Phil. Trans. R. Soc. Lond. A (1996) **354**, 1715–1750
Printed in Great Britain

the atmosphere, oceans, biosphere and so forth) a complete *ab initio* analysis is well out of reach. The best one can do is to develop models with adequate predictive power. A consistent feature of weather/climate data is that they are aperiodic and their deviations from periodicity cannot be explained by conventional linear models of time series analysis. With this as the background, three types of models have evolved. (1) A system perturbed by stochastic noise. Here, one can only hope to predict the future weather or climate in a statistical sense. (2) A deterministic system (chaotic or non-chaotic) with many degrees of freedom: in this case, accurate prediction (short term if chaotic) would be possible in principle. However, building the model would be rather unfeasible if detailed mechanisms are not known. (3) A chaotic (deterministic) system with only a few degrees of freedom; although predictability would be limited to short times, modelling without a knowledge of all details would in principle be feasible.

A paradigm that has been discussed in the literature over the past decade is the so-called *weather/climate attractor* conjecture, which states that weather/climate systems have low-dimensional attractors. Several studies corroborating this conjecture have appeared (Nicolis & Nicolis 1984; Fraedrich 1986; Tsonis & Elsner 1990; Keepenne & Nicolis 1989; Maasch 1989; Sharifi *et al.* 1990). In all these reports the Grassberger–Procaccia procedure (Grassberger & Procaccia 1983) for estimating the correlation dimension (D_2) of an attractor was applied to several data sets that are representative of weather or climate variations. A summary of the type of data used, the number of points in the data strings and the estimated dimension is given in table 1. Reported values have typically fallen between 3 and 8. There have also been reports negating the *weather/climate attractor* conjecture (Grassberger 1986; Tsonis & Elsner 1990; Lorenz 1991; Zeng *et al.* 1992). Part of the skepticism arises from the realization that dimension estimates spuriously saturate for short data records, leading to false conclusions about finite dimensional attractors. Several saturation limits have been obtained on the basis of heuristic arguments (Procaccia 1988; Smith 1988; Eckmann & Ruelle 1992). Eckmann & Ruelle show that the saturation dimension is of the order $2\log_{10} N$, where N is the number of points in the time series, but it must be noted that this estimate assumes data of good precision and scaling only over a decade. For the estimated D_2 in a physical experiment to be reliable, it must be well below this saturation limit. In view of such data requirements (both quantity and quality) most of the dimension estimates cited in table 1 are not satisfactory. Apart from the spurious saturation of dimension estimates, there are several other systematic errors (due to discretization, geometry of the attractor, edge effects and so on) which make it difficult to obtain reliable dimension estimates from experimental data. For a catalogue of potential difficulties, see Theiler (1990).

Before describing the specific goals of this work, it seems useful to put it in context by addressing the following two questions: (1) What does one mean by a 'weather attractor' or 'climate attractor'? (2) What dimension values can be considered low? *Weather* refers to the localized state of the atmosphere (involving temperature, humidity, wind velocity, and so forth), with time scales ranging from a few hours (e.g. thunder storms) to a few weeks (e.g. planetary-scale waves in the atmospheric circulation). *Climate*, on the other hand, refers to the long-term state (a few decades to millions of years) of the atmosphere and is usually characterized by representative aggregates (such as suitably defined averages) of weather systems. Strictly speaking, the relevant phase space for weather and climate attractors would be the space of all global fields of atmospheric temperature, humidity and wind velocity and so forth.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 1. *Summary of the reports of low-dimensional estimates of D_2 for weather/climate data* (δ^{18}O is the ratio of the rare ^{18}O and the common ^{16}O isotope. To a first order, it is a proxy of global ice volume (Fraedrich 1986).)

references	data type	approximate data points	estimated D_2
Nicolis & Nicolis (1984)	δ^{18}O	500	3.1
Fraedrich (1986)	local surface pressure	5500	3–4
	relative sunshine duration	11 000	3–4
	zonal wave amplitude	3500	3–4
	δ^{18}O	200	3–4
Keppenne & Nicolis (1989)	500 mbar geopotential height	9000	7.5
Tsonis & Elsner (1989)	vertical wind velocity	4000	7.3
Maasch (1989)	δ^{18}O	200	4–6
Sharifi *et al.* (1990)	rainfall	4000	4

To study the properties of the attractor one needs to resolve the smallest time scale of interest and encompass enough of the largest scales. For example, the shortest time scale of interest for the weather attractor would be an hour (say), with the largest time scale of the order of several weeks. For climate attractors, typical short and large time scales could be a year and few million years (associated with ice ages), respectively. Data spanning such large periods of time, although available in some *proxy* form (Nicolis & Nicolis 1984), are unfortunately very limited in extent and have the further drawback of not resolving time scales on human experience. On the other hand, data spanning a few decades or centuries, with resolution on the order of a day, are easily available; however, these data are not long enough to encompass features of the climate attractor. These conflicting demands between the resolution and the record length may appear to make the situation for the climate quite hopeless. However, the issue that makes further study meaningful is that climate phenomena of disparate time scales are reasonably stationary with respect to their own characteristic time scales. Accordingly, parts of the 'global' climate attractor can be thought of (in very loose terms) as 'local' attractors, which can then be studied as objects of interest in themselves. It is in this sense that the term 'local climate attractor' is interpreted in this paper. Since the 'local' attractors might indeed be quite different depending upon the nature of different data sets (time scales and sampling locations), the issue is really about the existence of *some* climate systems which have low-dimensional characterizations.

With regard to the dimension of the weather/climate attractors, it must be said first that the phase space is infinite dimensional in the strict sense. It is generally the case, however, that in dissipative systems the dimension of the attractor is less than that of the phase space in which it resides. Recent work has shown the existence of finite-dimensional (Hausdorff dimension) attractors in dissipative systems with infinite-dimensional phase space (Mallet-Paret 1976; Foias & Témam 1979; Ruelle 1982; Constantin *et al.* 1985). This gives some reason to believe that the conjecture about low-dimensionality of weather/climate attractors may not be absurd. The issue of what constitutes 'low-dimensionality' should be discussed in the light of the fact

that studying systems of dimension greater than ten, say, is fraught with practical difficulties. Only smaller dimensions could therefore be considered usefully 'low'.

The goal of this paper is to assess the feasibility of low-dimensional characterization of atmospheric temperature data. We examine data sets consisting of single-point daily maximum and minimum temperature ($T_{\max}$ and $T_{\min}$) in the atmosphere spanning about 70 years. We are thus looking at a local climate system with the largest characteristic time of the order of a decade. The use of daily maxima or minima in reconstructing the attractor is believed to produce a smoothed version of it without introducing undue distortions on time scales of interest; the primary reason for this belief is that the smoothing works on a time scale of the order of a day, which is almost certainly finer than the finest resolution required for climate attractors. This issue will be revisited in the text. Along with dimension estimates, we examine several characterizations of temperature data (all computed from the phase space reconstruction of the attractor); these include the global singular value decomposition (SVD) (Broomhead & King 1986), local SVD (Broomhead *et al.* 1987), an extension of SVD based on calculating nonlinear correlations among the amplitudes of the global singular vectors (Healey 1994) and a test for recognizing determinism in a time series (Wayland *et al.* 1993). SVD provides an 'optimal' basis for the data in phase space and the number of 'dominant' basis vectors can be taken as an upper bound to the dimension. Global SVD is calculated using the whole data set and does not always indicate low-dimensionality because nonlinear correlations among the amplitudes of the global basis vectors 'excite' new basis vectors thereby spuriously increasing the total number of 'dominant' vectors. Such correlations could be reduced by examining local bases (local SVD) or could be uncovered by correlation fits using spline-like functions. The test for recognizing determinism in a time series needs some explaining since one is looking here for a relation between a statistic which depends upon the dynamics on the attractor and its dimension, which is a geometric (measure-theoretic) property. This relationship is not known in general, although it is believed that a deterministic attractor of high dimensionality would appear stochastic in a low-dimensional reconstruction. On the other hand, if the experimental data appear deterministic in a low-dimensional reconstruction then they are likely to have a low-dimensional characterization.

It should be pointed out that, so far, only dimension calculations and global SVD have been applied to weather/climate data. The application of empirical orthogonal decomposition and principal component analysis (variants of global SVD) in studying weather/climate data can be traced back to Lorenz (1956). There have been numerous studies since then, making use of singular vectors as expansion bases, and as filters to identify the mean drift and important structural components of the data. Implications of global SVD on dimensionality have been explored only recently. Fraedrich (1986) and Vautard & Ghil (1989) examined proxy data indicative of global ice ages and used the amplitudes of singular vectors and statistical tests to identify the dominant vectors. These authors reported, respectively, on the possibility of a four-dimensional and a ten-dimensional characterization of their data sets. Further phase space analysis of such small, non-stationary data sets (100–500 points) would, of course, be devoid of much meaning. Indeed, from the perspective of weather/climate attractors, the relevance of the reported results is unclear.

The outline of the paper is as follows. In § 2, the data sets of daily maximum and minimum temperature are given preliminary probing by displaying their time trace, spectra, histograms and delay-time plots. Phase space reconstruction of dynamical

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 2. *Summary of the breaks in $T_{\max}$ and $T_{\min}$ data*

length of break	1	2	4	5	6	11	20	23	30	31	43	53	91
number of breaks	11	5	1	1	1	1	1	1	1	1	2	1	1

systems using time series of a single variable is discussed in § 3. Since the data are likely to be noisy we have used a noise reduction technique based on phase space reconstruction (Kostelich & Yorke 1990) as a pre-processing step. The methodology and results of this pre-processing are described in § 4. Briefly, the noise reduction scheme works by modelling the local dynamics on the attractor by linear maps and then adjusting the data to make them more consistent with the local linear maps. The noise reduction *might* be considered effective without being detrimental if this adjustment moves points in phase space by small distances (say less than 10% of the attractor extent, which is defined as the maximum inter-point distance on the reconstructed attractor). It turns out that the $T_{\max}$ and $T_{\min}$ data require larger adjustments. Hence, the pre-processing step was deemed ineffective for these data and the rest of the analysis was done with the original data. In § 5 two practical methods of calculating dimensions, namely the Grassberger–Procaccia algorithm (Grassberger & Procaccia 1983) for the correlation dimension (D_2) and the nearest neighbour algorithm (Badii & Politi 1985) for the information dimension (D_1), are described. Estimates of D_1 for the temperature data are reported (but D_2 could not be estimated due to limitations of the data). The results do not indicate low-dimensionality. However, the dimension estimates with a noisy and small data set can only be taken as a rough guide. The SVD techniques and the test of determinism of a time series are introduced in § 6 to explore further the issue of low-dimensionality. These techniques are evaluated in the first half of § 7 via their performance on control data (the x-variable of the Hénon map and the local maxima of the y-variable of the Lorenz flow). In the second half of that section, we report the results of these tests for the temperature data. Summary and conclusions are presented in § 8.

The principal conclusion of the investigation is that a low-dimensional characterization of the atmospheric temperature data is *unlikely*.

2. The data

The time series of daily maximum and minimum temperature, sampled at Berkeley (California), were obtained from the US Climatic Data Center, Asheville, North Carolina. The temperature data, recorded in degrees Fahrenheit to the nearest integer, cover a time span of about 70 years. There are some breaks in the data (identical for $T_{\max}$ and $T_{\min}$) as shown in table 2. For the sake of efficiency we have decided to treat the whole data set (*ca.* 25 000 points) as contiguous. Any attempt to fill the breaks using some interpolation would introduce extraneous correlations and is to be avoided. The number of 'bad points' in the reconstructed phase space (i.e. the number of points affected by these breaks) is at the worst about 4% of the total number of points. The calculation of the number of 'bad points' and its implications on the phase space reconstruction along with those of using a non-standard reconstruction procedure, are discussed in the next section. Among the various data displays shown in this section, the breaks marginally affect the power spectrum and the delay-time plots. The frequency axis in the power spectrum has to be rescaled (multiplied by

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

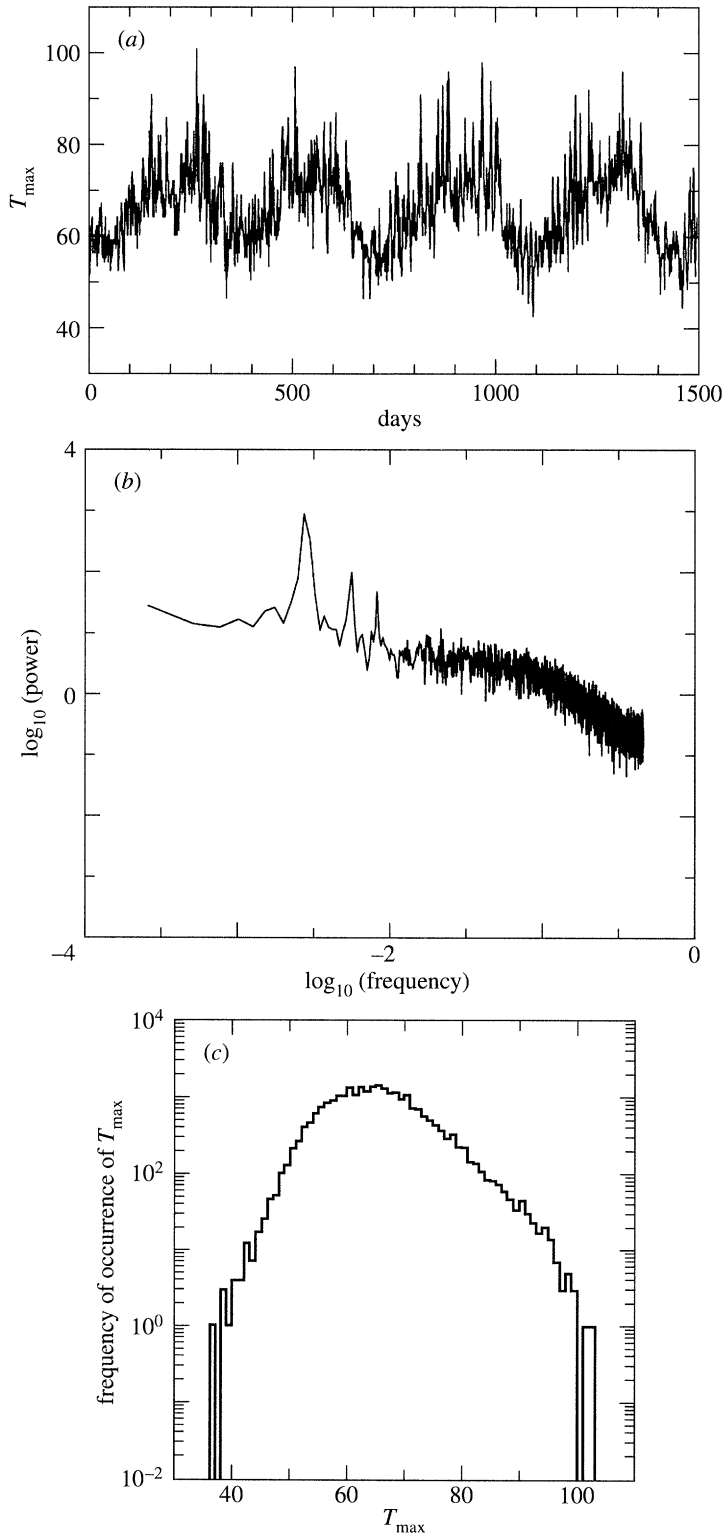

Figure 1. For description see opposite.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

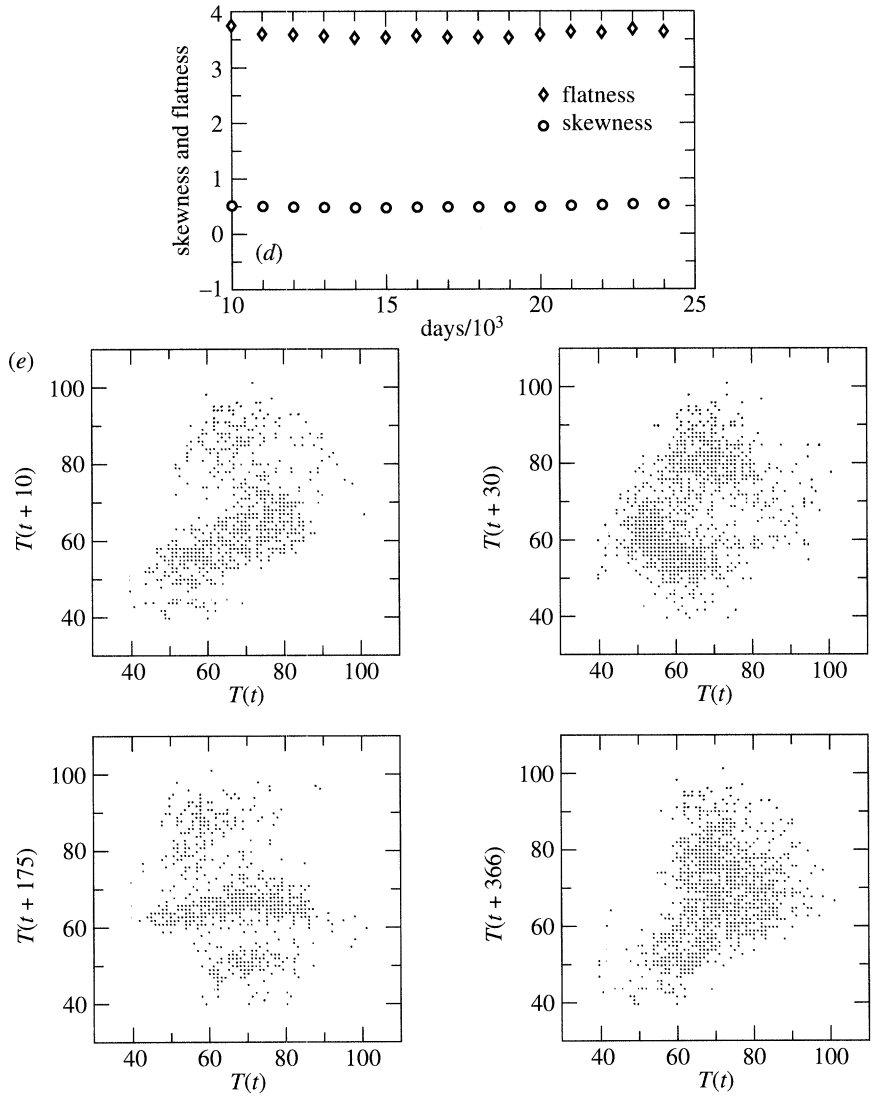

Figure 1. Plots for the $T_{\max}$ data consisting of *ca.* 25 000 points. (*a*) A sample of the time trace. (*b*) The power spectrum. The frequency is in $(\text{days})^{-1}$ and the spectrum was computed in blocks of 2^{12} points and averaged over six blocks. (*c*) The amplitude histogram with bin size = 1. (*d*) The variation of the skewness factor $\langle (T_{\max} - \overline{T}_{\max})^3 \rangle / \sigma^3_{\max}$, and the flatness factor $\langle (T_{\max} - \overline{T}_{\max})^4 \rangle / \sigma^4_{\max}$ with the data record length. $\overline{T}_{\max}$ is the mean and $\sigma^2_{\max}$ is the variance. (*e*) Delay-time plots for delay-times of 10, 30, 175 and 366 days. 10 000 points have been used for each plot.

the ratio of the time span of the data set to the total number of points in it) in order to display it in units of $(\text{days})^{-1}$. The delay-time plots undergo only the innocuous change of having a small number of points perturbed from their true values.

Figures 1 and 2 show, respectively for the $T_{\max}$ and $T_{\min}$ data, the following features: (*a*) a typical segment of the time trace, (*b*) power-spectral density, (*c*) amplitude histogram, (*d*) the skewness and flatness factors and (*e*) delay-time plot (i.e. plot of $T(t)$ versus $T(t+\tau)$, where τ is the delay-time). The notable features of both

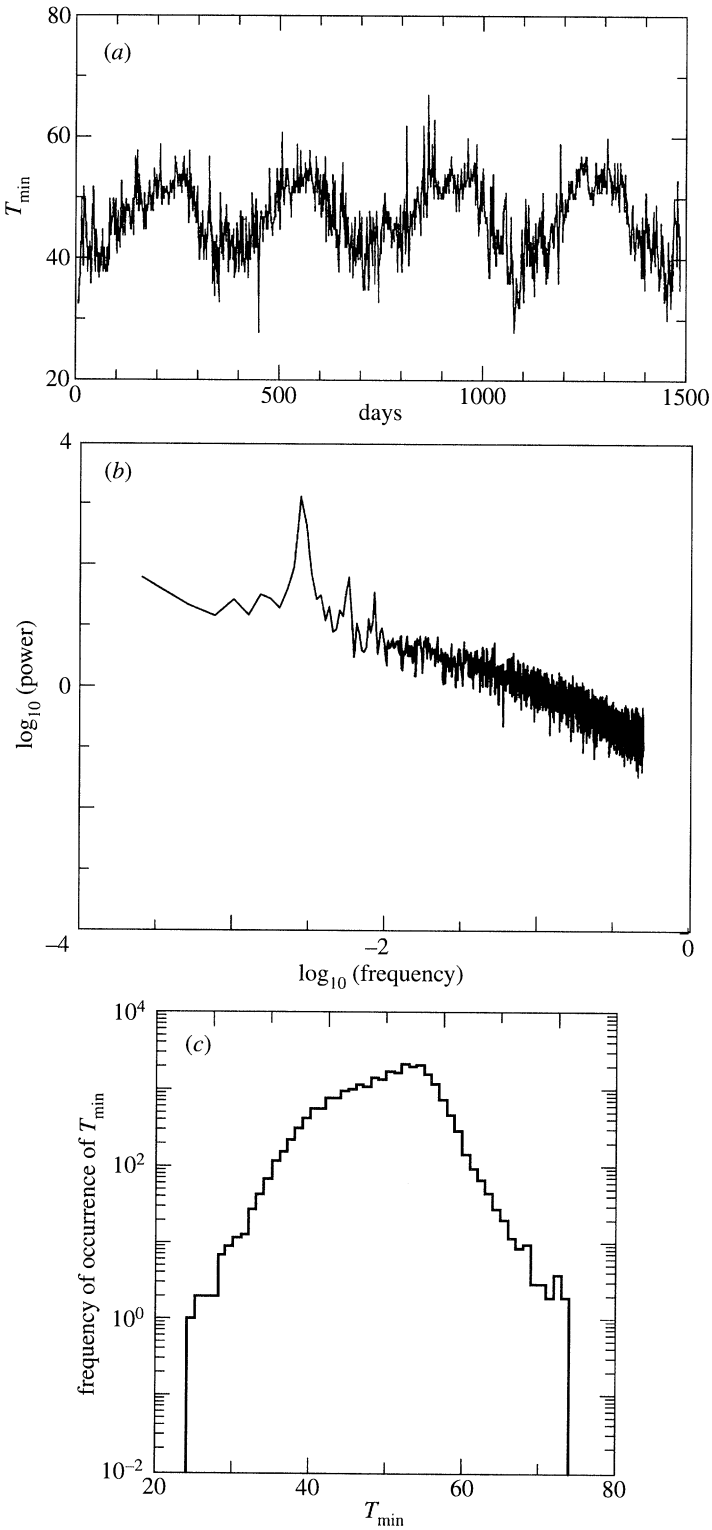

Figure 2. For description see opposite.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

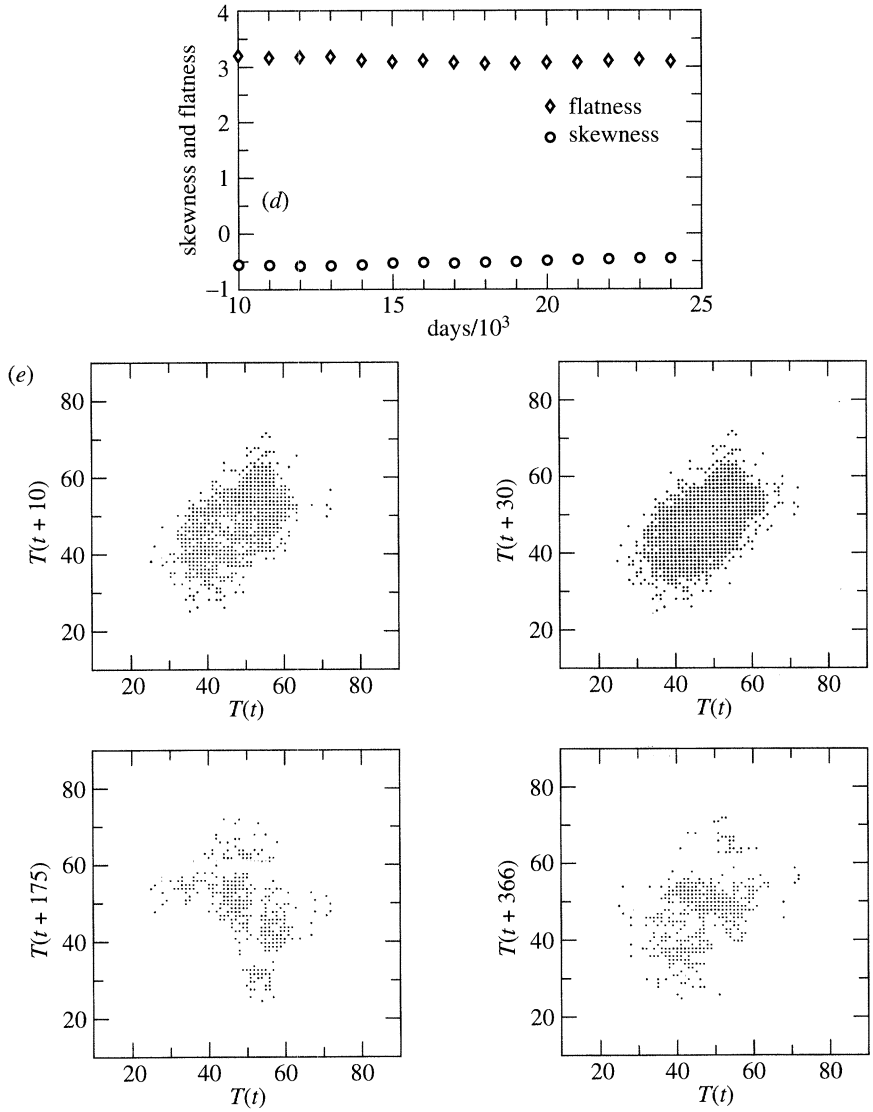

Figure 2. Plots for the $T_{\min}$ data consisting of *ca.* 25 000 points. (*a*) A sample of the time trace. (*b*) The power spectrum. The frequency is in $(\text{days})^{-1}$ and the spectrum were computed in blocks of 2^{12} points and averaged over six blocks. (*c*) The amplitude histogram with bin size = 1. (*d*) The variation of the skewness factor $\langle(T_{\min} - \overline{T}_{\min})^3\rangle/\sigma^3_{\min}$, and the flatness factor $\langle(T_{\min} - \overline{T}_{\min})^4\rangle/\sigma^4_{\min}$ with the data record length. $\overline{T}_{\min}$ is the mean and $\sigma^2_{\min}$ is the variance. (*e*) Time-delay plots for delay-times of 10, 30, 175 and 366 days. 10 000 points have been used for each plot.

time traces (figures 1*a*, *b*) are: (i) periodicity with a period approximately equal to a year and (ii) large amplitudes of small scale fluctuations. These features appear in the power spectra (figures 1*b* and 2*b*) in the form of peaks (at $f = 1/366$ $(\text{days})^{-1}$, $1/175$ $(\text{days})^{-1}$ and $1/119$ $(\text{days})^{-1}$) and the broadband structure. The fundamental frequency corresponds to the one-year periodicity in the climate associated with the rotation of the Earth. The two near-harmonics do not correspond to known cycles and their presence is quite likely indicative of the nonlinear nature of the data. The his-

tograms of the temperature (figures 1*c* and 2*c*) highlight the skewness of the data. The time series of maximum temperature has mean ($\overline{T}_{\max}$) $\approx 64\,^{\circ}$Fahrenheit, standard deviation ($\sigma_{\max}$) $\approx 8.1\,^{\circ}$Fahrenheit, skewness factor ($\langle(T_{\max}-\overline{T}_{\max})^3\rangle/\sigma^3_{\max}$) ≈ 0.5 and flatness factor ($\langle(T_{\max}-\overline{T}_{\max})^4\rangle/\sigma^4_{\max}$) ≈ 3.6. The time series of minimum temperature has mean ($\overline{T}_{\min}$) $\approx 50\,^{\circ}$, standard deviation ($\sigma_{\min}$) $\approx 5.7\,^{\circ}$, skewness factor ($\langle(T_{\min}-\overline{T}_{\min})^3\rangle/\sigma^3_{\min}$) ≈ -0.45 and flatness factor ($\langle(T_{\min}-\overline{T}_{\min})^4\rangle/\sigma^4_{\min}$) ≈ 3.1. The variation of the skewness and flatness factors with the record length (figures 1*d* and 2*d*) show that they are reasonably converged. The delay-time plots (figures 1*e* and 2*e*) show no structure even for values of delay-time corresponding to the fundamental frequency and the first near-harmonic in the Fourier spectrum.

As mentioned at the beginning of this section $T_{\max}$ and $T_{\min}$ have been recorded to the nearest integer. This is a large discretization level (1.5% of the range of the $T_{\max}$ data and 2% of the range of the $T_{\min}$ data) and we take it as our operational definition of measurement error.

3. Phase space reconstruction

In order to extract any information about the attractor from the time series of an observable, one needs to set up a map $f : A \to \mathbb{R}^n$, where A is the attractor (in phase space) and $\mathbb{R}^n$ is the n-dimensional Euclidean space, such that $P(A) = P(f(A))$, where $P(A)$ refers to some property of A and $f(A)$ is the *reconstruction* of A. Packard *et al.* (1980) advocated using independent coordinates from a time series to form the map f. They suggested using the value $x(t)$ of the time series and its time derivatives $\dot{x}(t)$, $\ddot{x}(t)$, ..., as independent coordinates. Another set of independent coordinates, $\{x(t), x(t+\tau), \ldots, x(t+(n-1)\tau)\}$ is attributed in that work to a communication from D. Ruelle. This latter procedure is described more fully by Eckmann & Ruelle (1985). The coordinates are called delay-coordinates, τ is called the delay-time and the points in the delay-reconstruction are called delay-vectors. The delay-coordinate map is easy to implement numerically and hence often used for reconstruction.

Having found the form of the map, one needs to know the values of n and τ for which a desired property P will be preserved for a typical dynamical system and a typical observation function (i.e. the observed time series). To preserve qualitatively the global orbit structure on the attractor, thereby preserving all dynamical properties (relating to the sequencing of points on the orbit) and measure-theoretic properties (relating to some probability measure on the attractor) of the system, almost all delay-coordinate maps are good enough when n is greater than twice the dimension of A (Sauer *et al.* 1991). (This result is true under certain hypotheses which are vacuously satisfied by aperiodic time series. Also, 'dimension' in this result refers specifically to the boxcounting dimension.) This result is an extension of an earlier result of Takens (1981). If, instead of preserving the global orbit structure, one is interested only in the dimension (a measure-theoretic property), then a smaller n may suffice. It has been proved that n not less than the dimension of A is sufficient to preserve the dimension when the relevant dimension is the Hausdorff dimension (Hunt *et al.* 1993) or the correlation dimension (Sauer & Yorke 1993).

Since the attractor dimension is the unknown in practice, the theorems just mentioned cannot be used to estimate the value of n required to preserve the dimension. However, we are assured that if n is *large enough* then the reconstruction will have the dimension of the attractor. Thus measurement of a dimension of a dynamical

system from a time series has to involve a series of reconstructions for increasing values of n till the 'dimension' saturates, assuming of course that this saturation is not an artifact of the finiteness of the data.

For data with finite precision, the value of the delay-time τ plays a practical role in phase space reconstruction. Small values of τ (implying that delay-coordinates are strongly correlated) lead to a phase space portrait which is clustered around the diagonal of the phase space. Large values of τ (implying that delay-coordinates are almost uncorrelated) will lead to a space-filling portrait. In both cases, the estimation of attractor properties will be difficult and might lead to erroneous results. There have been various suggestions for an optimum value of τ, ranging from the simple – for example, the time required for the autocorrelation function of the time series to reach 1/e or zero (Atten *et al.* 1984) – to the more sophisticated – for example, the first minimum of the mutual information between delay-coordinates (Fraser & Swinney 1986). Our choice of τ has been influenced significantly by the number of 'bad points' (N_{b}) in the reconstructed phase space. Let b_{l} represent the breaks of length less than or equal to τ and b_{g} represent the breaks of length greater than τ. If the spacing between consecutive breaks is larger than $(n-1)\times\tau$ (as it is for the temperature data, for all values of n and τ considered here), then

$$N_{\mathrm{b}} = (n-1)\times[\text{sum of the lengths of } b_{\mathrm{l}} + \tau \times \text{number of } b_{\mathrm{g}}]. \tag{3.1}$$

This formula shows that N_{b} increases with τ for a fixed n. Hence, based solely on the criterion of minimizing N_{b}, the choice of τ would be 1. As a compromise between 'optimal phase space reconstruction' and 'minimal number of bad points' we have chosen $\tau = 10$. For this value the autocorrelation is *ca.* 0.43 and and the number of 'bad points' in the worst case ($n = 9$) is about 4% of the total number of points on the reconstructed attractor.

In principle, all techniques based on phase space reconstruction can be carried out making use of only the 'valid' delay-vectors (those that do not go across breaks in the data). However, such sophistication comes at a price, and its value has to be assessed for each specific test separately. The initial step of any phase space technique (based on delay-embedding) is the formation of delay-vectors from a given time series. This requires that detailed information be provided, along with the time series, about the break positions and break lengths as well. Tests in which the dynamics of the attractor have to be reconstructed will suffer from the added expense of having to label all the points which do not have images (because of the breaks), so that the algorithm does not use them while computing local dynamics. These bookkeeping overheads may well overwhelm the overall computational time of tests depending, of course, on the specific test. One then has to ask the question whether the gain (which relates to the elimination of the error incurred due to the assumption of contiguity) is commensurate with the added overhead.

In order to answer this question, we shall characterize the various tests by certain qualitative attributes and we will discuss the gain and the price as functions of these attributes. The characterization shall be in terms of: (1) the number of points used – small (a few thousand), or large, (2) whether dynamic (temporal sequence of points), or geometric (location of points) information is required and (3) whether or not the tests involve some form of global averaging.

The bookkeeping overheads will be negligible for tests requiring small data sets, but significant for the others – most certainly for those requiring dynamic information such as noise reduction and the test for determinism. As regards the error incurred in

the results of these tests, there is a sharp difference between the tests which require dynamic information and those that require only geometric information. The error in case of the 'geometric' tests will primarily depend on the number of 'bad' delay-vectors and not on the magnitude of perturbation they represent. In case of 'dynamic' tests the magnitude of the perturbations might play a significant role. Finally, the errors are going to be small for the tests which involve global averaging.

Tests which require small data sets, global SVD and nonlinear SVD, could and should be carried out with only 'valid' delay-vectors. In our work, these tests have been carried out with a section of data of length 5000, having only two breaks, each of length one. Consequently, the errors in the results of these tests can be considered to be negligible. Among the tests which require large data sets, those that are 'geometric' (dimensions and local SVD) will have small errors if the number of 'bad' delay-vectors is small. The test of determinism involves a global average of a local statistic and, for this reason, a small number of 'bad' delay-vectors will introduce only a small error.

The noise reduction technique of Kostelich & Yorke has the *potential* for suffering adversely, even by the small number of 'bad' points we have estimated (4% in the worst case). At the same time, because of the large number of data points required (with the requirement becoming particularly critical for larger reconstruction dimensions) and of the need to estimate local dynamics, the bookkeeping overheads will be rather high. It is unavoidable that large atmospheric data sets (with long time spans) will have breaks in them. As longer data sets become available and the computing power increases – so that larger reconstruction dimensions can be used for phase space reconstruction – attempts will have to be made to build algorithms to identify and suppress heuristically the influence of 'bad' delay-vectors. To a large extent such heuristics are built into the noise reduction technique of Kostelich & Yorke. Details will be presented in the next section.

Finally, we need to justify the sampling of the consecutive local maxima or local minima of a single variable. This may be thought to be equivalent in some sense, to sampling a time series of a certain Poincaré section of the flow, and numerical evidence from Lorenz flow and Rössler flow suggests that this is essentially the case. A Poincaré section of an attractor preserves all the dynamical properties of the attractor and has dimension one less than the dimension of the attractor. To monitor the possible artifacts of using such a non-standard reconstruction, we have used a time series of local maxima of a single variable of the Lorenz flow as one of the control signals to be compared with the temperature data with respect to all the tests of low-dimensionality. This will be described in § 7.

4. Noise reduction

(*a*) *Methodology*

Heuristically, noise is the difference between the 'true' system and its experimental observation. To attempt a more precise definition one needs to restrict the class of systems under consideration. For deterministic systems, two concepts of noise can be made precise: measurement noise and dynamic noise. Suppose that a noise-free deterministic dynamics is given by the map

$$\boldsymbol{y}_{n+1} = g(\boldsymbol{y}_n). \tag{4.1}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The measurement noise ($\boldsymbol{r}_n$) is defined as the difference between the exact orbit ($\boldsymbol{y}_n$) and the measured orbit ($\boldsymbol{x}_n$),

$$\boldsymbol{r}_n = \boldsymbol{y}_n - \boldsymbol{x}_n. \tag{4.2}$$

In general, $\boldsymbol{r}_n$ may be deterministic or stochastic; and, if stochastic, it may be correlated or uncorrelated with $\boldsymbol{y}_n$. Dynamic noise, on the other hand, is a feedback process wherein the system is perturbed during its evolution,

$$\boldsymbol{x}_{n+1} = g(\boldsymbol{x}_n + \boldsymbol{r}_n). \tag{4.3}$$

In this case, an orbit satisfying the exact dynamics might simply not exist. The 'shadowing problem', which deals with the existence and search of 'exact' orbits, is a very difficult one. Importantly, although the two types of noises are associated with different mechanisms (and hence might require different schemes to remove them), they may not be distinguishable *a posteriori* based only on the data. Almost all noise-reducing schemes aim to remove additive measurement noise, with special operational definitions for $\boldsymbol{r}_n$.

Any practical noise-reduction scheme has to assume that the 'exact' orbit can be recovered. The scheme should have: (1) an operational definition of noise, (2) an algorithm for separating the noise (as defined in step (1)), and (3) an estimate of the effect of noise-removal on the signal. If there is *a priori* knowledge of some characteristics of the system, a natural way to define noise would be to relate it to the part of the signal devoid of those characteristics. For example, in the case of band-limited signals, noise could be defined as a collection of high frequency components. This is the operational definition of noise used in low-pass Fourier filtering. As another example, in the case of low-dimensional dynamical systems one could relate noise to high-dimensional behaviour (deterministic or stochastic). In the absence of any *a priori* knowledge of the exact system, one has to *assume* an operational definition of noise. The assumption should be guided by some testable hypothesis about the system. For the atmospheric temperature data, a plausible hypothesis (and one which we wish to test) is that the local climate attractor (associated with the temperature data) is low-dimensional. It would be reasonable to choose a noise-reduction scheme aimed at removing noise from a low-dimensional system. Several (essentially similar) schemes have been proposed to remove (measurement) noise from low-dimensional systems (for reviews, see Grassberger *et al.* 1993; Kostelich & Schreiber 1993). We use the method proposed by Kostelich & Yorke (1990).

The goal is to get an approximation for the underlying 'true' attractor from its 'blurred' reconstruction. Kostelich & Yorke propose to achieve this by using local linear approximations of the map governing the dynamics of the reconstructed phase space, g. The linear approximation (which they call the Eckmann–Ruelle linearization) can be computed by carrying out a linear least-squares fit to the data points $(\boldsymbol{x}_i^{(\mathrm{ref})}, \boldsymbol{y}_i^{(\mathrm{ref})})$ with a linear estimator, where $\boldsymbol{x}_i^{(\mathrm{ref})}$ are the delay-vectors in a small neighbourhood of $\boldsymbol{x}^{(\mathrm{ref})}$, and $\boldsymbol{y}_i^{(\mathrm{ref})}$ are their respective images under g. In the next step, Kostelich & Yorke perturb sections of the orbit (defined as sets of delay-vectors ordered temporally, $\{\boldsymbol{x}_i\}_1^p$) to a new one $\{\hat{\boldsymbol{x}}_i\}_1^p$ so that it is more 'consistent' with the dynamics of the locally linear maps. The measure of inconsistency (which is to be minimized) is taken to be

$$\sum_{i=1}^{p} \{w\|\hat{\boldsymbol{x}}_i - \boldsymbol{x}_i\|^2 + \|\hat{\boldsymbol{x}}_i - L_{i-1}(\hat{\boldsymbol{x}}_{i-1})\|^2 + \|\hat{\boldsymbol{x}}_{i+1} - L_i(\hat{\boldsymbol{x}}_i)\|^2\} \tag{4.4}$$

where w is a weighting factor (which depends upon the confidence one has in the accuracy of the original values), L is the linear map, p is the length of the trajectory section adjusted in a single adjustment step and $\|\cdot\|$ denotes the Euclidean norm. The minimization problem is solved using least-squares.

The size of the neighbourhoods in which the local linear maps are computed have to be small so that nonlinearities do not introduce large errors. At the same time the neighbourhoods have to be large enough to provide a sufficient number points to compute the linear maps (involving a least-squares fit). Reasonable values of the size of the neighbourhoods are between 10% and 20% of the attractor extent. Points whose images do not remain close to the image of the reference point are discarded as they probably lie on another fold of the attractor. Such restrictions on the size of the neighbourhoods of the reference point and its image will prevent 'bad' points which are large perturbations of the trajectory from being used in linear map computations. The influence of the 'bad' points is further suppressed by forms of local averaging during linear map computations and trajectory adjustment. The least-squares fit of the local maps is overdetermined, i.e. the number of points used in the computations (50–100 in our work) are more than the n^2 unknowns (n being the reconstruction dimension). This reduces the influence of the few neighbouring 'bad' points in determining the local linear maps. In the trajectory adjustment criterion of equation (4.4), a large value of p will reduce the effect of 'bad' points in determining adjustment distances.

An upper limit is placed on the distance that a point can move during trajectory adjustment (5–10% of the attractor extent), since large adjustments would indicate that the local linear approximations do not work well, either because the data are too noisy and/or the reconstruction dimension is not large enough to preserve dynamical properties of the attractor. If a large number of points (say more than 5–10%) need to be moved beyond the specified limit then the scheme should be abandoned.

(*b*) *Results*

The noise reduction algorithm discussed above was applied to the $T_{\max}$ and $T_{\min}$ data with different sets of reconstruction dimension n (3 to 7) and delay-time τ (1 to 30). The maximum adjustable distance was set to 10% of the attractor extent. The value of the weighting factor w in equation (4.4) was taken to be 1.0 and 0.8, and the value of p was taken to be 24. Linear maps were computed using 50–100 points inside neighbourhoods of radii equal 20% of the attractor extent. It was also required that the maximum distance between the images of the selected points and the image of the reference point be less than 20% of the attractor extent. In all the cases, it was found that a large number of points needed to be moved by more than the preset maximum adjustable distance. In case of the higher dimensional reconstructions ($n \geqslant 5$) there was the added problem of the failure of linear map computations at several points because of a lack of sufficient neighbouring points. Noise reduction was also performed on a smaller section of the maximum temperature time series for reconstruction dimensions of three and four, with various values of the delay-time. The 17 000-point long section contains only three large breaks, of lengths 11, 43 and 23 respectively. The results of the noise reduction are qualitatively similar to that of the whole time series. *Thus the preprocessing is ineffective for these data, and the rest of the analysis will be done using the original data.* The failure of the noise-reduction scheme may already be an indication that a low-dimensional characterization of the data might not be possible.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

5. Dimension calculation

The first step in attempting to show that a dynamical system is low-dimensional is obviously to measure its dimension. In this section we describe the various existing definitions of 'dimension' and point out the practical schemes to compute the 'dimension' from an experimental time series. At the end of the section the dimension estimates of the temperature data are given.

(*a*) *A brief review of dimensions*

It is best to think about the dimension in terms of the intuitive geometric notion of topological dimension. Although the definition of the topological dimension and the tools used to calculate it are far from trivial, we do have a intuitive feel of the (topological) dimension of manifolds (sets which locally look like Euclidean spaces). Thus, a curve is one dimensional, a surface is two dimensional, and so on. Many simple attractors are manifolds and fit into our intuitive scheme of things very well. For example, a stationary time-independent equilibrium (fixed point) has dimension zero, a stable periodic oscillation (limit cycle) has dimension one, and a doubly periodic attractor (2-torus) has dimension two. But chaotic (strange) attractors are often not manifolds (such sets are usually called fractals (Mandelbrot 1983)) and cannot be characterized with sufficient accuracy using the notion of topological dimension. For example, both a straight line (manifold) and a space filling curve (fractal) have the same topological dimension. Clearly, such a characterization is too coarse to be useful for fractals. Throughout this century people have attempted to develop useful notions of dimensions of a fractal set. Most notions start with a covering of the fractal set with nice Euclidean sets like cubes or spheres and investigate the scaling properties of some measure on the cover of the fractal set by increasing the observational resolution (i.e. by decreasing the size of the covering cubes or spheres). Broadly speaking, the relevant definitions of fractal dimension fall into two groups. In the first group, called the geometric dimensions, the scaling variable is the cardinality of the cover. This is a purely geometric prescription and treats the 'fractal' as a set of points in phase space. It does not necessarily relate to the natural measure supported on the fractal set (i.e. the relative frequency of occurrence of any point in the set). In the second group, called the generalized dimensions, the scaling of powers of the natural measure supported on the fractal set is considered. The correlation dimension (D_2) and the information dimension (D_1), for example, belong to this category.

For purposes of this study we need to keep in mind, first, that all of these dimensions agree with our intuitive notion of dimension for manifolds but yield non-integer values for fractal sets. Secondly, numerical computation of dimensions of attractors has more or less been limited to variants of two algorithms: the Grassberger–Procaccia algorithm (Grassberger & Procaccia 1983), which estimates D_2, and the nearest neighbour algorithm (Badii & Politi 1985), which estimates D_1. The reason is that most definitions of dimensions (especially the geometric dimensions) do not lend themselves to efficient algorithms except when the 'dimension' of the attractor is very small. It can be shown that $D_2 \leqslant D_1$ (Hentschel & Procaccia 1983), but the numerical values of the two dimensions are usually within the measurement precision for attractors reconstructed from experimental data. This will be especially true in the present case because of the poor precision of the data.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

(b) Correlation dimension (D_2)

The estimate of D_2 according to the Grassberger–Procaccia algorithm is given by the slope of the plot $\log C(r)$ versus $\log r$ for small r, where the correlation sum $C(r)$ is given by the number of pairs of points on the reconstructed attractor with mutual distance not more than r. There are several practical considerations in the estimation of D_2, which have been highlighted over the years by many researchers based on heuristic arguments and empirical evidence. We list all the considerations we have taken into account. (1) $C(r)$ is computed by averaging the two-point distance statistics over 1250 reference points (which is 5% of the total number of points on the attractor), selected randomly with respect to the natural measure of the attractor. (2) Since the scaling is expected to hold good only for small values of r, we have limited the upper end of the scaling range to a fraction of the attractor extent ($r_{\max}$). Based on the full-range $C(r)$ curve at three randomly selected points we take $r_{\max}$ to be 25% of the attractor extent. (3) Discretization of the data produces steps in the $C(r)$ curve with ordinate levels at $C(r + \epsilon/2)$, where ϵ is the level of discretization. The correct scaling can be obtained if $C(r)$ is calculated only at the mid point of the steps. The net effect of discretization is to reduce the effective number of points on the $\log C(r)$ versus $\log r$ plot. (4) To avoid spurious dimension estimates due to high correlation between points on the attractor, the pairs of points considered for the two-point distance statistics are separated by more than 85 units of the sampling time of the time-series (this being equal to the first zero-crossing of the autocorrelations for the $T_{\max}$ and $T_{\min}$ data).

(c) Information dimension (D_1)

The estimate of D_1 according to the nearest neighbour algorithm is given by the negative of the inverse slope in the plot of $\log\langle\delta_p(k)\rangle$ versus $\log k$. Here, $\delta_p(k)$ is the distance from a reference point of its pth nearest neighbour among a collection of k points in the reconstructed attractor, and $\langle\cdot\rangle$ represents the average over all the reference points (1250 points). By choosing a large enough value of p, i.e. a nearest neighbour which lies outside the range dominated by noise or discretization, one can obtain better scaling than for D_2. A comparative study Kostelich & Swinney (1987) suggests that the nearest neighbour algorithm has better convergence than the correlation dimension algorithm (at least for dimensions less than 10). But the nearest neighbour algorithm also has difficulties. In particular, two features need to be considered carefully: (1) the average distance of the reference nearest neighbour should not be larger than 5–10% of the attractor extent, and (2) the estimation of D_1 becomes very sensitive to errors in the calculation of the slope of the plot $\log\langle\delta_p(k)\rangle$ versus $\log k$, because of the inverse relation between D_1 and the slope.

(d) Results of the dimension calculations

Figures 3a, b are plots of $\log C(r)$ versus $\log r$ for the $T_{\max}$ and $T_{\min}$ data respectively with reconstruction dimensions (n) 3–9 and delay-time (τ) equal to 10. The high level of discretization (1.5% of the attractor extent for $T_{\max}$ data and 2% of the attractor extent for $T_{\min}$ data) reduces the effective number of points to 15 and 12 respectively (if we restrict the upper end of the scaling range to 25% of the attractor extent). No scaling is apparent for either data set.

Figures 4a, b are plots of $\log\langle\delta_p(k)\rangle$ versus $\log k$ for the $T_{\max}$ and $T_{\min}$ data respectively with $n = 3$–9 and $\tau = 10$. For the $T_{\max}$ data the scalings have been shown for

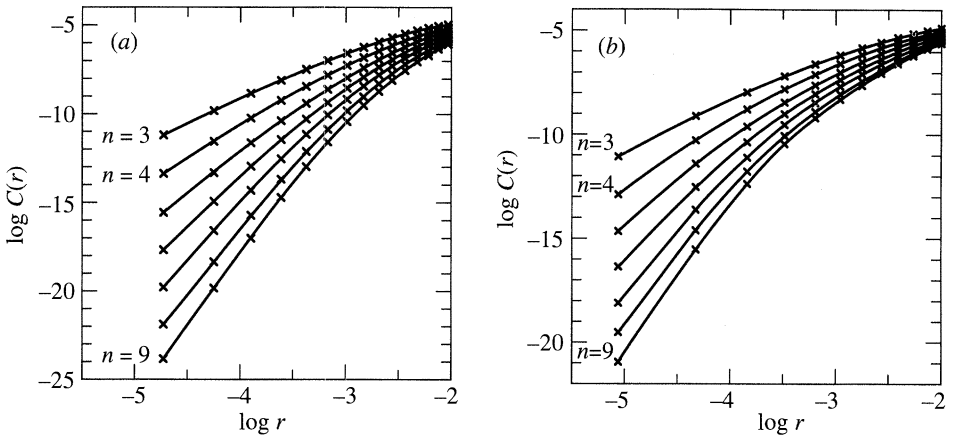

Figure 3. $\log C(r)$ versus $\log(r)$. (*a*) $T_{\max}$ data, reconstruction dimension $n = 3$–9, and delay-time $\tau = 10$. (*b*) $T_{\min}$ data, $n = 3$–9, and $\tau = 10$. $C(r)$ has been calculated using 1250 reference points. $r = r_{\max} - (k - \frac{1}{2}) \times \epsilon$, where $r_{\max} = 25\%$ of the attractor extent, $\epsilon(= 1)$ is the discretization level and k is increased from 1 until r attains the smallest value greater than 2ϵ ($k = 15$ for $T_{\max}$ and $k = 12$ for $T_{\min}$).

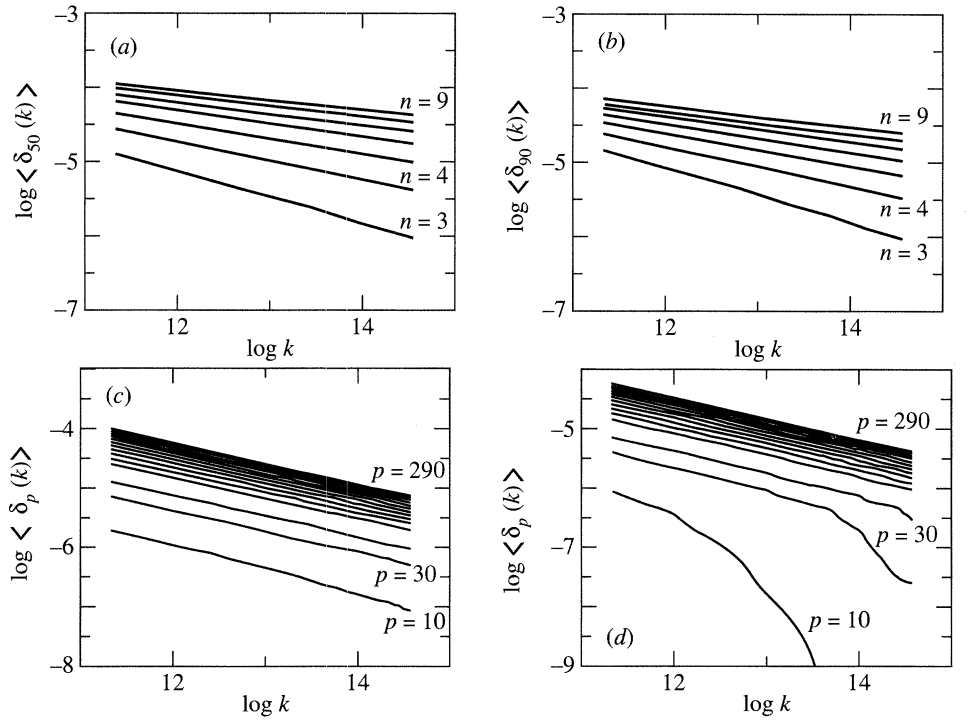

Figure 4. $\log\langle\delta_p(k)\rangle$ versus $\log k$. (*a*) $T_{\max}$ data, reference nearest neighbour $p = 50$, reconstruction dimension $n = 3$–9, and delay-time $\tau = 10$. (*b*) $T_{\min}$ data, $p = 90$, $n = 3$–9, and $\tau = 10$. (*c*) $T_{\max}$ data, $p = 10$–290 in steps of 20, $n = 3$, and $\tau = 10$. (*d*) $T_{\min}$ data, $p = 10$–290 in steps of 20, $n = 3$, and $\tau = 10$. The average $\langle\cdot\rangle$ was taken over 1250 reference points. There are 32 points on each line.

the 50th nearest neighbour (i.e. $p = 50$) and for the $T_{\min}$ data the scalings have been shown for the 90th nearest neighbour. The scalings were obtained for 15 different

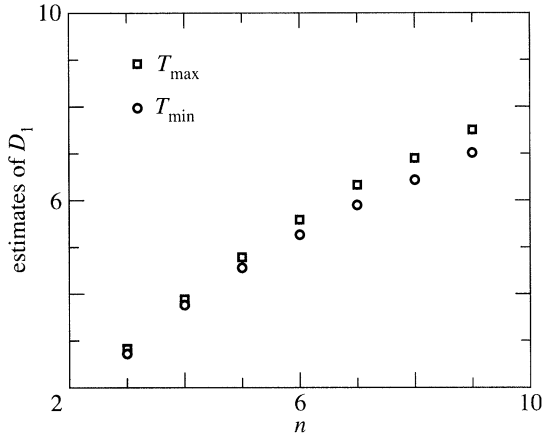

Figure 5. Information dimension D_1 versus the reconstruction dimension n for the $T_{\max}$ and for $T_{\min}$ data. D_1 is the negative of the inverse slope of the plots in figures 4*a*, *b*. The slopes have been calculated by linear regression.

nearest neighbours (i.e. $p = 10$–290 in steps of 20). Most of them give a value of D_1 close to that calculated from figures 4*a*, *b*. This can be seen from figures 4*c*, *d*, which show the scalings for the 15 different nearest neighbours for $n = 3$ and $\tau = 10$. The average distance of the reference nearest neighbours is always less than 7% of the attractor extent and hence the chances of spurious scaling, arising because of the nearest neighbours being too far away, are small. The estimated values of D_1 are shown in figure 5. The dimension estimates do not level off with n, indicating that a low-dimensional attractor does not exist.

In summary, no linear scaling region could be identified for the correlation sum ($C(r)$) and hence D_2 could not be estimated. Estimates of D_1 were obtained on the basis of the scaling of the nearest neighbours; they do not show any tendency to level off. In an attempt to corroborate this finding we examine *well-converged* characterizations of the attractor which could be related to the dimension of the attractor.

6. Potential indicators of low-dimensionality

Our objective is to find characterizations of the attractor which can be obtained reliably from a small data set (small *vis-à-vis* the requirements for the computation of D_2) and which could be used as *potential* indicators of low-dimensionality.

A natural candidate for such a characterization would be a coarse descriptor of the geometry of the attractor. The singular value decomposition (SVD) technique of linear algebra provides a basis for a minimal (linear) subspace containing the reconstructed attractor. If this subspace is a proper subspace of $\mathbb{R}^n$ then its dimension will be an upper bound to the attractor dimension. As we shall see, global SVD is too coarse an indicator of low-dimensionality for chaotic maps such as the Hénon map. The problem has to do with nonlinear correlations among the amplitudes of the basis vectors. We show that local SVD and a procedure for computing correlations among the amplitudes of the basis vectors can be used as fine tuners of the global SVD test. A less obvious *potential* indicator of low-dimensionality is a statistic to test for determinism in a time series. The idea here is that if a time series appears

deterministic in a low-dimensional reconstruction then it is likely to have a low-dimensional characterization.

(*a*) *Singular value decomposition (SVD)*

(i) *Global SVD*

Broomhead & King (1986) introduced singular value decomposition in the context of delay-coordinate reconstructions. The methodology consists of the following essentials.

First the trajectory matrix $\boldsymbol{X}$ is defined as

$$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^{\mathrm{T}}, \tag{6.1}$$

where $\boldsymbol{x}_i$ is a delay-vector, N is the total number of delay-vectors and the superscript 'T' represents the transpose operation. Each delay-vector (a row of $\boldsymbol{X}$) is a point on the reconstructed attractor in $\mathbb{R}^n$, n being the reconstruction dimension. It will be helpful to think of $\boldsymbol{X}$ as a linear map from $\mathbb{R}^N$ into $\mathbb{R}^n$, and observing that the row space of $\boldsymbol{X}$ (subspace of $\mathbb{R}^n$ containing all linear combinations of the rows of $\boldsymbol{X}$) is the smallest subspace containing the attractor. The singular value decomposition of the $N \times n$ matrix $\boldsymbol{X}$,

$$\boldsymbol{X}_{N\times n} = \boldsymbol{V}_{N\times n}\boldsymbol{\Sigma}_{n\times n}\boldsymbol{U}^{\mathrm{T}}_{n\times n}, \tag{6.2}$$

constructs two bases: the columns of $\boldsymbol{U}$ (called the singular vectors) spanning $\mathbb{R}^n$ and the columns of $\boldsymbol{V}$ (called the left singular vectors) spanning $\mathbb{R}^N$. $\boldsymbol{\Sigma}$ is a diagonal matrix and its entries σ_i (called singular values) are non-negative and given by convention in decreasing order. The σ_i are the root-mean-square of the projections of the delay-vectors onto the corresponding singular vectors and in this sense are their *amplitudes*. Since the singular vectors corresponding to the non-zero σ_i (m in number) span the row space of $\boldsymbol{X}$, the dimension of the smallest subspace containing the attractor is m.

Our test will therefore comprise of the following steps: (1) construct trajectory matrix $\boldsymbol{X}$ from the time series data, (2) compute SVD of $\boldsymbol{X}$, and (3) if $m < n$, conclude that the dimension of the attractor not more than m. If m fails to be less than n then we cannot draw any conclusions. An alternative way for obtaining m is the calculation of the spectrum of the covariance matrix, $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ (dimension $n \times n$). This procedure is variously known as Karhunen–Loêve decomposition, empirical orthogonal decomposition, proper orthogonal decomposition and principal component analysis (Lumley 1970; Devijver & Kittler 1982; Sirovich 1989). The singular values of $\boldsymbol{X}$ are the square roots of the eigenvalues of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ and the respective singular vectors and eigenvectors coincide. Working with $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is computationally advantageous since in practical applications $N \gg n$. But this comes at a certain price, the most important one being the loss of precision in determining the singular vectors. In the procedure for uncovering (nonlinear) correlations among the amplitudes of the singular vectors (to be discussed in § 6 *a* (iii), the projection of the delay-vectors onto the singular vectors will play an important role. Now, it turns out that reasonably converged singular values can be obtained with small sections of the time series (*ca.* 5000). Hence, the computational advantage gained by working with $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$, meagre as it is in this case, is worth sacrificing for the precision.

Implementation of the SVD test hinges on the possibility of recognizing the number of non-zero singular vectors unambiguously. In the presence of noise, all the singular values are non-zero and for a general case there is no objective criterion for deciding

the number of non-zero singular values associated with the 'true' system. Where there is an abrupt decrease in the magnitudes of the singular values (i.e. there exists some $i < n$ such that $\sigma_1, \ldots, \sigma_i \gg \sigma_{i+1}, \ldots, \sigma_n$) it *might be* reasonable to take the dominant ones as the non-zero singular values. Otherwise we should refrain from drawing definite conclusions.

Global SVD test is a poor indicator of low-dimensionality as it fails to identify some well-known low-dimensional chaotic systems, like the Hénon map. This is illustrated and discussed in § 7 *a*. The failure of the SVD test in such a case highlights a generic problem of representing a time series from a nonlinear system in terms of global modes (modes being basis vectors or basis functions). First, consider the representation of a time series from a linear system. Trivially, the amplitudes of the excited modes (those with non-zero amplitudes) will be uncorrelated and the number of excited modes will be equal to the dimension of the system. Nonlinearities lead to the excitation of new modes whose amplitudes are (nonlinearly) correlated to the amplitudes of the modes that produced them. Consequently, in case of a nonlinear dynamical system (with a finite-dimensional attractor) the number of excited global modes will always overestimate the dimension.

(ii) *Local SVD*

Instead of applying SVD to the whole reconstructed attractor, it is applied to points in a neighbourhood of a reference point. When the attractor happens to be a manifold, scaling of singular values with the size of the neighbourhood can be used to calculate the topological dimension (Broomhead & King 1986). For fractal sets, a criterion has been put forward in Broomhead *et al.* (1991) to recognize local directions in which the natural measure of the attractor is a fractal, and prescriptions have been provided to interpret the scaling of singular values. However, a problem with using a local scaling method is the need to access local neighbourhoods for a large range of neighbourhood sizes in order to get a substantial scaling range. Given the finite amount of data and the fact that reconstructions can produce quite convoluted copies of the attractor as the reconstruction space grows bigger, it is not always possible to get a substantial scaling region. With these considerations in mind, we have decided not to attempt to make use of the full potential of the local SVD. Similar to the global SVD test, the number of non-zero (local) singular values, m_l, will be identified. If m_l turns out to be less than n then m_l would be an upper bound for the (pointwise) dimension at the reference point. The reason for expecting $m_l < m$ is that locally the attractor will be approximated better by linear spaces, i.e. represented by fewer linear modes.

(iii) *Correlations among singular vectors*

Local SVD is one way of getting around the limitations of a global basis in representing a nonlinear time series. Another way, suggested by Healey (1994), is to use radial basis functions (Powell 1985) to uncover the nonlinear correlations among the amplitudes of the excited singular vectors. To test for correlations, sets of 'independent' and 'dependent' amplitudes have to be chosen. A systematic way of chosing such sets would be to use the natural ordering of the singular vectors (given by the decreasing order of the singular values) and to evaluate the dependence of the amplitude of the $(k+1)$th singular vector on the amplitudes of the first k singular vectors.

A standard procedure for determining correlation between two sets of data points

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

is to use a linear least-squares fit with appropriate basis functions; the deviation of the fit from the data will be a quantitative measure of the correlation between the variables representing the two data sets (the deviation being smaller for good correlations). Obviously, the deviation will depend on the particular basis functions and it is to be understood that the choice of these is made so as to minimize the deviation. A practical implementation of such a procedure will require basis functions that approximate a large class of correlations, splines in one dimension being an example. Radial basis functions are generalizations of one-dimensional splines to higher dimensions. Examples are $\phi(r) = r$, r^3 and $r^2 \log r$, where r refers to the Euclidean distance.

In our problem the data points to be fitted are $(c_{nk}, c'_{n(k+1)})$ where

$$c_{nk} = (p_{n1}, p_{n2}, \ldots, p_{nk})^{\mathrm{T}}$$

is the vector of the projections of the nth delay-vector on the first k singular vectors and $c'_{n(k+1)} = p_{n(k+1)}$ is the projection of the nth delay vector on the $(k+1)$th singular vector. Formally, the problem can be expressed in matrix notation as

$$\boldsymbol{C}'_{N\times 1} = \boldsymbol{\Phi}_{N\times p}\boldsymbol{\Omega}_{p\times 1}, \tag{6.3}$$

where $\boldsymbol{C}'_j = c'_{j(k+1)}$, $\Phi_{nl} = \phi(\|c_{nk} - c_{lk}\|)$ (where the lth delay-vectors are randomly chosen) and $\boldsymbol{\Omega}$ is the matrix of fitted parameters. A little matrix algebra will show that the linear least-squares fit can be expressed as

$$\boldsymbol{C}'_{\mathrm{fit}} = \boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}', \tag{6.4}$$

where $\boldsymbol{S}$ is the matrix of left singular vectors of $\boldsymbol{\Phi}$. The standard deviation between $\boldsymbol{C}'$ and $\boldsymbol{C}'_{\mathrm{fit}}$, $\sigma'_{k+1} = \|\boldsymbol{C}' - \boldsymbol{C}'_{\mathrm{fit}}\|$, provides a quantitative estimate of the nonlinear dependence of the amplitude of the $(k+1)$th singular vector on those of the k most-energetic singular vectors. The efficacy of this method depends on the range of functional forms that can be estimated by the radial basis functions which, though large, does not cover all possible functional forms. Hence, conclusions can be drawn only when nonlinear correlations are uncovered.

(*b*) *Determinism in a time series*

The test for recognizing determinism in a time series (Wayland *et al.* 1993) is based on the idea of 'phase space continuity' for deterministic systems, which is that nearby points in phase space will remain nearby for short times. This test develops a statistic to measure the observed continuity in the reconstructed attractor and can be taken to be an indicator of determinism. If a time series appears deterministic in a low-dimensional reconstruction, then it is most likely that the time series will have a low-dimensional characterization. Both high-dimensional and stochastic systems will appear stochastic in a low-dimensional reconstruction.

Let x_0 be an arbitrary reference point on the reconstructed attractor and $x_1, \ldots, x_l$ be the l nearest neighbours of x_0. Let $y_0, \ldots, y_l$ represent their images. The normalized standard deviation of the translation vectors $(v_j = y_j - x_j)$,

$$e_{\mathrm{trans}} = \frac{1}{l+1}\sum_{j=0}^{l} \|v_j - \langle v\rangle\|^2 / \|\langle v\rangle\|^2, \tag{6.5}$$

where $\langle\cdot\rangle$ represents the average over $l+1$ points and $\|\cdot\|$ represents the Euclidean

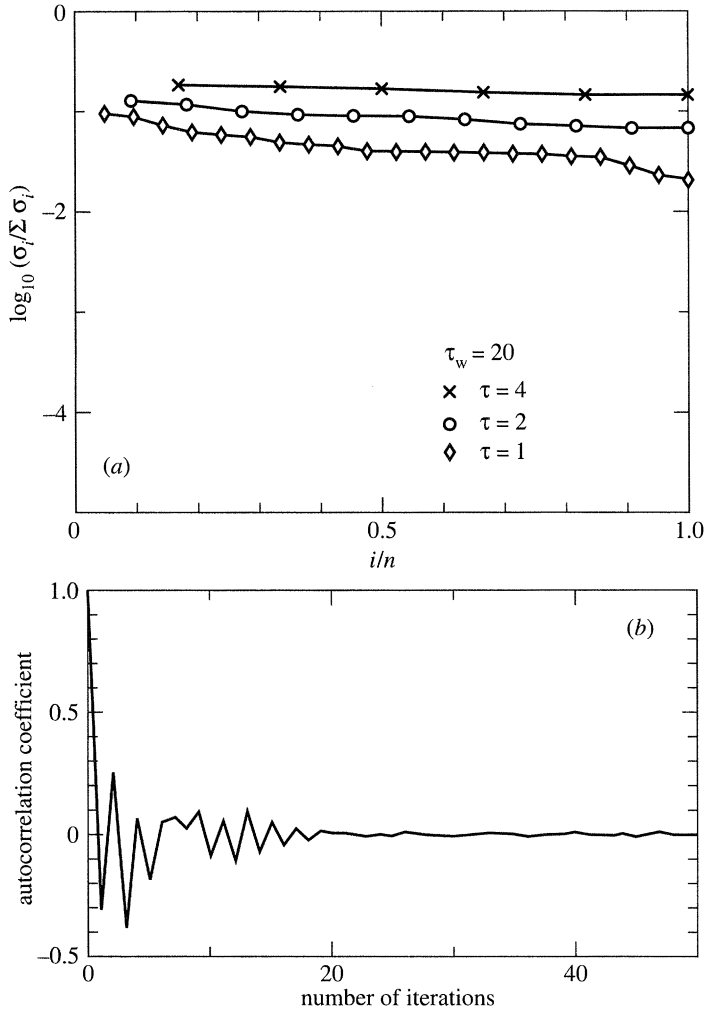

Figure 6. Plots for a time series of the x-variable of the Hénon map. (a) The normalized global singular values $\log_{10}(\sigma_i/\sum\sigma_i)$, versus the scaled reconstruction dimension i/n, with $\tau_w = 20$ and τ taking on values 4, 2 and 1. The time series was scaled to have zero mean and 5000 data points were used in the computations. (b) The autocorrelation function $(\sum_{i=1}^{n} x(i)x(i+k)/\sum_{i=1}^{n} x(i)x(i))$ with $n = 20\,000$ and $k = 1, 2, \ldots, 50$.

norm, is a measure of 'phase space continuity' around x_0. If the time series is deterministic then the v_j will be nearly equal and the translation error e_{trans} will be small. For a stochastic series, the v_j will be oriented randomly and e_{trans} will be large. A global measure of continuity will be the median value of a set of e_{trans} evaluated at points distributed randomly on the attractor. The average of the medians obtained from several independent reconstructions of the attractor, $\langle \text{median}(e_{\text{trans}})\rangle$, provides a robust statistic of the continuity of orbits on the reconstructed attractor.

7. Results from low-dimensionality tests

(a) *Quality of the indicators*

The quality of the *potential* indicators of low-dimensionality is evaluated on the basis of their performance on a set of low-dimensional data. The control data, as

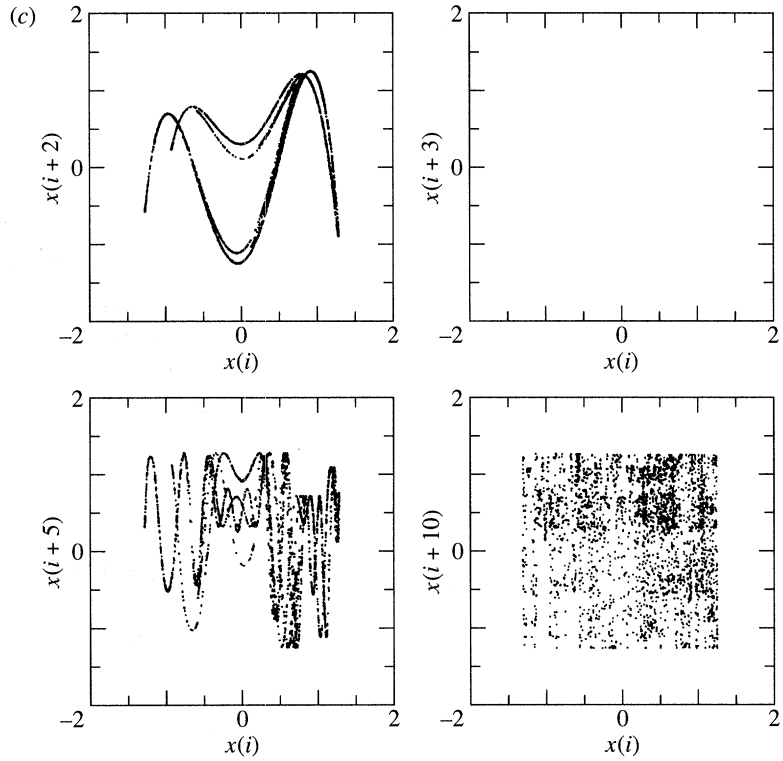

Figure 6. *Cont.* (*c*) Delay-time plots for delay-times of 2, 3, 5 and 10. 5000 points have been used for each plot.

they will be referred to subsequently, are the x-variable of the (chaotic) Hénon map and the local maxima of the y-variable of the (chaotic) Lorenz flow. All the results of these control data will be compared with those of their respective 'stochastic surrogates', formed by randomizing the Fourier phases. This comparison will give us an indication of the ability of the tests to distinguish between two sets of 'similar' data, but having very different dimensions.

The parameter values used for the (chaotic) Hénon map,

$$x_{i+1} = 1 - ax_i^2 + y_i, \qquad y_{i+1} = bx_i, \tag{7.1}$$

are $a = 1.4$ and $b = 0.3$. The first 1000 iterates of the x-variable, starting from the initial conditions $x = 0.2$ and $y = 0.2$, were discarded. The parameter values used for the (chaotic) Lorenz flow,

$$\dot{x} = \sigma(y - x), \qquad \dot{y} = rx - y - xz, \qquad \dot{z} = -bz + xy, \tag{7.2}$$

are $\sigma = 10$, $b = 8/3$ and $r = 28$. This set of differential equations was solved with a simple Euler integration scheme with a time step of 0.01 and initial conditions $x = 1$, $y = 1$ and $z = 1$. All the local maxima of the y-variable time series were sampled after discarding the first 1000 values. A set of dimension calculations (D_1) gave estimates of about 1.2 and 1 for the control data from (7.1) and (7.2) respectively. These estimates are consistent with previously known results. In comparison, one might expect that the 'stochastic surrogates' will be space filling and assume the dimension of the reconstruction space.

Figure 6*a* shows the result of the global SVD test on the control data generated from (7.1). Following Broomhead & King (1986) we plot $\log_{10}(\sigma_i/\sum\sigma_i)$, where σ_i are the singular values, against the scaled reconstruction dimension (i/n). The window length τ_w $((n-1)\times\tau)$ is fixed at 20 and τ takes on the values 4, 2 and 1. All the singular values are the same and the global SVD fails to identify the low-dimensional attractor. The reason for the failure is the complication of the geometry of the reconstructed attractor caused by nonlinear correlations among the amplitudes of the singular vectors (see discussion on global SVD). Because of the chaotic nature of the map the output decorrelates in a few iterates of the map (figure 6*b*) leading to complicated geometry in the reconstructed phase space (figure 6*c*). Hence the delay-coordinates will be decorrelated and the resulting reconstruction will appear to be space filling. Globally, the reconstruction will be similar to that for an uncorrelated stochastic process and the singular value spectrum will be flat.

Now, it is known that global SVD cannot distinguish among different dynamic processes with the same power spectrum. This is because the singular values are eigenvalues of the covariance (see § 6 *a* (i)), which depends totally on the autocorrelation of the data (provided the data record length is long enough to assure approximate stationarity). Hence the Lorenz flow data, which has a nearly 'white spectrum' like the Hénon map data, and their respective 'stochastic surrogates' have the same global singular value spectrum. In view of the present failure of global SVD, it should be pointed out that there are indeed cases where global SVD gives useful answers. Broomhead & King (1986) obtained an upper bound of 4 to the phase space containing the reconstructed attractor for the time series of the x-variable of the Lorenz flow. This signal, unlike our control data, is band-limited, and it is therefore possible to choose combinations of τ_w and τ such that the delay coordinates are not decorrelated.

Figure 7*a* shows the normalized local singular values at three different points on the reconstructed attractor, picked randomly according to the natural measure of the reconstructed attractor, for the x-variable of the (chaotic) Hénon map and its 'stochastic surrogate'. Unlike the global singular values, the local singular values for the Hénon map are not of the same magnitude. While this observation leads to no definite conclusion (since none of the singular values are actually zero), the decay of the local singular value spectra is an indication of low-dimensionality. By comparison, the 'stochastic surrogate' has flat local singular value spectra and is unlikely to have a low-dimensional characterization. The local maxima of the y-variable of the (chaotic) Lorenz flow and its 'stochastic surrogate' show similar results (figure 7*b*).

For the nonlinear correlation test, following Healey (1994), we took $\phi(r) = r^3$ as the radial basis function and p (number of random centres) to be 100. The test uncovers some of the nonlinear dependence of the amplitudes of the higher order singular vectors, for both control data (figures 8*a*, *b*). On the other hand, the 'stochastic surrogates' show no such dependence (figures 8*c*, *d*).

Figure 9*a* is a plot of $\langle\text{median}(e_{\text{trans}})\rangle$ (as defined in § 6 *b*) versus the reconstruction dimension, for the Hénon map and its 'stochastic surrogate'. The plot shows that $\langle\text{median}(e_{\text{trans}})\rangle$ of the control data is much smaller than for their 'stochastic surrogates', indicating that the former is likely to be a sample from a deterministic flow/map. The increase in $\langle\text{median}(e_{\text{trans}})\rangle$ with the reconstruction dimension, is related to the increase in 'randomness', brought about by fast decorrelation of the time series (see discussion on global SVD of the Hénon map). The control data from the Lorenz map shows similar behaviour (figure 9*b*).

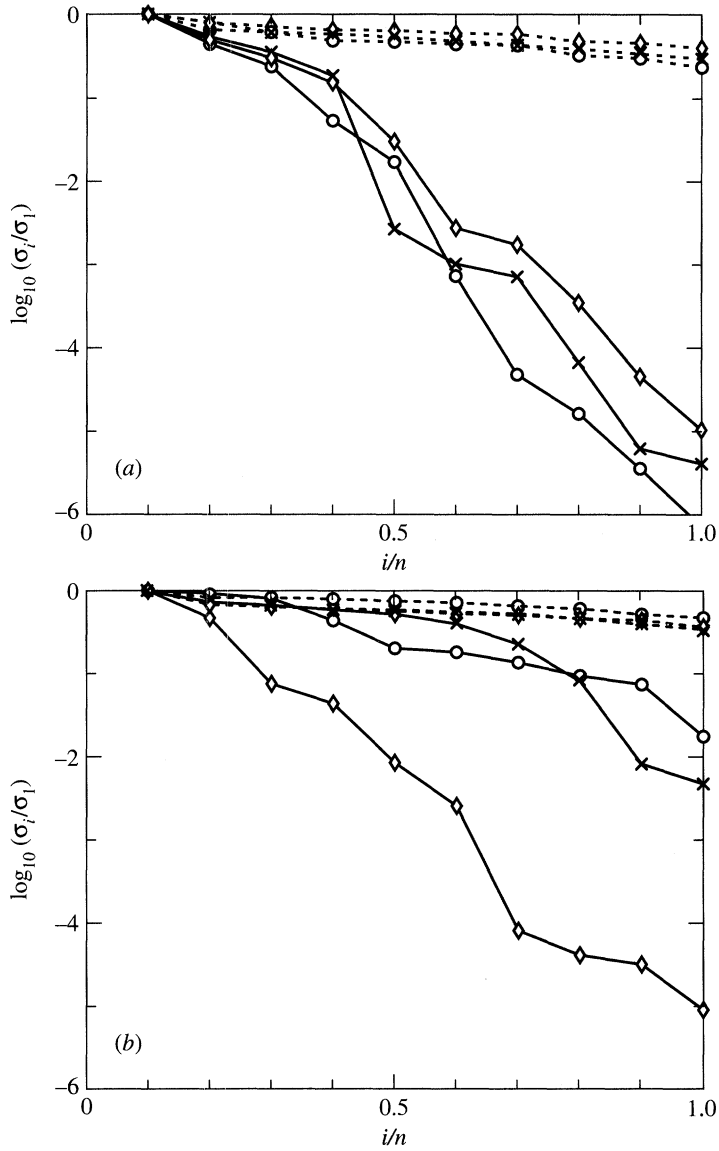

Figure 7. Normalized local singular values versus the scaled reconstruction dimension i/n. (a) Control data from the Hénon map (dark lines) and its 'stochastic surrogate' (dashed lines); (b) control data from the Lorenz map (dark lines) and its 'stochastic surrogate' (dashed lines). The three lines in each set (indicated by different markers) correspond to three different locations on the reconstructed attractor, chosen randomly according to its natural measure. In all the plots $\tau = 1$, $n = 10$ and the local neighbours are taken from a box size of linear dimension equal to 15% of the attractor extent. The number of local neighbours varies from 100 to 150.

What we learned from the analysis of the control data is that the SVD tests seem to be poor indicators of low-dimensionality. The global SVD test cannot be used to deduce the low-dimensionality of the control data. The local SVD test and the nonlinear correlation test give some indication of low-dimensionality. Strictly speaking, however, the strongest conclusion that could be drawn is that the control data have dimensions smaller than their 'stochastic surrogates'. The test for determinism,

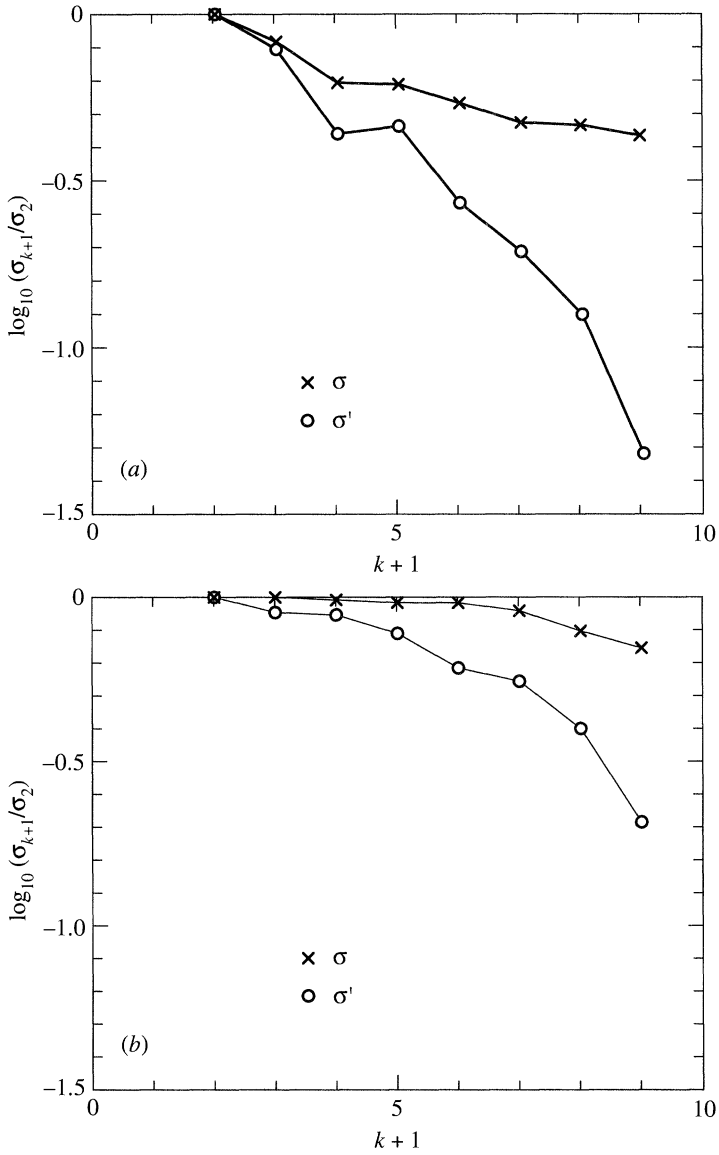

Figure 8. Nonlinear correlations among the amplitudes of the global singular vectors. The plots show σ_{k+1} and σ'_{k+1} for $\tau_w = 20$ and $\tau = 2$. The difference $\sigma_{k+1} - \sigma'_{k+1}$ is a measure of the nonlinear correlation between the amplitude of the $(k+1)$th singular vector and the first k singular vectors. (*a*) Control data from the Hénon map; (*b*) control data from the Lorenz map; (*c*) the 'stochastic surrogate' of (*a*); (*d*) the 'stochastic surrogate' of (*b*). 5000 data points were used in the computation of the nonlinear correlations.

clearly, is the best indicator of low-dimensionality. An alternative and, as it turns out, a more useful interpretation of the performance of *potential* indicators is the following: if a test does not indicate low-dimensionality, the confidence with which one could reject the hypothesis that a system has a low-dimensional characterization is highest for the test for recognizing determinism and lowest for the global SVD test. The interpretation of results of the tests on the T_{max} and T_{min} data should be made in this light.

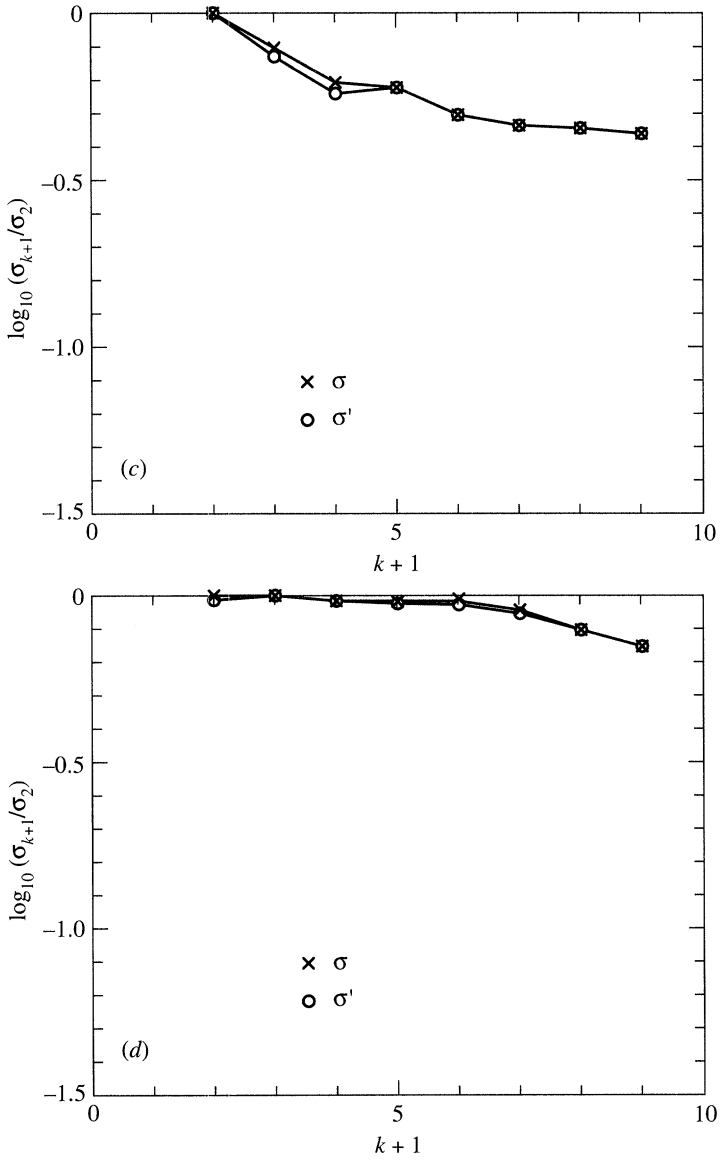

Figure 8. *Cont.*

(*b*) *Analysis of the temperature data*

(i) *SVD*

Figure 10 shows the result of the global SVD tests for the maximum temperature data (rescaled to have zero mean) with τ_w $(= (n-1) \times \tau)$ fixed at 50 and τ taking on values 10, 5 and 2. The computations were repeated for $\tau_w = 20$, $\tau = 5$ and 2 with no qualitative change. The spectra of singular values for the T_{max} and T_{min} data are nearly flat, although the first two singular values are slightly above the rest (especially in the $\tau = 2$ reconstruction). In view of the fact that the global singular values depend only the Fourier spectrum (discussed in §6 *a* (i)), their uniformity simply reflects the broadband spectra of the data.

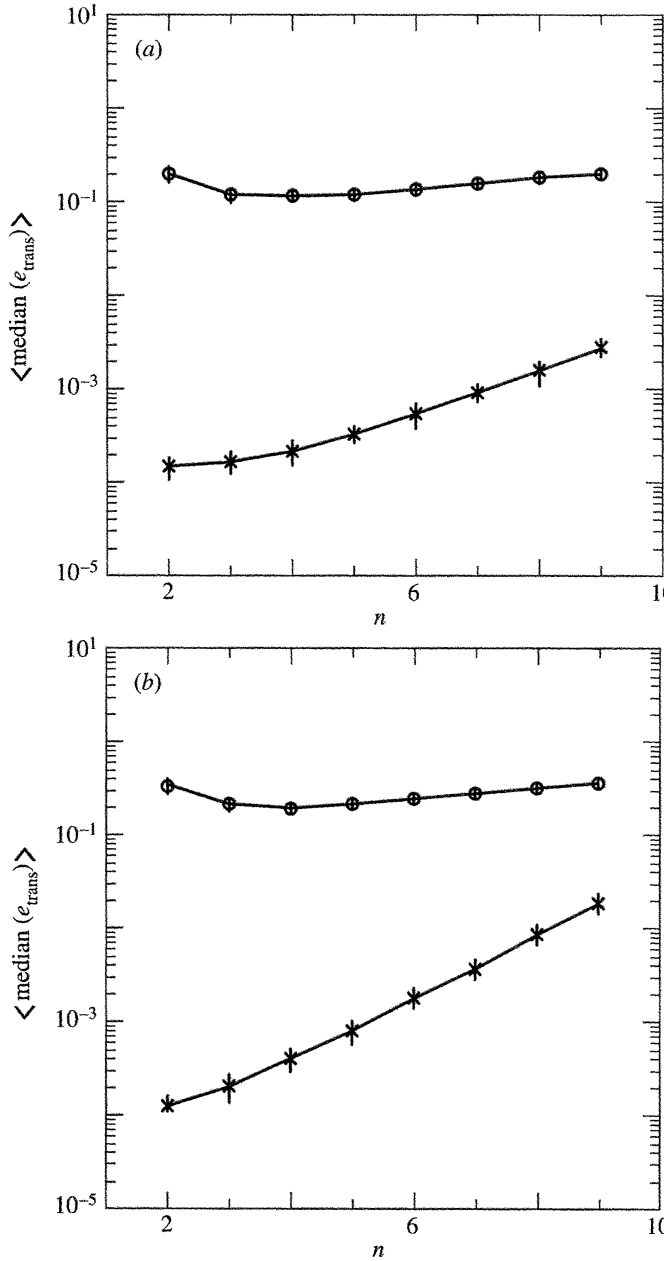

Figure 9. Average median translation error $\langle \text{median}(e_{\text{trans}}) \rangle$ versus the reconstruction dimension n. (a) Control data from the Hénon map (crosses) and its 'stochastic surrogate' (circles); (b) control data from the Lorenz map (crosses) and its 'stochastic surrogate' (circles). The plots are for $\tau = 1$ and $l = 4$. The median was obtained from a sample of 100 centres on the reconstructed attractor, chosen randomly according to its natural measure. The average $\langle \cdot \rangle$, was taken over 30 independent 1024-point samples of the data. The vertical lines in the plots represent ± 1 standard deviation.

It might be of interest to investigate whether one can associate the peaks in the Fourier spectrum, primarily the dominant peak corresponding to the annual cycle,

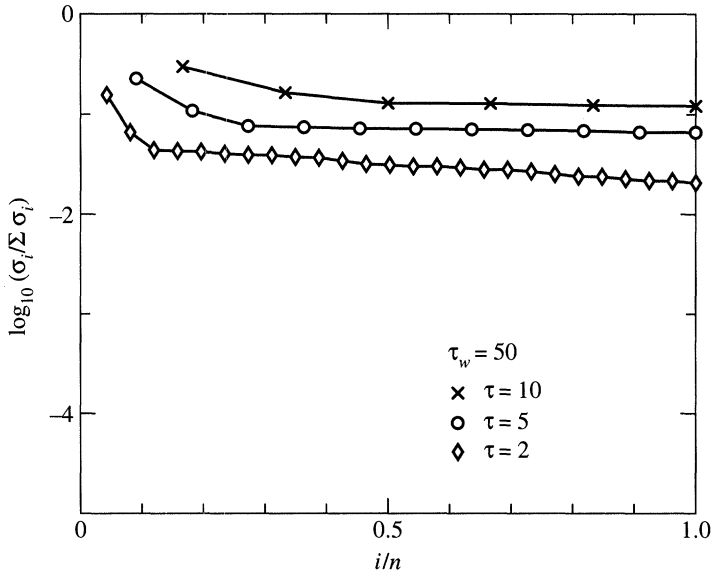

Figure 10. Normalized global singular values $\log_{10}(\sigma_i/\sum\sigma_i)$ versus the scaled reconstruction dimension i/n, for the $T_{\max}$ data (rescaled to have zero mean). 5000 data points are used to compute the singular values. This section of the data had only two breaks, each of length one. τ_w is fixed at 50 and τ takes on the values 10, 5 and 2.

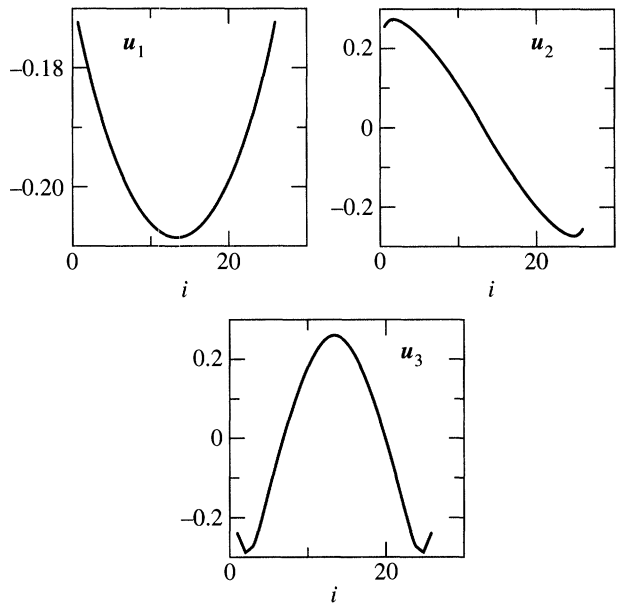

Figure 11. The first three singular vectors corresponding to the singular spectrum in figure 10 with delay-window $\tau_w = 50$ and delay-time $\tau = 2$. The delay-coordinate i goes from 1 to 26.

with the first two singular vectors. This would be in the spirit of earlier investigations of weather/climate data which made use of variants of global SVD to identify mean drifts and structural components. But our objective will be limited to identifying any trace of low-dimensional behaviour in the projection of the trajectory onto the subspace of the singular vectors. The first three singular vectors and the projection

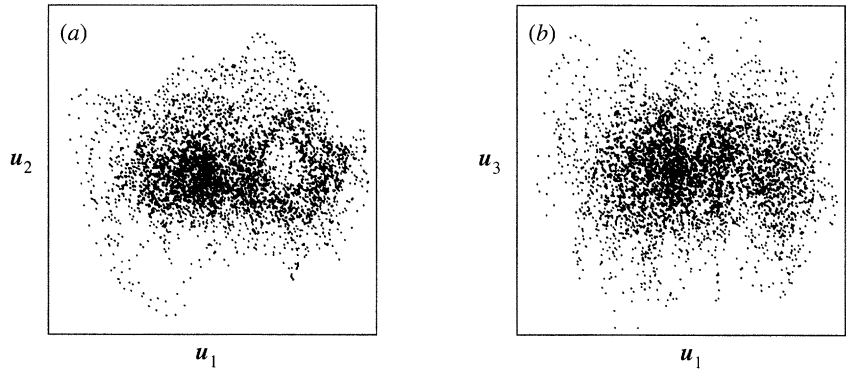

Figure 12. Projection of the reconstructed trajectory of the maximum temperature data onto the subspaces formed by the singular vectors shown in figure 11. (*a*) The projection of the trajectory onto the $(\boldsymbol{u}_1, \boldsymbol{u}_2)$-subspace; (*b*) the projection onto the $(\boldsymbol{u}_1, \boldsymbol{u}_3)$-subspace. The ith point on the $(\boldsymbol{u}_j, \boldsymbol{u}_k)$-plane is given by $(\boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{x}_i, \boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{x}_i)$, where $\boldsymbol{x}_i$ are the delay vectors and T represents transpose.

of the trajectory on the planes formed by singular vectors 1–2 and 1–3 are shown in figures 11 and 12. There seems to be no evidence of two-dimensional behaviour (limit cycle) in the 1–2 subspace. Furthermore, the projections on 1–2 and 1–3 subspaces are fairly identical. This argues against associating the two larger singular vectors with the annual cycle. Figure 13 shows smaller sections (corresponding to a year and to two years) of the projected trajectory. It appears that the phase portrait of each year can be identified with cycles in both the subspaces. It is thus clear that the considerable mismatch of cycles corresponding to the two years leads to the space-filling phase portraits in figure 12. A limit cycle (at least a noisy one) would probably be realizable if temperature at a point in the atmosphere were decoupled from (or very weakly coupled with) other state variables. The projection results are quite likely an indication of strong couplings and high-dimensional behaviour.

Figures 14*a*, *b* show the local singular values for the $T_{\max}$ and $T_{\min}$ data respectively, at three points on the attractor picked randomly according to the natural measure of the attractor. The local singular values are of similar magnitude, indicating that the local climate attractor is not low-dimensional. Unlike the control data, no nonlinear correlations are uncovered for the temperature data (figures 15*a*, *b*), indicating further that a low-dimensional characterization is *unlikely*.

(ii) *Testing determinism*

Figures 16*a*, *b* are plots of the average median translation error $\langle \mathrm{median}(e_{\mathrm{trans}}) \rangle$, versus the reconstruction dimension for the $T_{\max}$ and $T_{\min}$ data respectively, along with their corresponding 'stochastic surrogates'. The plots show that $\langle \mathrm{median}(e_{\mathrm{trans}}) \rangle$ for the original data and their 'stochastic surrogates' are the same and show no apparent dependence on the reconstruction dimension. This is a further indication, perhaps the strongest one, that a low-dimensional characterization for the atmospheric temperature data is *unlikely*.

8. Conclusions

The question which we set out to answer was: does the time series of daily maximum and minimum temperature ($T_{\max}$ and $T_{\min}$) have a low-dimensional charac-

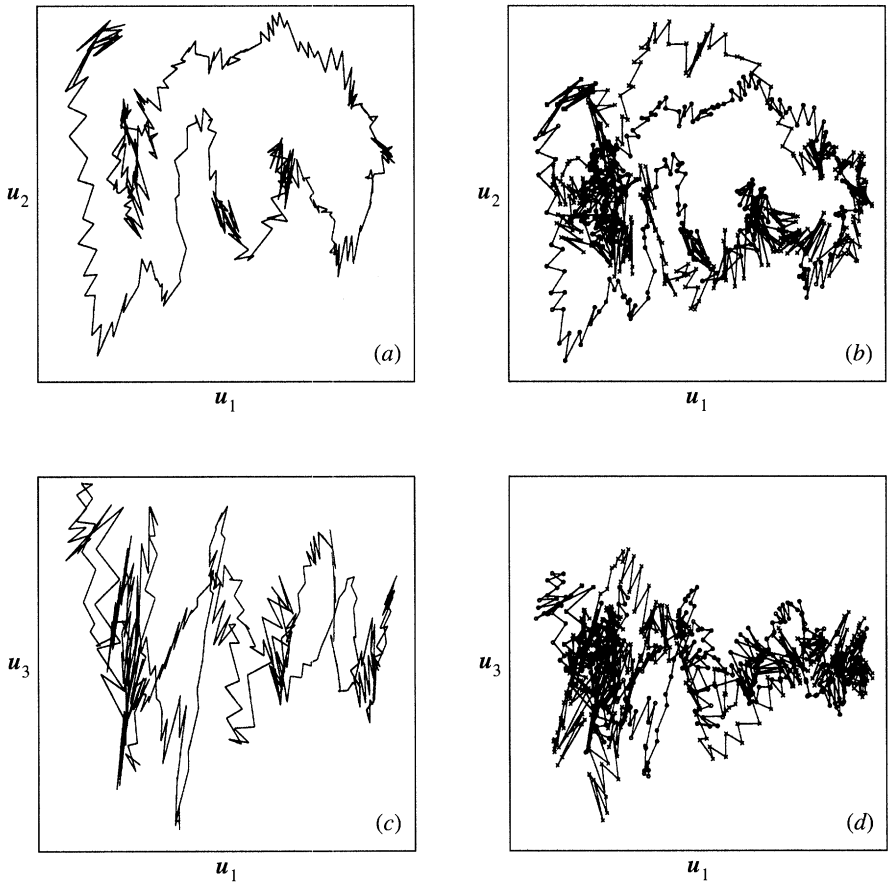

Figure 13. Small contiguous sections of the projected trajectory shown figure 12. (*a*), (*b*) Contiguous sections of the trajectory projected onto the $(\boldsymbol{u}_1, \boldsymbol{u}_2)$-subspace using trajectory points 1 to 370 and 1 to 730 respectively. (*c*), (*d*) The projections of the same parts of the trajectory onto the $(\boldsymbol{u}_1, \boldsymbol{u}_3)$-subspace. In the longer sections, (*b*) and (*d*), points 1 to 370 are indicated by circles and points 371 to 730 are indicated by crosses.

terization? We attempted to answer this question in two ways: (1) by estimating the dimension (direct method), and (2) by examining *potential* indicators of low-dimensionality (indirect method). The dimension estimates (via the nearest neighbour algorithm) do not show any tendency towards saturation with respect to the reconstruction dimension. Ideally, one would conclude from this that the dimension of the local climate attractor (the saturation value) will be larger than the maximum estimated value, which is 7.5 (for $T_{\max}$) and 7.0 (for $T_{\min}$). However, in view of the record length of the data, dimension estimates larger than 4 or 5 may not be reliable. Hence, based on the dimension estimates, the correct claim to make is that the dimension of the local climatic attractor is not smaller than 5, say. The indirect method comprises of using well-converged characterizations of the time series as potential indicators of low-dimensionality. Based on the outcome of these indicators (of which the test for determinism is the best) we conclude that it is *unlikely* that the local climate attractor has a dimension less than, say, 10.

Needless to say, these conclusions are appropriate for the *reconstructed attractor*. It is conceivable that the 'true' local climate attractor is low-dimensional, but is

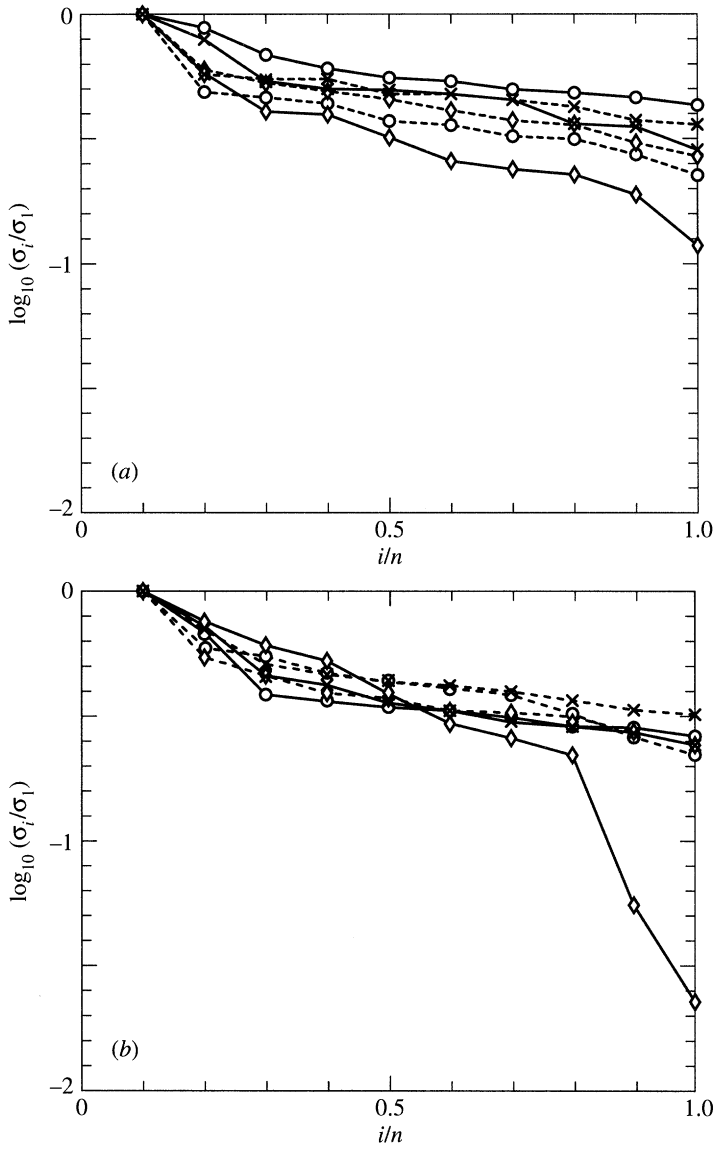

Figure 14. Normalized local singular values versus the scaled reconstruction dimension i/n. (a) The $T_{\max}$ data (dark lines) and its 'stochastic surrogate' (dashed lines); (b) the $T_{\min}$ data (dark lines) and its 'stochastic surrogate' (dashed lines). The three lines in each set (indicated by different markers) correspond to three different locations on the reconstructed attractor, chosen randomly according to its natural measure. In all the plots $\tau = 5$, $n = 10$ and the local neighbours are taken from a box size of linear dimension equal to 10% of the attractor extent. The number of local neighbours varies between 100 to 150.

'blurred' by noise which we are unable to remove. In view of the many sources of disturbance in the atmosphere, dynamic noise is likely to play an important part in determining the structure of the local climate attractor. In such a scenario, the 'true' local climate attractor will be very difficult to find and indeed might not exist. Thus, we could have a stochastic attractor with characteristics of the 'true' system, as for example in our case, a stochastic attractor with a strong periodic component.

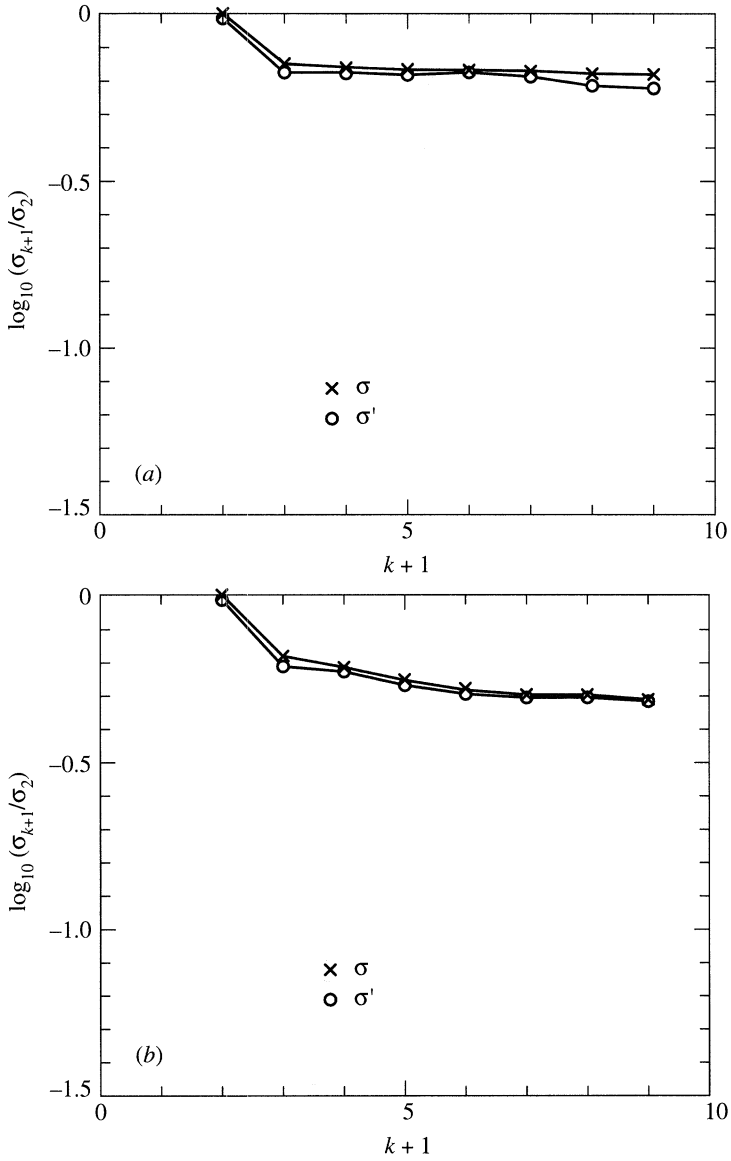

Figure 15. Nonlinear correlations among the amplitudes of the global singular vectors (similar to figure 8), with $\tau_w = 50$ and $\tau = 5$. (*a*) The T_{max} data; (*b*) the T_{min} data. The sections of data used are the same as that used for global SVD in figure 10.

The alternative to the stochastic attractor is a high-dimensional deterministic attractor. Because of practical difficulties, tools of dynamical system cannot be used to distinguish between the two alternative scenarios. Given the data length and computer resource requirements, it seems unlikely that such distinctions can be made in the near future. *The* distinguishing property of stochastic and deterministic (Markovian) systems is the transition probability, i.e. probability distribution of the possible future states given the present state. If the transition probability is a delta function then the system is deterministic, otherwise it is stochastic. Now, if the dynamics of

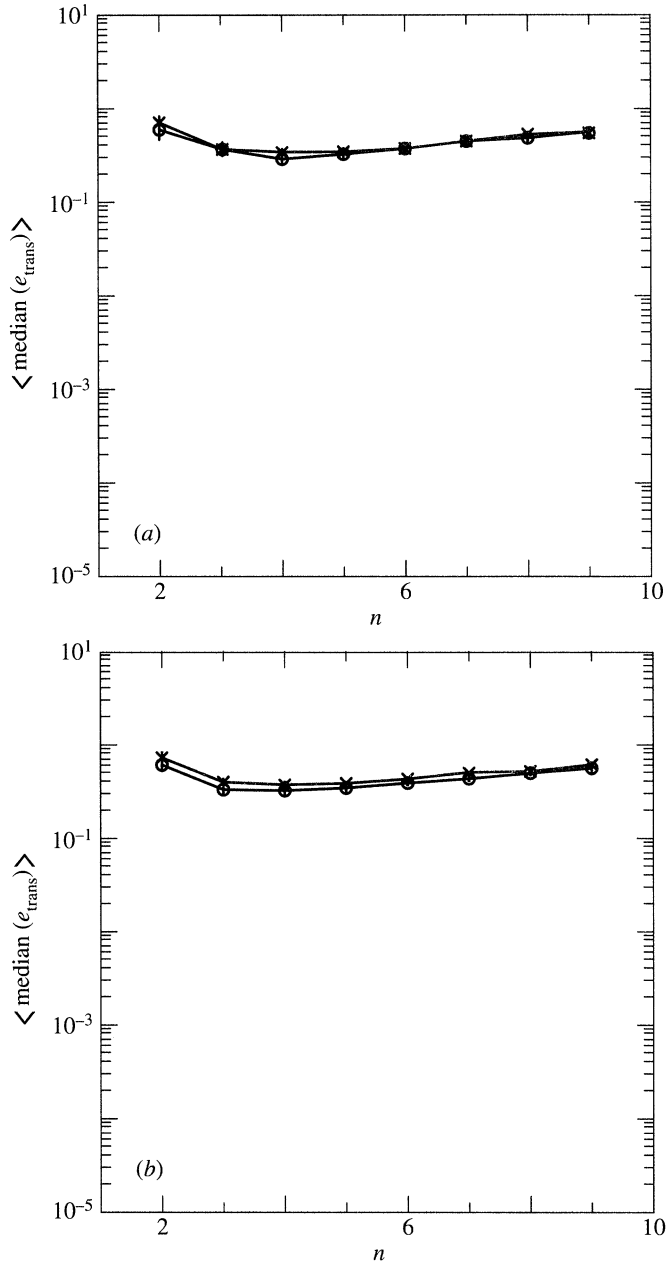

Figure 16. Average median translation error $\langle \text{median}(e_{\text{trans}}) \rangle$ versus the reconstruction dimension n. (a) The $T_{\max}$ data (crosses) and its 'stochastic surrogate' (circles); (b) the $T_{\min}$ data (crosses) and its 'stochastic surrogate' (circles). The plots are for $\tau = 10$ and $l = 4$. The median was obtained from a sample of 100 centres on the reconstructed attractor, chosen randomly according to its natural measure. The average $\langle \cdot \rangle$, was taken over twelve 2048-point non-overlapping sections of the data. The vertical lines in the plot represent ± 1 standard deviation.

the system cannot be reconstructed (because of high dimensionality) there seems little hope of distinguishing between a stochastic and high-dimensional attractor.

There are two directions of investigation, which *might* be worthwhile in the context

of the climate attractor. The first is to study the effects of dynamic noise on simple low-dimensional systems. The other direction of investigation would be to attempt an identification of spatial and temporal modes of the climate system via phase space reconstruction. After all, we are dealing with a spatially extended system and our investigations *via* tools of temporal chaos are complicated because of the coupling between spatial and temporal modes. Needless to say, the atmosphere is a far more complex beast than can be captured by single point measurements as attempted here.

We thank Professor D. D. Joseph who had some input in getting us interested in this problem, Professor Eric Kostelich for providing the noise reduction and nearest neighbour dimension codes and for his help in installing them, Dr. Richard Wayland for providing the code for the test for recognizing determinism in a time series, and Professor Dan Lathrop for some helpful suggestions. Finally, we thank one of the anonymous referees for his stimulating comments. The work was supported by the Air Force office of Scientific Research.

References

Atten, P., Caputo, J. G., Malraison, B. & Gagne, Y. 1984 Determination of attractor dimension of various flows. *J. Mech. Theor. Appl. Numero Spec.*, pp. 133–156.

Badii, R. & Politi, A. 1985 Statistical description of chaotic attractors: the dimension function. *J. Statist. Phys.* **40**, 725–750.

Broomhead, D. S. & King, G. P. 1986 Extracting qualitative dynamics from experimental data. *Physica* D **20**, 217–236.

Broomhead, D. S., Jones, R. & King, G. P. 1987 Topological dimension and local coordinates from time series data. *J. Phys.* A **20**, L563–L569.

Broomhead, D. S., Indik, R., Newell, R. C. & Rand, D. A. 1991 Local adaptive Galerkin bases for large-dimensional dynamical systems. *Nonlinearity* **4**, 159–197.

Constantin, P., Fois, C. & Témam, R. 1985 Attractors representing turbulent flow. *Mem. Am. Math. Soc.* **53**, 314.

Devijver, P. A. & Kittler, J. 1982 *Pattern recognition: a statistical approach.* New York: Prentice-Hall.

Eckmann, J. P. & Ruelle, D. 1985 Ergodic theory of chaos. *Rev. Mod. Phys.* **57**, 617–655.

Eckmann, J. P. & Ruelle, D. 1992 Fundamental limitations for estimating dimensions and lyapunov exponents in dynamical systems. *Physica* D **56**, 185–187.

Foias, C. & Témam, R. 1979 Some analytic and geometric properties of the solutions of the evolution Navier–Stokes equations. *J. Math. Pure Appl.* **58**, 339–368.

Fraedrich, K. 1986 Estimating the dimensions of weather and climate attractors. *J. Atmos. Sci.* **43**, 419–432.

Fraser, A. M. & Swinney, H. L. 1986 Independent coordinates for strange attractors from mutual information. *Phys. Rev.* A **33**, 1134–1140.

Grassberger, P. 1986 Do climatic attractors exist? *Nature* **323**, 609–612.

Grassberger, P. & Procaccia, I. 1983 Characterization of strange attractors. *Phy. Rev. Lett.* **50**, 346–350.

Grassberger, P., Hegger, R., Kantz, H., Schaffrath, C. & Schreiber, T. 1993 On noise reduction methods for chaotic data. *Chaos* **3**(2), 127–141.

Healey, J. J. 1994 Identifying finite dimensional behavior from broadband spectra. *Phys. Lett.* A **187**, 59–66.

Hentschel, H. G. E. & Procaccia, I. 1983 The infinite number of generalized dimensions of fractals and strange attractors. *Physica* D **8**, 435–444.

Hunt, B. R., Sauer, T. & Yorke, J. A. 1993 Prevalence: a translation-invariant 'almost every' on infinite dimensional spaces. *Bull. Am. Math. Soc. (N.S.)* **27**, 217–224.

Keepenne, C. L. & Nicolis, C. 1989 Global properties and local structure of the weather attractor over western europe. *J. Atmos. Sci.* **46**, 2356–2370.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Kostelich, E. J. & Schreiber, T. 1993 Noise reduction in chaotic time-series data: a survey of common methods. *Phys. Rev.* E **48**, 1752–1763.

Kostelich, E. J. & Swinney, H. L. 1987 Practical considerations in estimating dimensions from a time series. In *Chaos and related natural phenomena* (ed. I. Procaccia & M. Shapiro), pp. 141–156. New York: Plenum.

Kostelich, E. J. & Yorke, J. A. 1990 Noise-reduction: finding the simplest dynamical system consistent with the data. *Physica* D **41**, 183–196.

Lorenz, E. N. 1956 Empirical orthogonal functions and statistical weather prediction. Rep. no. 1. Statistical Forecasting Project, MIT.

Lorenz, E. N. 1991 Dimension of weather and climate attractors. *Nature* **353**, 241–244.

Lumley, J. L. 1970 *Stochastic tools in turbulence*. New York: Academic Press.

Maasch, K. A. 1989 Calculating climatic attractor dimension from δ^{18} records by the Grassberger–Procaccia algorithm. *Clim. Dynam.* **4**, 45–55.

Mallet-Paret, J. 1976 Negatively invariant sets of compact maps and an extension of a theorm of Cartwright. *J. Diffl Equat.* **22**, 331–348.

Mandelbrot, B. 1982 *The fractal geometry of nature*. San Fransisco: Freeman.

Nicolis, C. & Nicolis, G. 1984 Is there a climatic attractor? *Nature* **311**, 529–532.

Packard, N., Crutchfield, J., Farmer, D. & Shaw, R. 1980 Geometry from time series. *Phys. Rev. Lett.* **45**, 712–716.

Powell, M. J. D. 1985 In *Proc. IMA Conf. on Algorithms for the Approximation of Functions and Data*. RMCS Shrivenham.

Procaccia, I. 1988 Complex or just complicated? *Nature* **333**, 489–499.

Ruelle, D. 1982 Large volume limit of the distribution of characteristic exponents in turbulence. *Communs Math. Phys.* **87**, 287–302.

Sauer, T. & Yorke, J. A. 1993 How many delay coordinates do you need? *Int. J. Bifurc. Chaos* **3**, 737–744.

Sauer, T., Yorke, J. A. & Casdagli, M. 1991 Embedology. *J. Statist. Phys.* **65**, 579–616.

Sharifi, M. B., Georgakakos, K. P. & Rodriguez-Iturbe, I. 1990 Evidence of deterministic chaos in the pulse of storm rainfall. *J. Atmos. Sci.* **47**, 888–893.

Sirovich, L. 1989 Empirical eigenfunctions and low dimensional systems. In *New perspectives in turbulence* (ed. L. Sirovich), pp. 139–163. Springer.

Smith, L. A. 1988 Intrinsic limits on dimension calculations. *Phys. Lett.* A **133**(6), 283–288.

Takens, F. 1981 Detecting strange attractors in fluid turbulence. In *Dynamical systems and turbulence* (ed. D. Rand & L. S. Young), vol. 898, pp. 366–381. Berlin: Springer.

Theiler, J. 1990 Estimating fractal dimension. *J. Opt. Soc. Am.* A **7**, 1055–1073.

Tsonis, A. A. & Elsner, J. B. 1990 Comments on 'dimension analysis of climate data'. *J. Clim.* **3**, 1502–1505.

Vautard, R. & Ghil, M. 1989 Singular spectrum analysis in nonlinear dynamics, with applictions to paleoclimatic time series. *Physica* D **8**, 395–424.

Wayland, R., Bromley, D., Pickett, D. & Passamante, A. 1993 Determinism in a time series. *Phy. Rev. Lett.* **70**, 580–582.

Zeng, X., Pielke, R. A. & Ekyholt, R. 1992 Estimating the fractal dimension and the predictability of the atmosphere. *J. Atmos. Sci.* **49**, 649–659.

Received 22 February 1994; revised 31 May 1995; accepted 2 November 1995