# Chaos and Deterministic versus Stochastic Nonlinear Modeling

Martin Casdagli

SFI WORKING PAPER: 1991-07-029



**SANTA FE INSTITUTE**

# Chaos and Deterministic versus Stochastic Nonlinear Modeling

**Martin Casdagli**
Santa Fe Institute, 1120 Canyon Road
Santa Fe, New Mexico 87501

**Abstract**

An exploratory technique is introduced for investigating how much of the irregularity in an aperiodic time series is due to low-dimensional chaotic dynamics, as opposed to stochastic or high-dimensional dynamics. Nonlinear models are constructed with a variable smoothing parameter which at one extreme defines a nonlinear deterministic model, and at the other extreme defines a linear stochastic model. The accuracy of the resulting short-term forecasts as a function of the smoothing parameter reveals much about the underlying dynamics generating the time series. The technique is applied to a variety of experimental and naturally occurring time series data, and the results are compared to dimension calculations.

Keywords: Chaos, Dimension, Nonlinear modeling, Randomness, Time series.

## 1 Introduction

The nonlinear modeling and forecasting of time series data has a relatively recent history. The statistics community has constructed stochastic nonlinear models since about 1980; for a review see Tong(1990). Independently, the dynamical systems community, motivated by the phenomenon of chaos, has constructed deterministic nonlinear models since about 1987; see Crutchfield and MacNamara(1987), Farmer and Sidorowich(1987,1988) and Casdagli(1989) and references therein. In this paper, a forecasting algorithm is presented which constructs a range of models which attempts to "bridge the gap" between the stochastic and deterministic approaches. The forecasting algorithm is applied to a variety of experimental and naturally occurring time series data, in order to investigate whether the data exhibit low-dimensional chaotic behaviour, as opposed to high-dimensional or stochastic behaviour. The idea is that if models near the deterministic extreme give the most accurate short-term forecasts, then this is strong evidence for low-dimensional chaotic behaviour in the data. Similar ideas have recently been applied to economic data, and no evidence was found for nonlinear forecastability or low-dimensional chaos (Briggs(1990), Hseih(1991)).

Several of the time series analysed in this paper are generated from nonlinear dynamical systems which are expected to be high-dimensional, or highly stochastic. The forecasting algorithm can be used to investigate whether the nonlinearity underlying such time series can be detected by fitting a range of nonlinear stochastic models of low dimension. Preliminary results suggesting nonlinear forecastability in sunspot

data have been obtained by Farmer and Gibson using these ideas; see Casdagli et al.(1991b).

Rather than presenting a detailed application to one specific time series, in this paper results are presented for a wide variety of time series in order to provide the reader with a perspective of the domain of applicability of the above ideas. Careful hypothesis testing is avoided in favour of a more exploratory approach to data analysis, and the results in this paper should be viewed as preliminary investigations. The organisation of this paper is as follows. Section 2 contains a brief summary of ideas in dynamical systems which are relevant to time series analysis; for more information see the reviews by Eckmann and Ruelle(1985), Eubank and Farmer(1989), and Grassberger et al.(1991). Results in this paper will be interpreted using these concepts. In Section 3 a forecasting algorithm with a variable smoothing parameter is presented for nonlinear modeling, and its properties illustrated with numerically generated time series. In Section 4 the forecasting algorithm is applied to experimentally generated time series from electronic circuits, fluid turbulence and flame dynamics. In Section 5 the forecasting algorithm is applied to time series from speech, EEGs (ElectroEncephaloGrams), measles and sunspots. In Section 6 the results of the forecasting algorithm are compared to results from correlation dimension calculations, which have also been used to identify low-dimensional chaos. In Section 7 the main conclusions and open questions are summarised.

# 2 Chaos and deterministic modeling

Chaos is about the irregular behaviour of solutions to deterministic equations of motion, and has received much attention from mathematicians and physicists over recent years. The equations must be nonlinear to generate chaotic solutions, but apart from that can be remarkably simple. A nonlinear difference equation in one variable can generate chaos and an ODE (Ordinary Differential Equation) in three variables can generate chaos. Chaotic solutions exhibit broad band spectra, and masquerade as random time series when analysed using linear techniques. Chaotic solutions are only accurate for a length of time governed by the errors on initial conditions and the *Lyapunov exponent* of the system, which quantifies the exponential divergence of trajectories in chaotic systems. However, when considered in the underlying state space, in many cases chaotic solutions relax onto a *strange attractor* which has a fractal structure and typically a non-integral *dimension.*

Given that deterministic equations in a small number of variables can generate complicated behaviour, the question arises: how much of the complicated behaviour observed in nature can in fact be described by deterministic equations with a small number of variables? The obvious answer is none: at a small enough scale the laws of quantum mechanics apply, and a fully deterministic description is fundamentally impossible. On the other hand, at larger scales, many equations derived from the laws of physics are deterministic, and hold to an excellent degree of accuracy. Moreover, there are several examples of such equations which are nonlinear, have a small number

of variables, and exhibit chaotic behaviour. Examples range from equations describing mechanical and electrical systems to the motion of the orbit of Pluto. In these low degree of freedom systems, deterministic ODEs derived from the laws of physics can be used to numerically establish the existence of chaos, and to perform short-term forecasting and time series analysis.

More recently it has been shown that deterministic ODEs with a small number of variables can accurately model phenomena in the physics of many degree of freedom systems near the transition to chaos. Motivated by results in dynamical systems theory, Ruelle and Takens(1971) conjectured that the transition to turbulence observed in fluid dynamics experiments may be explained by a bifurcation to a low-dimensional strange attractor. This conjecture, which was in direct conflict with the high-dimensional Landau picture of turbulence, has been confirmed by the universality results of Fiegenbaum, mathematical results in PDEs (Partial Differential Equations) (for example see Doering et al.(1988) and references therein), as well as by numerical and physical experiments. Thus, near the transition to chaos, the many degrees of freedom are in fact coupled together coherently, and an enormous dimension reduction takes place.

The dimension of the attractors of many degree of freedom systems may increase rather rapidly after the transition to chaos. Thus the problem arises as to how to estimate the dimension of the strange attractor which underlies an irregular time series generated by a many degree of freedom system such as a fluid. Unfortunately, it is extremely difficult to estimate the dimension of attractors by numerical integration of PDEs. Dimension estimation is an important problem, because if the time series exhibits low-dimensional behaviour, then it should be possible to accurately model the underlying system with a small number of ODEs, rather than numerically intractable PDEs, and hence obtain insight into the behaviour of the system. This problem is also of great interest in systems for which the underlying physics is not understood.

Techniques of state space reconstruction were introduced by Packard et al.(1980) and Takens(1981), which show that it is, in principle, possible to address the above problem of dimension estimation by direct observations of a long enough time series of measurements of the system of interest, as follows. Suppose an observed scalar time series $x_1, x_2, ..$ is generated by a $D$-dimensional attractor of a deterministic dynamical system with $d$ degrees of freedom

$$s(t) = F^t(s(0)) \tag{1}$$

$$x_i = h(s(i\tau_s)) \tag{2}$$

where $s(t) \in \Re^d$ denotes the state of the system at time $t$, $F^t$ is the time-$t$ evolution of a smooth dynamics, $\tau_s$ is the sampling rate, and $h : \Re^d \rightarrow \Re$ is a scalar valued measurement function. Define the *delay vectors* $\underline{x}_i$ of *embedding dimension* $m$ and *delay time* $\tau$ (which by convention I take to be an integer in this paper) by

$$\underline{x}_i = x_i, x_{i-\tau}, .., x_{i-(m-1)\tau} \tag{3}$$

Takens proved that if $m > 2d$ then, subject to genericity assumptions on $F$, $h$ and $\tau_s$, a smooth dynamics is induced on delay vectors, so that for all integers $T$ and $\tau$, there exists a smooth map $f^T : \Re^m \to \Re$ such that

$$x_{i+T} = f^T(x_i, x_{i-\tau}, .., x_{i-(m-1)\tau}) \tag{4}$$

for all integers $i$. For geometrical illustrations of Takens' theorem, and generalisations to the case of low-level "observational noise" (i.e noise added to the right hand side of (2)), see Casdagli et al.(1991a). Takens' theorem implies that the delay vectors $\underline{x}_i$ fill out a "reconstructed" attractor in $\Re^m$ which is diffeomorphic to the original attractor. In particular, the dimension of the reconstructed attractor is $D$, independently of $m$, and is invariant with respect to the measurement function $h$. Stronger results of this kind have been established by Sauer et al.(1990). In particular, it is typically sufficient to take $m > 2D$ for (4) to hold, and if $D < m < 2D$, (4) holds for almost everywhere. By contrast, if the time series is generated by a stationary stochastic process then, for all values of $m$, a noise term is expected to appear on the right hand side of (4) and the delay vectors are expected to fill out a set of dimension $m$.

A variety of algorithms for time series analysis have been proposed to exploit the above mathematical results. In the cases $m = 1, 2, 3$ various graphical techniques can be used to investigate the geometry and dynamics of the state space reconstruction. For example, a "phase portrait" obtained by plotting $x_i$ against $x_{i-\tau}$ often provides much insight, and is used throughout this paper to display time series data. In some applications, graphical techniques give striking evidence for low-dimensional chaos. Examples include data from a dripping faucet (Shaw(1984)), experiments in fluid dynamics (see Eckmann and Ruelle(1985)) and biological oscillations (see Glass and Mackey (1988)). "Recurrence plots" introduced by Eckmann et al.(1987) are useful for the analysis of non-stationary data. The correlation dimension algorithm of Grassberger and Procaccia (1983) is a numerical technique for investigating the geometry of a reconstructed attractor in higher embedding dimensions than those available to most graphical techniques, and will be considered in Section 6.1.

In this paper, attention is focused on forecasting algorithms, which attempt to approximate the unknown nonlinear map $f$ in (4) from time series data $x_1, .., x_N$. In practice there is always a noise term on the right hand side of (4) due to observational errors and, for example, thermodynamic fluctuations. However, if the effects of noise are small, and if the dimension $D$ of the attractor is low, then with a modest amount of data (say $N \approx 10^D$), it is possible to obtain an accurate approximation to $f$. The approximation can be tested by making out of sample, short-term forecasts, which should be much more accurate than expected if there were a substantial stochastic component on the right hand side of (4). On the other hand, if the dimension $D$ is large, there will be insufficient data to approximate (4) with a deterministic model with $m > D$, and one is forced to approximate (4) with a (possibly nonlinear) stochastic model with $m < D$. In this weak sense, a high-dimensional deterministic system is equivalent to a stochastic system. Moreover, it can be shown that, for high-dimensional chaotic systems with low-level observational noise, a large noise term in

(4) can be induced by the process of state space reconstruction from a scalar time series, so that even a knowledge of the functional forms of $F$ and $h$ in (1,2) does not allow for accurate short-term forecasting of the time series, irrespective of the length of the time series (Casdagli et al.(1991a)). In this strong sense, a high-dimensional chaotic system is equivalent to a stochastic system. The forecasting algorithm presented in Section 3 is designed to explore a wide variety of models from the nonlinear deterministic extreme to the linear stochastic extreme, in order to investigate the dimensionality of the time series.

The properties of low-dimensional chaotic systems with low levels of noise are now rather well understood, even mathematically. However, although low-dimensional chaos is expected to be common in few degree of freedom systems, it is expected to be rare in many degree of freedom systems, except near the transition to chaos. Consequently, the attention of many physicists has shifted to the study of high-dimensional systems. Several idealised models have been studied numerically, and a variety of interesting "collective phenomena" observed, such as spatio-temporal chaos (see Crutchfield and Kaneko(1988) and references therein), robust intermittency (Keeler and Farmer(1986)) and self-organised criticality (Bak et al.(1987)). There have been some preliminary attempts to fit such models to spatio-temporal data. The phenomena discovered in these idealised models may also have qualitative implications for the analysis of scalar time series from many degree of freedom systems. In particular, some of the models provide potential alternatives to fractional Brownian motion models, which have been used to investigate "long-memory" time series with slowly decaying power spectra. However, it is not clear how to develop quantitative tests for time series analysis which distinguish between these alternatives. In addition to its use for detecting low-dimensional chaos, the forecasting algorithm presented in Section 3 can be used to investigate whether time series from high-dimensional deterministic systems can be approximately modeled with low-dimensional nonlinear stochastic models, as has been suggested in the case when there are large, spatially coherent structures in the system (Broomhead et al.(1990)).

# 3 A forecasting algorithm

## 3.1 Description

The idea behind the forecasting algorithm used in this paper is straightforward: construct piecewise-linear approximations to the unknown function $f^T$ of (4) using a variable number $k$ of neighbours. A small value of $k$ corresponds to a deterministic approach to modeling (Farmer and Sidorowich(1987)). The largest value of $k$ corresponds to fitting a stochastic linear autoregressive model. Intermediate values of $k$ correspond to fitting nonlinear stochastic models. Nonlinear stochastic models of a similar type have been constructed by Tong and Lim(1980), with relatively large $k$ values, in the case of short time series. The algorithm described below is a computationally efficient implementation of this idea; the longest computations presented in

this paper took a few hours of CPU time on a SUN4 workstation.

(A) Divide the time series into two separate parts; a *fitting* set $x_1, .., x_{N_f}$ and a *testing set* $x_{N_f+1}, .., x_{N_f+N_t}$.

(B) Choose an embedding dimension $m$, a delay time $\tau$, and a forecasting time $T$.

(C) Choose a delay vector $\underline{x}_i$ with $i \geq N_f$ for a $T$-step ahead forecasting test.

(D) Compute the distances $d_{ij}$ of the test vector $\underline{x}_i$ from the delay vectors $\underline{x}_j$, $1 + (m-1)\tau \leq j \leq N_f - T$ in the fitting set, using the maximum norm for computational efficiency.

(E) Order the distances $d_{ij}$, find the $k$ nearest neighbours $\underline{x}_{j(1)}, .., \underline{x}_{j(k)}$ of $\underline{x}_i$, and fit an affine model of the form (5), where the parameters $\alpha_0, .., \alpha_m$ are computed by ordinary least squares.

$$x_{j(l)+T} \approx \alpha_0 + \sum_{n=1}^{m} \alpha_n x_{j(l)-(n-1)\tau} \qquad l = 1, .., k \tag{5}$$

Vary the number of nearest neighbours $k$ at several representative values between $2(m+1)$ and $N_f - T - (m-1)\tau$. I used the "heapsort" algorithm described in Press et al.(1988) to order the distances $d_{ij}$. The parameters $\alpha_0, .., \alpha_m$ can be obtained by solving the normal equations for the linear system (5), which can be updated recursively as $k$ is increased and solved using LU decomposition for computational efficiency.

(F) Use the fitted model (5) to estimate a $T$ step ahead forecast $\hat{x}_{i+T}(k)$ for the test vector $\underline{x}_i$, and compute its error $e_i(k) = |\hat{x}_{i+T}(k) - x_{i+T}|$.

(G) Repeat steps (C) through (F) for all $i$ in the testing set, and compute the normalised RMS forecasting error

$$E_m(k) = (\sum_i e_i^2(k))^{1/2}/\sigma \tag{6}$$

where $\sigma$ is the standard deviation of the time series. In long, finely sampled time series, I often space the delay vectors $\underline{x}_i$ used for testing $N_s > 1$ units of the sampling time apart. This reduces the amount of CPU time for a fixed amount of data, without significantly affecting statistical averages.

(H) Choose the delay time $\tau$ and the forecasting time $T$ by discretion. In chaotic systems, they should not be chosen too large. In finely sampled continuous time systems, they should not be chosen too small. In this paper, $\tau$ is chosen by trial and error, in order to obtain low values for $E_m(k)$. For a review of theoretical issues involved in choosing $\tau$, see Casdagli et al.(1991a). Finally, vary the embedding dimension $m$, and study the curves $E_m(k)$ as a function of $k$.

## 3.2 Numerical example

In order to test the algorithm, noise free time series were generated from numerical simulations of deterministic dynamical systems. Independent identically distributed Gaussian noise of variance $\epsilon^2$ was added to the noise free time series, using a pseudo-random number generator, to simulate observational noise. Define the *noise level* to be the ratio of $\epsilon$ to the standard deviation of the noise free time series. Fig. 1a shows $E_m(k)$ curves for an example with a 3.6 dimensional attractor and a 2% noise level. The noise free time series was obtained by sampling a numerical solution to the delay differential equation (7) of Mackey and Glass(1977) with parameter $\Delta = 30$, initial condition $x(t) = 0.9$ for $0 \leq t \leq \Delta$, and sampling rate $\tau_s = 6$.

$$\frac{dx(t)}{dt} = -0.1x(t) + \frac{0.2x(t-\Delta)}{1+[x(t-\Delta)]^{10}} \tag{7}$$

Fitting and testing sets were chosen with $N_f = 40000$, $N_t = 400$ and $N_s = 1$. Greatly improved forecasts of nonlinear models over linear models are obtained as the embedding dimension $m$ is increased, until a "plateau" is reached at $m = 6$. The most accurate forecasts are obtained with values of $k$ near the deterministic extreme, which allows one to correctly conclude that the time series is essentially low-dimensional.

Fig. 1b shows $E_5(k)$ curves for an example with an attractor of dimension 1.7, for noise levels 0%, 2%, 20% and 100%. The noise free time series was obtained from a model of the form (1,2) with the dynamics given by the map $f : \Re^2 \to \Re^2$ of Ikeda(1979) defined by

$$f(x,y) = (1 + a(x\cos t - y\sin t), a(x\sin t + y\cos t)) \tag{8}$$

where $t = 0.4 - 6.0/(1+x^2+y^2)$ and with the parameter $\mu = 0.9$. The measurement function was taken to be $h(x,y) = x$. It can be seen from Fig. 1b that greatly improved forecasts are obtained over the linear model for the noise levels 0% and 2%, and these time series are correctly identified as essentially low-dimensional. However, when the noise level is increased to 20%, the forecastability is lost to the extent that one can only conclude that the time series is nonlinear. It could be either low-dimensional and noisy, or high-dimensional. Finally, at a noise level of 100%, it is difficult to detect any nonlinearity at all from an inspection of Fig. 1b. It would be interesting to apply the forecasting algorithm to high-dimensional numerical time series data, but in the remainder of this paper I concentrate on analysing experimental and naturally occurring time series data.

## 3.3 Remarks

Theoretical properties of the $E_m(k)$ curves may be summarised as follows. In the case that the noise is negligible, forecasting errors are dominated by the inaccuracies in fitting a piecewise linear model to the smooth map $f$ in (4), and the scaling law

$$E_m(k) \sim C(k/N_f)^{2/D} \quad k/N_f \to 0 \tag{9}$$

is anticipated for $m > 2D$, where $C$ is a "curvature" constant and $D$ denotes the dimension of the attractor; compare Farmer and Sidorowich(1987) and Casdagli(1989). In general, a fractal attractor has a range of dimensions associated with it, and it can be argued that the precise value of $D$ in the scaling law (9) is given by $D_q$, where $q$ is the solution to $(q-1)D_q = -4$, and $D_q$ are the Renyi dimensions of the attractor (see Badii and Politi(1985) for definitions and similar arguments). The scaling law (9) can be shown to simplify if the RMS average (6) is replaced by the geometric mean; the appropriate value of $D$ in (9) is then the information dimension $D_1$. The dashed lines in Fig. 1 were obtained by substituting known values for $D_1$ into the scaling law (9), since, for each of these attractors, the relevant Renyi dimensions are close to $D_1$. The constant $C$ is difficult to estimate from first principles, and in Fig. 1 is chosen arbitrarily to illustrate the scaling behaviour (9) of $E_m(k)$ with $k$. In this regime of negligible noise, large values of $k$ should be avoided to obtain accurate forecasts, particularly when the dimension $D$ of the attractor is small.

In the case where observational noise of level $\epsilon$ is added to (2), a noise term is induced on the right hand side of (4), which enforces a limitation to forecastability as $N_f \to \infty$, and the scaling law (9) breaks down. This effect can be computed in a small $\epsilon$ approximation, and the horizontal line "l" in Fig. 1b was computed from the functional forms of $f$ and $h$ for this example (see Casdagli et al.(1991a)). In this noise dominated regime, small values of $k$ should be avoided to obtain the most accurate forecasts due to problems of statistical estimation (see Casdagli(1991)). An optimal choice of $k$ will in general involve a compromise between nonlinear deterministic effects and the effects of noise and statistical estimation. Similar considerations apply to optimising the choice of the embedding dimension $m$. Of course, if the underlying dynamics generating the time series is not understood, the parameters $k$ and $m$ should be varied to investigate the unknown dynamics, and this is the approach taken in Sections 4 and 5.

From a more practical point of view, there are some pitfalls in drawing conclusions from an inspection of the $E_m(k)$ curves as follows. First, in extreme situations, low-dimensional chaos can be misidentified as stochastic behaviour. In fact, all of the time series analysed in Fig. 1 came from low-dimensional deterministic dynamics, since the pseudo-random number generator that I used was based on the deterministic process $\epsilon_{i+1} = a\epsilon_i + b \pmod 1$. However, $a$ was taken very large, and the forecasting algorithm requires a very large amount of time series data to discover such rules. Pseudo-random number generators are somewhat pathological examples of low-dimensional chaos, and such misidentification problems are not expected with time series from natural processes, if an appropriately high sampling rate is used. Second, the results of the forecasting algorithm should be carefully interpreted in the case of non-stationary time series. In the applications to experimental time series presented in Section 4, it is implicitly assumed that the experiments are sufficiently well controlled so that stationarity holds to a good degree of approximation. Other problems with the interpretation of results from the forecasting algorithm will be discussed as they arise in later sections of the paper.

# 4 Experimental time series

## 4.1 Coupled diodes

Electrical circuits containing diodes are a useful source of experimental time series for testing ideas in nonlinear dynamics (for example see Gunaratne et al.(1989)). Diodes have a nonlinear, approximately deterministic, response to applied electrical currents. The forecasting algorithm of Section 3 was applied to time series data generated by measuring currents in circuits containing $n$ coupled diodes, for $n = 2, 4, 6$. Since these are essentially few degree of freedom systems, any irregular behaviour observed in the time series data is expected to be mostly due to low-dimensional chaos. The time series were sampled at the frequency of the periodic driving voltage, so it is natural to fix the delay time $\tau = 1$ and the forecasting time $T = 1$. Fig. 2a (resp. 2c, 2e) shows a phase portrait for the case $n = 2$ (resp. $n = 4, 6$). Fig. 2b (resp. 2d, 2f) shows the corresponding $E_m(k)$ curves. The fitting and testing sets were chosen with $N_f = 20000$, $N_t = 400$ and $N_s = 1$.

Evidently there is much structure in the phase portraits. The attractor of Fig. 2a appears to display a wide range of Renyi dimensions. In this case, the $E_m(k)$ curves were also computed using a geometric mean in place of (6); these are the dotted curves in Fig. 2b. This was the only case where I used the geometric mean. As remarked in Section 3.3, the geometric mean has some nice theoretical properties, but in practice I have often found it to have poor convergence properties as the size of the testing set is increased. It can be seen from Fig. 2b that deterministic (low $k$) models give substantial improvements in predictive accuracy over linear models, and the $n = 2$ time series is identified as low-dimensional chaos, with $D_1 \approx 3.5$ according to the scaling law (9). In the case $n = 4$, it can be seen from Fig. 2d that the improvement in predictive accuracy of the nonlinear models over linear models has declined considerably, and nonlinear stochastic models with $k$ values towards the deterministic extreme give about a 40% improvement in predictive accuracy. Finally, in the case $n = 6$, it can be seen from Fig. 2f that the nonlinear stochastic models at intermediate values of $k$ give only about a 10% improvement in predictive accuracy over linear models.

In this example, the results of the forecasting algorithm are only unambiguous in the case $n = 2$. One might conjecture that the reason for the decay in predictive accuracy as $n$ is increased from 2 to 6 is not only that the dimension of the reconstructed attractor is increased, but that the dynamics induced on the reconstructed attractor is more complicated, especially in view of the low sampling rate. On the other hand, one might conjecture that stochastic, thermal effects in diodes become more important as $n$ is increased. In Section 6.1 result are presented from dimension calculations and an attempt is made to resolve this dichotomy. It would also be interesting to address the above dichotomy by attempting to forecast the time series using information in the ODEs which are believed to accurately model diodes (see Brorson et al.(1983)). Algorithms have been developed by Geweke(1989) and Breeden and Hubler(1990) which might be useful for such problems.

## 4.2 Fluid turbulence

Experiments in fluid turbulence provide a rich source of time series data, due to the potentially enormous number of degrees of freedom available. The forecasting algorithm of Section 3 was applied to time series believed to be at the two extremes of weak turbulence and strong turbulence. The first time series, shown in Fig. 3a, was generated by measuring temperature differences between two plates enclosing a Rayleigh-Benard convection experiment near the transition to turbulence (see Haucke and Ecke(1987)). The second time series, shown in Fig. 3c, was generated by measuring a component of the velocity at a point in a Taylor-Couette experiment at a highly turbulent regime (Reynolds number $\approx 10^5$). Since these are many degree of freedom systems, low-dimensional chaos is only expected for the first time series, which is near the transition to turbulence. The application of the forecasting algorithm is complicated due to the absence of a natural choice for $\tau$ and $T$. A systematic variation of these parameters remains to be done. The choices used for Fig. 3 seem to be reasonable. Figs. 3a and 3c show phase portraits, and Figs. 3b and 3d show corresponding $E_m(k)$ curves. The fitting and testing sets were chosen with $N_f = 20000$, $N_t = 400$ and $N_s = 10$.

For the weakly turbulent time series, Fig. 3a reveals that the motion is confined to a toroidal region of phase space. In Fig. 3b deterministic (low $k$) models clearly give substantial improvements in predictive accuracy over linear models, and this time series is identified as low-dimensional chaos, with $D \approx 3.0$ according to the scaling law (9). This is consistent with the Ruelle-Takens conjecture, and the results of Hauke and Ecke(1987) and Farmer and Sidorowich(1987) on this time series. It can be seen from Fig. 3b that to obtain the most accurate forecasts, $m = 20$ is needed, which is surprisingly large. If the time series indeed has a low noise level then, from the results of Sauer et al.(1991), $m = 2D + 1 \approx 7$ should be sufficient.

By contrast, for the highly turbulent time series, Fig. 3c reveals little structure other than occasional intermittent bursts. In Fig. 3d, linear models are seen to be superior to deterministic models, and nonlinear stochastic models at intermediate values of $k$ give only about a 3% improvement in predictive accuracy over linear models. In this case the time series is identified as high-dimensional with possibly a small amount of nonlinearity which can be exploited by nonlinear stochastic modeling in low embedding dimensions $m$. However, it is not clear that a linear stochastic model is adequate to describe qualitative features of this time series, because of the large intermittent bursts apparent in Fig. 3c. As mentioned at the end of Section 2, such intermittent bursts often occur as the result of collective phenomena in high-dimensional systems, and are expected to be a fundamentally nonlinear effect.

## 4.3 Flame dynamics

Experiments in flame dynamics have been performed by Gorman and Robbins(1991) in order to investigate spatio-temporal chaos in such systems. Phase portraits of time series obtained from measurements of the light intensity at a point in the flame are

shown in Fig. 4. Spectral methods (extensively used by Gorman and Robbins) indicate that the time series shown in Fig. 4a comes from a quasiperiodic, non-chaotic regime, and indeed the phase portrait resembles a two-dimensional torus. For the time series shown in Fig. 4c, the spectrum is broad band, and direct observations of the spatio-temporal motion of the flame (which have been recorded on videotape) reveal a strongly spatially coherent motion, so that this time series is expected to exhibit low-dimensional chaos. For the time series shown in Fig. 4e, the spectrum is slowly decaying, and direct observations of the spatio-temporal motion of the flame reveals complicated patterns of about 20 interacting pockets of flame activity, known as "cellular flames", so that this time series is expected to be of rather high dimension. The structures of the phase portraits shown in Fig. 4 are consistent with these observations. Again, the application of the forecasting algorithm of Section 3 is complicated due to the absence of a natural choice for $\tau$ and $T$. The choices used in Fig. 4 seem to be reasonable. The fitting and testing sets were chosen with $N_f = 15000$, $N_t = 400$ and $N_s = 10$. Since the total length of each time series was $N \approx 200000$, only every tenth point was retained before applying the forecasting algorithm in the cases where the time series were finely sampled.

The results shown in Fig. 4f for cellular flames are similar to those obtained in the case of the highly turbulent fluid shown in Fig. 3d. Even though the turbulence is not fully developed in the case of cellular flames, the underlying nonlinear dynamics is not of a form which can be detected using this forecasting algorithm. The results shown in Figs. 4b and 4d are more difficult to interpret. Fig. 4b at first seems to contradict the graphical evidence for approximately deterministic motion on a two-dimensional torus shown in Fig. 4a, since the improvements in accuracy of nonlinear models over linear models is only about 50%. However, this is not surprising, since the non-chaotic motion in this example is sufficiently simple that it might be expected to be well approximated by a linear model. On the other hand, it can be seen from Fig. 4d that the nonlinear stochastic models at intermediate values of $k$ give more than 100% improvement in predictive accuracy over linear models. Similar behaviour occurred in the $E_m(k)$ curves for the low-dimensional chaotic example of Fig. 1b at a 20% noise level, and in the 4-coupled diodes example of Fig. 2d. At this stage I can only conclude that the time series shown in Fig. 4c is either low-dimensional with a moderate amount of noise (this could be dynamical noise, i.e. noise added to the right hand side of (1), rather than observational noise, at a level of say 20%), or of moderate dimension (say $D \approx 6$). Further information is required to resolve this dichotomy. Since direct observations of the spatio-temporal motion reveal strong spatial coherence in this example, I suspect that the low-dimensional possibility is more likely. A quantitative analysis of the spatio-temporal time series data could potentially resolve this issue, but unfortunately the data is not yet available in digitised form.

# 5 Observational time series

The time series analysed in Section 4 were generated by carefully controlled experiments on physical systems. By contrast, in this section I analyse time series generated by a variety of "naturally occurring" systems which are less well understood, and which in some cases cannot be controlled at all.

## 5.1 Speech

Improved techniques for the short-term forecasting of speech time series are of potential commercial use in transmission coding, and there is evidence for nonlinear structure in speech time series (Kumar and Mullik(1990), Townshend(1990)). The forecasting algorithm of Section 3 was applied to a time series of pressure fluctuations recorded from the sentence "We lost the golden chain". A phase portrait of part of the time series is shown in Fig. 5a. The large amplitude, low-frequency, oscillations correspond to vowels. Fig. 5c shows a phase portrait of the time series after it has been passed through a moving pole linear filter which is designed to extract both linear and non-stationary structure from the original time series. The corresponding $E_m(k)$ curves shown in Figs. 5b and 5d were obtained by taking $N_f = 20000$, $N_t = 400$ and $N_s = 60$. In Fig. 5b (resp. 5d) the best nonlinear stochastic models outperform linear models by a factor of about 50% (resp. 100%). A clear interpretation of these results is complicated due to the non-stationarity of this short sentence (this is evident in Fig. 5d, since linear models give a value of $E_m(k)$ significantly greater than 1). The effects of non-stationarity would be expected to become less important for longer time series of speech data. At this stage, I would hesitate to conclude from the results of Fig. 5 that the time series exhibits evidence of low-dimensional chaos. However, Fig. 5d does support the result of Townshend that a moving pole linear filter is unable to extract all the forecastable structure from the original time series. There are also good physical reasons for suspecting that speech time series are produced by filtering non-stationary inputs through a few degree of freedom nonlinear filter, so it is not surprising that substantial nonlinear effects show up in Fig. 5.

## 5.2 Electroencephalograms

There has been much recent interest in applying ideas from dynamical systems to the analysis of EEGs and other physiological time series (see Babloyantz and Destexhe(1986), Mayer-Kress and Layne(1987), Mayer-Kress et al.(1988) and Babloyantz(1989)). The forecasting algorithm of Section 3 was applied to two such time series. Fig. 6a (resp. 6c) shows a phase portrait of an EEG time series from a patient who is resting with eyes closed (resp. the same patient with fluroxene induced anesthesia). Observe that Fig. 6a appears more structured than Fig. 6c. The corresponding $E_m(k)$ curves shown in Figs. 6b and 6d were obtained by taking $N_f = 12000$, $N_t = 400$ and $N_s = 8$, and reveal very little evidence for nonlinear forecastability. Also, observe that linear models suffice to distinguish between these two time series,

and give a difference of about 60% in forecastability. I conclude that there is very little evidence for low-dimensional chaos in these time series. Indeed, these time series come from a many degree of freedom system and, except in certain diseases, are expected to exhibit high-dimensional behaviour. As in the case of fully developed fluid turbulence, this does not necessarily imply that linear models are adequate to describe the time series. Graphical techniques and dimension calculations which have been used for purposes of *signal classification* do seem to reveal interesting nonlinear structure in some EEG time series (see references at the beginning of this subsection).

## 5.3 Measles

Sugihara and May(1990) have applied deterministic forecasting algorithms to a variety of short time series. In the case of the monthly New York measles time series, they claimed to find evidence for deterministic chaos, as well as nonlinearity. Fig. 7a shows a phase portrait of the measles time series, and a strong cyclical dynamics is apparent, somewhat reminiscent of the Rossler attractor. Fig. 7b shows the results of applying the forecasting algorithm of Section 3 with $N_f = N_t = 216$ and $N_s = 1$. The best nonlinear stochastic models outperform linear models by a factor of about 15%. Since the test set consisted of 216 points, a careful hypothesis test would be expected to yield a high statistical significance for evidence of nonlinearity in the time series. This is indeed what was found by Sugihara and May. However, I do not consider this improvement in accuracy to be large enough to conclude that deterministic chaos is relevant to the measles time series. Indeed, recent results of Ellner(1991) indicate that the forecasting results of Sugihara and May are consistent with the more plausible alternative of an underlying stochastic population growth model, rather than deterministic chaos. The nonlinearity in the population growth model considered by Ellner may be removed by a logarithmic transformation. After performing such a transformation on the measles time series and computing $E_m(k)$ curves, I found that most of the evidence for nonlinearity in the original time series disappeared. Since chaos cannot be removed by such a transformation, this supports the alternative of Ellner.

Following the conventions of Sugihara and May, the forecasting algorithm of Section 3 was also applied to the first differenced measles time series, with $N_f = N_t = 216$ and $N_s = 1$, and using the correlation coefficient $\rho_m(k)$ between predicted and actual values of the time series in place of the RMS error $E_m(k)$ of (6). A phase portrait of the first differenced time series is shown in Fig. 7c, and the $\rho_m(k)$ curves are shown in Fig. 7d. Also shown in Fig. 7d is the result of using a step function approximation with $k$ neighbours (obtained by averaging over $k$ neighbours rather than fitting an affine function of the form (5)), and the result of Sugihara and May's simplex algorithm. The best predictions are obtained with stochastic piecewise-linear models, and increase the correlation coefficient from linear models by about 20% (the $E_m(k)$ curves for the first differenced measles data gave similar results, and assuming that the variance of the predicted values is approximately equal to the variance of the actual values, it follows that $E_m^2(k) \approx 2(1 - \rho_m^2(k))$). Also, it can be seen from Fig. 7d

that if step function approximation is used, then much lower values of $k$ give the best predictions. The more sophisticated the function approximation technique used, the more likely it is that stochastic or high-dimensional effects will be identified. This is because a more sophisticated function approximation technique is likely to break down faster as the smoothing parameter is decreased, due to its ability to fit to the noise. Sugihara and May's simplex technique is essentially a kernel density technique, and is more robust at small $k$ than the piecewise-linear technique of Section 3. Although the results shown in Fig. 7d appear to provide stronger evidence for nonlinearity than those shown in Fig. 7b, I still do not consider this improvement in accuracy to be large enough to conclude that the measles time series exhibits low-dimensional chaos.

### 5.4 Sunspots

The annual sunspot time series has been analysed by several authors (see for example Tong(1990) and Weigend et al.(1990)), and strong evidence found for nonlinearity. A phase portrait is shown in Fig. 7e and a strong cyclical dynamics is apparent, corresponding to the famous "11 year cycle". Existing physical models for the interior of the sun have been unable to explain this effect. Fig. 7f shows the results of applying the forecasting algorithm of Section 3, with $N_f = N_t = 143$ and $N_s = 1$. Nonlinear stochastic models are found to give about a 20% improvement over linear models (compare Casdagli et al.(1991b)). However, I do not consider this improvement in accuracy to be large enough to conclude that the sunspot time series exhibits low-dimensional chaos. In this case, a more reasonable alternative is that the underlying nonlinear dynamics is high-dimensional, but with a sufficient amount of coherent structure that can be exploited by a low-dimensional nonlinear stochastic model.

## 6 Relationship to other work

In this section I discuss the relationship of the forecasting algorithm of Section 3 to some existing techniques of nonlinear time series analysis. The emphasis is on *supplementing* the existing techniques, rather than *competing* with them. When only one technique is used to analyse a time series, the results are expected to be at best incomplete, and at worst misleading.

### 6.1 Dimension calculations

The idea of varying neighbourhood sizes is intrinsic to dimension calculations, which attempt to estimate the dimension $D$ of an attractor directly. When *carefully interpreted*, dimension calculations on long time series can reliably distinguish low-dimensional from high-dimensional or stochastic behaviour in a stationary system; see for example Mayer-Kress(1986), Mayer-Kress(1988), Abraham et al.(1989) and Theiler(1991). For statistical issues, see Theiler(1990) and Smith(1991). Note that results from dimension calculations on time series from non-stationary systems must

be interpreted using different concepts, because time series from several stochastic non-stationary systems (for example Brownian motion) are known to have finite dimension (see for example Mandelbrot(1985), Osborne and Provenzale(1989) and Falconer(1990)). As noted in Section 3, forecasting algorithms should also be carefully interpreted when applied to time series from non-stationary systems.

The correlation dimension algorithm of Grassberger and Procaccia(1983) is defined as follows. First define the *correlation integral* $C_m(r)$ by

$$C_m(r) = \#\{(i,j) : \| \underline{x}_i - \underline{x}_j \| < r\}/N^2 \tag{10}$$

where $N$ is the length of the time series, $\underline{x}_i$ and $\underline{x}_j$ are $m$-dimensional delay vectors, $\| \cdot \|$ denotes the maximum norm, and "#" denotes "the number of points in the set". Then if the attractor is of (correlation) dimension $D_2$, and ignoring the effects of noise, it follows that $C_m(r)$ obeys the scaling law

$$C_m(r) \sim r^{\min(m,D_2)} \quad N \to \infty, \quad r \to 0 \tag{11}$$

Fig. 8a shows $C_m(r)$ curves for the case of 6 coupled diodes. Fig. 8b shows corresponding $\nu_m(r)$ curves which estimate the slopes in Fig. 8a and are defined by

$$\nu_m(r) = (\log C_m(r') - \log C_m(r))/(\log r' - \log r) \tag{12}$$

where $r'$ is taken so that $C_m(r')/C_m(r) \approx 5$, following a (personal) recommendation of Theiler. Fig. 8c (resp. 8d) shows $\nu_m(r)$ for 2 (resp. 4) coupled diodes. I interpret the results shown in Fig. 8 as follows. Fig. 8c (resp. 8d) indicates convergence to $D_2 \approx 2.5$ (resp. 3.5), according to the scaling law (11). Fig. 8b is more open to interpretation, but perhaps indicates $D_2 \approx 4.5$. The forecasting results of Fig. 2 can now be re-interpreted. I conjecture that the reason for the unimpressive nonlinear forecasts in Figs. 2d and 2f is not that the dimension of the underlying attractor is high, but that the dynamics induced on the reconstructed attractor is complicated due to the low sampling rate. Of course, I might have misinterpreted the results of the dimension calculations. As suggested in Section 4.1, information in the underlying ODEs believed to accurately model the coupled diode system might be useful in supporting or refuting the above conjecture.

Figs. 9a and 9b show $\nu_m(r)$ curves for the fluid turbulence time series of Fig. 3. The delay times $\tau$ were the same as those used for Figs. 3a and 3c. Also, no distances were computed between delay vectors $\underline{x}_i$ and $\underline{x}_j$ with $|i-j| < \tau$ in evaluating the correlation integral (10), to avoid spurious low-dimensional effects (for example see Theiler(1986)). Fig. 9a indicates low-dimensional behaviour with $D_2 \approx 3.0$. Fig. 9b indicates high-dimensional behaviour. The forecasting results of Figs. 3b and 3d provide useful extra information to support the results of these dimension calculations. In the case of the highly turbulent time series, a small amount of extra structure was picked up by the forecasting algorithm which was not apparent in the dimension calculation. Figs. 9c and 9d show $\nu_m(r)$ curves for the flame dynamics time series of Figs. 4a and 4c, with the same conventions that were used for the fluid turbulence

time series. The dimension calculations have not converged to a low dimension, but when compared to Fig. 9b do not appear to be consistent with high-dimensional behaviour either. This conclusion is supported by the forecasting results of Figs. 4b and 4d.

Dimension calculations have been applied to short time series with some controversy (for example see Grassberger(1986) and Ruelle(1990)). However, Brock et al.(1986) have developed a statistical test (the "BDS" test), based on dimension calculations which can be used reliably on short time series. The results are used to *reject* various null hypotheses about the time series, rather than *accept* evidence for chaos. See also Theiler et al.(1991). In this paper I have avoided hypothesis testing in favour of a more exploratory approach. However, it is essential to perform hypothesis tests in the case of short time series to obtain statistically convincing results. Forecasting algorithms have been applied to economic time series by Hseih(1991) to decide between various alternative hypotheses when the BDS test rejects a null hypothesis. Lee et al.(1989) have compared the power of the BDS test against other statistical tests for nonlinearity. Some of these results show that the BDS test may identify nonlinear structures, such as conditional heteroskedasticity in economic time series, which are of a form which cannot be detected by the forecasting strategy of Section 3. It would therefore be interesting to apply the BDS test to the case of the long, high-dimensional time series of Figs. 3c, 4e, 6a and 6c, to test the adequacy of linear stochastic models for these systems.

## 6.2 Lyapunov exponents and multi-step forecasting

Algorithms for estimating Lyapunov exponents in deterministic systems have been developed by Sano and Sawada(1985), Wolf et al.(1985) and Eckmann et al.(1986). The algorithms estimate the rate of divergence of trajectories in deterministic approximations to the map $f$ in (4). Algorithms for estimating Lyapunov exponents in nonlinear stochastic systems have been developed by McCaffrey et al.(1991). A related approach is to investigate the behaviour of the average forecasting errors $E_m(T)$ as the forecasting time $T$ increases (Farmer and Sidorowich(1987)). Fig. 10 shows $E_m(T)$ curves for fluid turbulence and flame dynamics time series. The map $f^T$ in (4) was approximated directly, rather than iteratively. Optimal values of $m$ and $k$ were taken according to the results of Figs. 3 and 4, and found to be robust as $T$ is varied. Also shown in Fig. 10 are results obtained with linear models.

Fig. 10a shows a slow rate of decay of predictive accuracy for the nonlinear deterministic model, and provides strong evidence for a weakly chaotic system (compare Farmer and Sidorowich(1987)). The results in Fig. 10b might be confused with evidence for low-dimensional chaos, since short-term forecasts decay rapidly. However, close to optimal results are obtained with a linear model, and are consistent with high-dimensional stochastic behaviour. The results in Fig. 10c might be confused with evidence for a weakly chaotic system. However, since forecasts can be made accurately for long times $T$ using both linear and nonlinear models, these results are strong evidence for a non-chaotic deterministic system, perturbed by noise. The

results in Fig. 10d are more difficult to interpret. The rapid decay in the accuracy of forecasts indicates either a low-dimensional chaotic system or a high-dimensional stochastic system. The main new insight provided by Fig. 10 is the distinction drawn between the multi-step forecastability of the flame dynamics time series of Fig. 4a and 4c.

Sugihara and May(1990) have investigated decays of nonlinear forecasts with time on short time series, and have been able to convincingly distinguish chaotic behaviour from limit cycles with additive noise. They concluded that the New York measles time series is chaotic. However, the conclusion should have been that the measles time series is not adequately described by a limit cycle with additive noise. As argued by Ellner(1991), a more plausible alternative to chaos in the measles time series is that it is described by a limit cycle with *multiplicative* noise. Hopefully, the remarks of this subsection emphasise the difficulties in inferring low-dimensional chaos from estimates of the rate of divergence of trajectories alone.

## 6.3 Other forecasting algorithms

Piecewise linear approximation is one of several multivariate function approximation techniques which can be used for approximating the function $f$ in (4). Other, more computationally intensive, techniques have been shown to give more accurate forecasts on numerical time series and some observational time series. Examples include neural nets (Lapedes and Farber(1988), Weigend et al.(1990,1991)), radial basis functions (Casdagli(1989)) and locally weighted regression (Stokbro and Umberger(1991), Mead et al.(1991)). "Smoothing" parameters can be varied in all of these techniques, and the results of this paper suggest that a systematic variation of such parameters from deterministic to stochastic extremes could usefully be applied to time series analysis. With the current availability of computer resources, I believe such function approximation techniques provide promising alternatives to the more traditional nonparametric regression techniques of statistics based on kernel density estimation (Silverman(1986)).

In this paper the underlying state space for nonlinear modeling was reconstructed using delay vectors (3). Broomhead and King(1986) proposed a reconstruction technique based on PCA (Principal Components Analysis), which attempts to improve on delay coordinate reconstructions. Fraser(1989) has shown that this technique does not always work well. Theoretical investigations into general reconstruction techniques have been made by Casdagli et al.(1991a) and a local version of PCA proposed. The Broomhead and King technique has been applied by Townshend(1990) to the modeling of speech time series, and a local version of PCA has been applied by Hunter(1991) to the modeling of time series from mechanical vibrations and ice ages. The results presented in this paper could be used as a benchmark to assess the practical advantages of these more general reconstruction techniques for forecasting, compared to delay coordinate reconstructions.

Forecasting algorithms have been applied to the problem of noise reduction by several authors; for a review see Grassberger et al.(1991). The forecasting algorithm

of Section 3 can be used for noise reduction by "two-sided forecasting", with $T = -(m-1)\tau/2$ (and $m$ odd), and investigating the $E_m(k)$ curves. Finally, Packard(1990) has developed a genetic algorithm for searching for special regions of forecastability in state space reconstructions, which is of potential use in high-dimensional systems. In the case of some economic time series, it has been found that if attention is focused on special segments of the time series, then there can be a small, but statistically significant, amount of nonlinear forecastability (LeBaron(1991), Weigend et al.(1991)). The global average (6) used in this paper is unlikely to be sensitive to such effects, but might perhaps be modified.

# 7 Conclusions

The main results of this paper are as follows. Low-dimensional chaos was identified by obtaining accurate short-term forecasts for time series from two coupled diodes and a fluid near the transition to turbulence. Nonlinearity, but not low-dimensional chaos, was identified in time series from four and six coupled diodes, and time series from flame dynamics near the transition to turbulence. Nonlinearity was also identified in time series from speech, measles and sunspots, confirming the results of others. Linear forecasts were found to be close to optimal for time series from fluids and flames far from the transition to turbulence, and for EEG time series. Several of these conclusions could have been guessed at by inspection of the phase portraits, and the use of graphical techniques should not be underestimated. The above results are mostly consistent with prior expectations from dynamical systems theory. Although the phenomenon of chaos may be ubiquitous, low-dimensional chaos is only expected to be relevant to few degree of freedom systems (for example the coupled diodes system in this paper), or many degree of freedom systems (for example the fluid and flame systems in this paper) near the transition to chaos.

The results from the forecasting algorithm used in this paper were found to be a useful supplement to, and often less ambiguous than, results produced by dimension calculations. Unless carefully interpreted, dimension calculations often indicate low-dimensional behaviour when it is not there. By contrast, the results from the forecasting algorithm presented in this paper only rarely indicate low-dimensional behaviour. However, in certain cases, dimension calculations can pick up interesting structure unavailable to forecasting algorithms. For example, in the case of the four coupled diodes time series, strong evidence for low-dimensional chaos was suggested by dimension calculations even though the short-term forecasts for this time series were not particularly accurate. Nonlinear dynamical systems are capable of such a wide variety of behaviour that the use of a single technique of time series analysis should not be relied upon too heavily.

A number of questions were raised in the course of these investigations as follows. First, the conclusions reached on the four and six coupled diodes time series would be clarified by attempting to obtain short-term forecasts using the underlying deterministic equations believed to accurately model the system. In other cases where

high levels of nonlinearity have been identified by "statistical" forecasting techniques, this approach might aid in the development or improvement of more traditional "first principles" models. Second, the results obtained for the flame dynamics time series near the transition to turbulence remain ambiguous. It would be desirable to analyse spatio-temporal time series data in order to obtain clearer conclusions about the underlying dynamics. Third, it would be interesting to use the BDS test in the case of the time series from fluids and flames far from the transition to turbulence, and for the EEG time series, to test for the adequacy of linear models. It might be expected that the collective phenomena observed by physicists in idealised models of many degree of freedom systems should also occur in these time series, and it would be desirable to develop tests which can detect such phenomena in a time series. Finally, if the objective is to obtain optimal forecasting accuracy, a wide range of nonlinear forecasting strategies remain to be explored.

# Acknowledgements

I have had useful conversations with Howell Tong and Clive Granger. I would like to thank Paul Linsay for supplying the diode data, Dan Lathrop, J.Fineberg and H.Swinney for supplying the Taylor-Couette data, Mike Gorman, M.el-Hamdi and Kay Robbins for supplying the flame data, Brent Townshend for supplying the speech data, Gottfried Mayer-Kress for suppling the EEG data, George Sugihara for supplying the measles data and James Theiler for supplying the dimension calculation code. I would like to thank Richard Smith and James Theiler for a critical reading of the manuscript. I am grateful to partial support from the National Institute of Mental Health under grant 1-R01-MH47184-01.

# References

N.B.Abraham, A.M.Albano, A.Passamante, and P.E.Rapp (1989) *Measures of Complexity and Chaos*, Volume 208 of *NATO Advanced Science Institute Series.* Plenum, New York

A.Babloyantz (1989) Some remarks on nonlinear data analysis of physiological time series. In: *Measures of Complexity and Chaos*, Eds. Abraham et al., Volume 208 of *NATO Advanced Science Institute Series.* Plenum, New York

A.Babloyantz and A.Destexhe (1986) Low-dimensional chaos in an instance of epilepsy. *Proceedings of the National Academy of Science*, **53**, 3513-3517

R.Badii and A.Politi (1985) Statistical description of chaotic attractors: The dimension function. *J. Stat. Phys.*, **40**, 725

P.Bak, C.Tang, and K.Wiesenfeld (1987) Self-organized criticality: An explanation of $1/f$ noise. *Phys. Rev. Lett.* **59**, 381

J.L.Breeden and A.Hubler (1990) Reconstructing Equations of Motion from experimental data with hidden variables. *Phys. Rev. A*, **42**, 5817-5826

K.Briggs (1990) Improved methods for the analysis of chaotic time series. Mathematics Research Paper 90-2, La Trobe University, Melbourne, Australia

W.Brock, W.Dechert and J.Scheinkman (1987) A test for Independence based on the correlation dimension. University of Wisconsin-Madison SSRI Paper 8702

D.S.Broomhead and G.P.King (1986) Extracting qualitative dynamics from experimental data. *Physica D*, **20**, 217

D.S.Broomhead, R.Indik, A.C.Newell and D.A.Rand (1990) Local adaptive Galerkin bases for large dimensional dynamical systems. Submitted to *Nonlinearity*

S.D.Brorson, D.Dewey and P.S.Linsay (1983) Self-replicating attractor of a driven semiconductor oscillator. *Phys. Rev. A*, **28**, 1201-1203

M.Casdagli (1989) Nonlinear prediction of chaotic time series. *Physica D*, **35**, 335

M.Casdagli (1991) Nonlinear forecasting, chaos and statistics. In: *Modeling complex phenomena*, Eds. L.Lam and V.Naroditsky. Springer-Verlag, New York, to appear.

M.Casdagli, S.Eubank, J.D.Farmer, and J.Gibson (1991a) State space reconstruction in the presence of noise. *Physica D*, to appear.

M.Casdagli, D.DesJardins, S.Eubank, J.D.Farmer, J.Gibson, N.Hunter and J.Theiler (1991b) Nonlinear modeling of chaotic time series: theory and applications. In: *EPRI workshop on Applications of chaos*, Eds. J.Kim and J.Stringer, to appear.

J.P.Crutchfield and K.Kaneko (1988) Phenomenology of spatio-temporal chaos. In: *Directions in Chaos, Vol. I.* Ed. Hao Bai-Lin, World Scientific, Singapore

J.P.Crutchfield and B.S.McNamara (1987) Equations of motion from a data series. *Complex Systems*, **1**, 417-452

C.R.Doering, J.D.Gibbon, D.D.Holm and B.Nicolaenko (1988) Low dimensional behaviour in the complex Ginzburg-Landau equation. *Nonlinearity* **1**, 279-309

J.P.Eckmann, S.Kamporst, D.Ruelle and S.Ciliberto (1986) Lyapunov exponents from a time series. *Phys. Rev.* A **34**, 4971

J.P.Eckmann and D.Ruelle (1985) Ergodic Theory of Chaos and Strange Attractors. *Rev. Mod. Phys*, **57**, 617

J.P.Eckmann, S.Kamporst and D.Ruelle (1987) Recurrence Plots of Dynamical Systems. *Europhysics Letters*, 4, 973-977

S.Ellner (1991) Detecting low-dimensional chaos in population dynamics data: a critical review. In: *"Does chaos exist in ecological systems"*, Eds. J.Logan and F.Hain, U. Virginia Press, to appear.

S.Eubank and J.D.Farmer (1989) An introduction to chaos and randomness. In: *1989 Lectures in complex systems Vol II*, Ed. E.Jen, Addison-Wesley, Redwood City, CA

K.Falconer (1990) *Fractal Geometry.* Wiley, Chichester, UK

J.D.Farmer and J.J.Sidorowich (1987) Predicting chaotic time series. *Phys. Rev. Lett.*, **59**(8), 845-848

J.D.Farmer and J.J.Sidorowich (1988) Exploiting chaos to predict the future and reduce noise. In: *Evolution, Learning and Cognition*, Ed. Y.C. Lee, World Scientific, Singapore

A.M.Fraser (1989) Reconstructing attractors from scalar time series: A comparison of singular system and redundancy criteria. *Physica D*, **34**, 391-404

J.Geweke (1989) Inference and forecasting for chaotic nonlinear time series. Technical report, Duke University

L.Glass and M.C.Mackey (1988) *From clocks to chaos: the rhythms of life.* Princeton University Press, Princeton

M.Gorman and K.Robbins (1991) Real-time identification of flame dynamics. In: *EPRI workshop on Applications of chaos*, Eds. J.Kim and J.Stringer, to appear.

P.Grassberger (1986) Do climatic attractors exist? *Nature*, **323**, 609

P.Grassberger and I.Procaccia (1983) Characterization of strange attractors. *Phys. Rev. Lett.*, **50**, 346

P.Grassberger, T.Schreiber and C.Schaffrath (1991) Nonlinear time sequence analysis. Preprint, Univ. of Wuppertal

G.Gunaratne, P.Linsay and M.Vinson (1989) Chaos beyond onset: A comparison of theory and experiment. *Phys. Rev. Lett.*, **63**, 1

H.Haucke and R.Ecke (1987) Mode locking and chaos in Rayleigh-Benard convection. *Physica D* **25**, 307

D.Hseih (1991) Chaos and Nonlinear dynamics: Application to Financial Markets. *J. Finance*, to appear.

N.F.Hunter (1991) Application of nonlinear time series models to driven systems. In: *Nonlinear Modeling and Forecasting*, Eds. M.Casdagli and S.Eubank, Addison-Wesley, Redwood City, CA, to appear.

K.Ikeda (1979) Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. *Opt. Commun.* **30**, 257

J.D.Keeler and J.D.Farmer (1986) Robust space-time intermittency and $1/f$ noise. *Physica D* **23**, 413

A.Kumar and S.Mullick (1990) Attractor dimension, entropy and modeling of phoneme time series. Preprint, Indian Institute of Technology

A.S.Lapedes and R.Farber (1988) How neural nets work. In: *Evolution, Learning and Cognition*, Ed. Y.C. Lee, World Scientific, Singapore

B.LeBaron (1991) Nonlinear Forecasts for the S&P Stock Index. In: *Nonlinear Modeling and Forecasting*, Eds. M.Casdagli and S.Eubank, Addison-Wesley,

Redwood City, CA, to appear.

T.Lee, H.White and C.Granger (1989) Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. Preprint, University of California, San Diego

B.B.Mandelbrot (1985) *The fractal geometry of nature.* Freeman, San Fransisco.

M.C.Mackey and L.Glass (1977) Oscillation and Chaos in physiological control systems. *Science*, **197**, 287

G.Mayer-Kress (1986) *Dimensions and Entropies in Chaotic Systems* , Vol. 32, *Springer Series in Synergetics*, Springer-Verlag, New York

G.Mayer-Kress (1988) Applications of dimension algorithms to experimental chaos. In: *Directions in Chaos, Vol. I.* Ed. Hao Bai-Lin, World Scientific, Singapore

G.Mayer-Kress and S.P.Layne (1987) Dimensionality of the Human Electroencephalogram, *Annals of the New York Academy of Sciences,* **504**, 62-86

G.Mayer-Kress, F.Yates, L.Benton, M.Keidel, W.Tirsch, S.Poppl and K.Geist (1988) Dimensional analysis of nonlinear oscillations in brain, heart and muscle. In: *Nonlinearity in biology and medicine*, Eds. A.Perelson, B.Goldstein, M.Dembo and J.Jacquez, Elsevier, Amsterdam

W.C.Mead, R.D.Jones, Y.C.Lee, C.W.Barnes, G.W.Flake, L.A.Lee and M.K.O'Rourke (1991) Prediction of chaotic time series using CNLS net– Example: The Mackey-Glass equation. In: *Nonlinear Modeling and Forecasting*, Eds. M.Casdagli and S.Eubank, Addison-Wesley, Redwood City, CA, to appear.

D.McCaffrey, S.Ellner, A.Gallant and D.Nychka (1991) Estimating Lyapunov exponents with nonparametric regression. Preprint, North Carolina State University.

A.Osborne and A.Provenzale (1989) Finite correlation dimension for stochastic systems with power-law spectra. *Physica D*, **35**, 357-381

N.H.Packard (1990) A genetic learning algorithm for the analysis of complex data. *Complex Systems* **4**, 543-572

N.H.Packard, J.P.Crutchfield, J.D.Farmer, and R.S.Shaw (1980) Geometry from a time series. *Phys. Rev. Lett.*, **45**, 712-716

W.H.Press, B.P.Flannery, S.A.Teukolsky and W.T.Vetterling (1988) *Numerical recipes in C.* Cambridge University Press, Cambridge

D.Ruelle (1990) Deterministic chaos: the science and the fiction. *Proc. R. Soc. London* A **427**, 241

D.Ruelle and F.Takens (1971) On the nature of turbulence. *Comm. Math. Phys.*, **20** 167-192

M.Sano and Y.Sawada (1985) Measurement of the Lyapunov spectrum from chaotic time series. *Phys. Rev. Lett.*, **55**, 1082

T.Sauer, J.A.Yorke and M.Casdagli (1990) Embedology. *J. Stat. Phys.*, to appear.

R.Shaw (1984) *The dripping faucet as a model dynamical system.* Aerial Press, Santa Cruz, CA 95060

B.W.Silverman (1986) *Kernel density estimation techniques for statistics and data analysis.* Chapman Hall, London

R.L.Smith (1991) Optimal estimation of fractal dimension. In: *Nonlinear Modeling and Forecasting*, Eds. M.Casdagli and S.Eubank, Addison-Wesley, Redwood City, CA, to appear.

K.Stokbro and D.Umberger (1991) Forecasting with weighted maps. In: *Nonlinear Modeling and Forecasting*, Eds. M.Casdagli and S.Eubank, Addison-Wesley, Redwood City, CA, to appear.

G.Sugihara and R.May (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in a data series. *Nature*, **344**, 734-741

F.Takens (1981) Detecting strange attractors in fluid turbulence. In: *Dynamical Systems and Turbulence*, Eds. D.Rand and L.-S.Young, Springer-Verlag, New York

J.Theiler (1986) Spurious dimension from correlation algorithms applied to limited time series data. *Phys. Rev. A*, **34**, 2427-2432

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

J.Theiler (1991) Some comments on the correlation dimension of $1/f^{\alpha}$ noise. *Phys. Lett. A*, to appear

J.Theiler, B.Galdrikian, A.Longtin, S.Eubank and J.D.Farmer (1991) Using surrogate data to detect nonlinearity in time series. In: *Nonlinear Modeling and Forecasting*, Eds. M.Casdagli and S.Eubank, Addison-Wesley, Redwood City, CA, to appear.

H.Tong (1990) *Nonlinear Time Series Analysis: A Dynamical Systems Approach.* Oxford University Press, Oxford

H.Tong and K.S.Lim (1980) Threshold autoregression, limit cycles and cyclical data. *J. Roy. Stat. Soc. B*, **42**, 245-292

B.Townshend (1990) Nonlinear prediction of speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, to appear.

A.Weigend, B.Huberman and D.Rumelhart (1990) Predicting the future: a connectionist approach. *International Journal of Neural Systems*, **1**, 193-209

A.Weigend, B.Huberman and D.Rumelhart (1991) Predicting sunspots and exchange rates with connectionist networks. In: *Nonlinear Modeling and Forecasting*, Eds. M.Casdagli and S.Eubank, Addison-Wesley, Redwood City, CA, to appear.

A.Wolf, J.Swift, H.Swinney and J.Vastano (1985) Determining Lyapunov exponents from a time series. *Physica D*, **16**, 285

# Figure Captions

Figure 1: *Forecasting errors for numerical data.* (a) Mackey-Glass equation. $E_m(k)$ for $m = 4, 5, 6, 7$. In this and subsequent figures, the embedding dimension $m$ of the $E_m(k)$ curves can be identified according to how far to the left of the figure the $E_m(k)$ curve starts. The most accurate forecasts are obtained with $m = 6$ or 7. (b) Ikeda map. $E_m(k)$ for $m = 5$ at noise levels 0%, 2%, 20% and 100%. Higher curves correspond to higher noise levels. The dashed lines illustrate the scaling law (9), and the horizontal line "1" corresponds to a *fundamental limit to predictability* due to noise amplification of the 2% noise level; see Section 3.3.

Figure 2: *Forecasting errors for n coupled diodes data.* (a),(c),(e) Phase portraits for the cases $n = 2, 4, 6$ with $\tau = 1$. In this and subsequent figures, not all of the data is shown for long time series. (b),(d),(f) $E_m(k)$ for $m = 3, 4, 5, 6, 7$ and $\tau = T = 1$. The dotted curves in (b) correspond to a geometric mean error. In (b) (resp. (d), (f)) the most accurate forecasts are obtained with $m = 5$ (resp. 7, 6).

Figure 3: *Forecasting errors for fluid turbulence data.* (a) (resp. (c)) Phase portrait for weak (resp. strong) turbulence, with $\tau = 20$ (resp. $\tau = 30$). (b) (resp. (d)) $E_m(k)$ for $m = 4, 8, 12, 16, 20$, $T = 20$ (resp. $T = 10$) and $\tau = 5$ (resp. $\tau = 2$). The most accurate forecasts are obtained with $m = 20$ (resp. 8).

Figure 4: *Forecasting errors for flame dynamics data.* (a) (resp. (c), (e)) Phase portraits for what is believed to be a non-chaotic (resp. low-dimensional chaotic, high-dimensional chaotic) system, with $\tau = 40$ (resp. $\tau = 10, 100$). (b) (resp. (d),(f)) $E_m(k)$ for $m = 4, 8, 12, 16, 20$, $T = 40$ (resp. $T = 10, 50$) and $\tau = 10$ (resp. $\tau = 2, 10$). The most accurate forecasts are obtained with $m = 20$ (resp. 20, 8).

Figure 5: *Forecasting errors for speech data.* (a) (resp. (c)) Phase portrait of original (resp. filtered) time series with $\tau = 2$ (resp. $\tau = 1$). (b),(d) $E_m(k)$ for $m = 4, 8, 12, 16, 20$ and $T = \tau = 1$. The most accurate forecasts are obtained with $m = 16$.

Figure 6: *Forecasting errors for EEG data.* (a),(c) Phase portraits with $\tau = 10$. (b),(d) $E_m(k)$ for $m = 4, 8, 12, 16, 20$, $T = 10$ and $\tau = 2$.

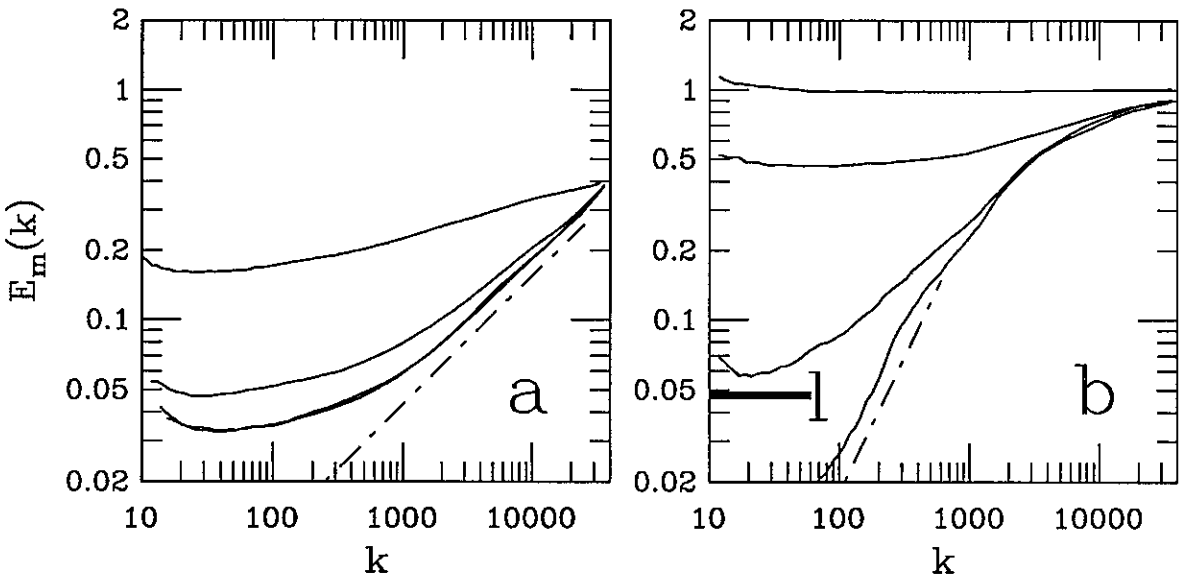
Figure 7: *Forecasting errors for measles and sunspot data.* (a) (resp. (c), (e)) Phase portrait of measles (resp. first differenced measles, sunspots) time series with $\tau = 2$ months (resp. $\tau = 1$ month, $\tau = 2$ years). (b) $E_m(k)$ for measles with $m = 2, 3, 4, 5, 6$ and $T = \tau = 1$. (d) Correlation coefficient $\rho_m(k)$ for first differenced measles with $m = 2, 3, 4, 5, 6$ and $T = \tau = 1$. The dashed line denotes results from a step function approximation technique with $m = 6$. The "X" denotes a result from Sugihara and May's simplex algorithm, with $k = m + 1 = 7$ neighbours. (f) $E_m(k)$ for sunspots with $m = 3, 4, 5, 6$ and $T = \tau = 1$.

Figure 8: *Dimension calculations for n coupled diodes data* for $m = 2, 3, .., 8$ and $\tau = 1$ (a) $C_m(r)$ curves for $n = 6$. Higher curves correspond to lower $m$. (b),(c),(d) $\nu_m(r)$ curves for $n = 6, 2, 4$. Higher curves correspond to higher $m$ until a saturation is reached.

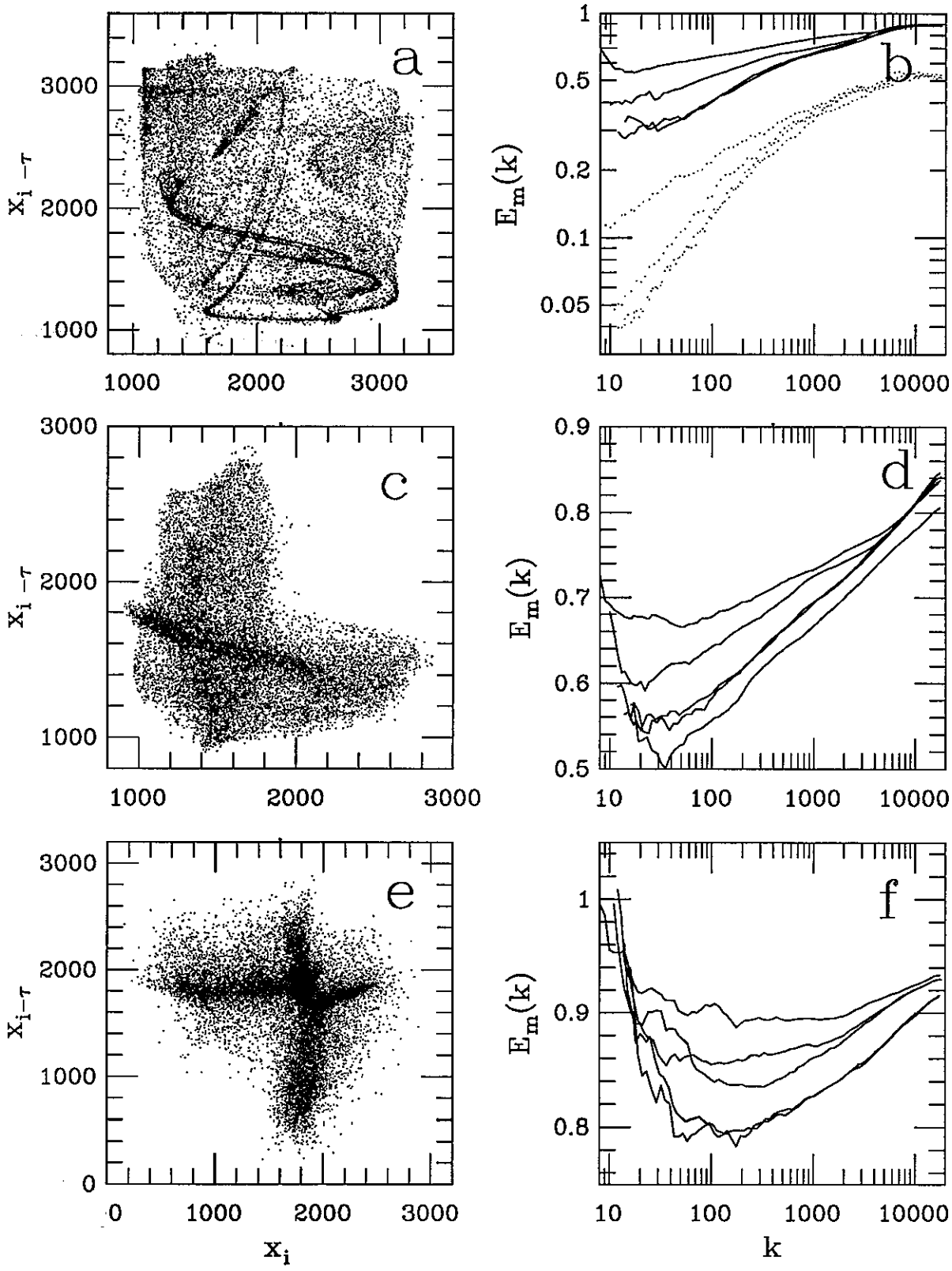
Figure 9: *Dimension calculations for fluid turbulence and flame dynamics data* for $m = 2, 3, .., 10$. Higher curves correspond to higher $m$ until a saturation is reached. (a) (resp. (b)) $\nu_m(r)$ curves for the case of weak (resp. strong) fluid turbulence with $\tau = 20$ (resp. $\tau = 30$). (c) (resp. (d)) $\nu_m(r)$ curves for what is believed to be a non-chaotic (resp. low-dimensional chaotic) flame dynamics time series, with $\tau = 40$ (resp. $\tau = 10$).

Figure 10: *Multi-step forecasting errors for fluid turbulence and flame dynamics data.* Dashed curves are for linear models, solid curves for nonlinear models. Also shown are segments of time series data on the same time scale. (a) (resp. (b)) $E_m(T)$ curves for the case of weak (resp. strong) fluid turbulence with $m, \tau, k = 20, 5, 42$ (resp. $m, \tau, k = 8, 2, 20000$). (c) (resp. (d)) $E_m(T)$ curves for what is believed to be a non-chaotic (resp. low-dimensional chaotic) flame dynamics time series, with $m, \tau, k = 20, 10, 300$ (resp. $m, \tau, k = 20, 2, 200$).

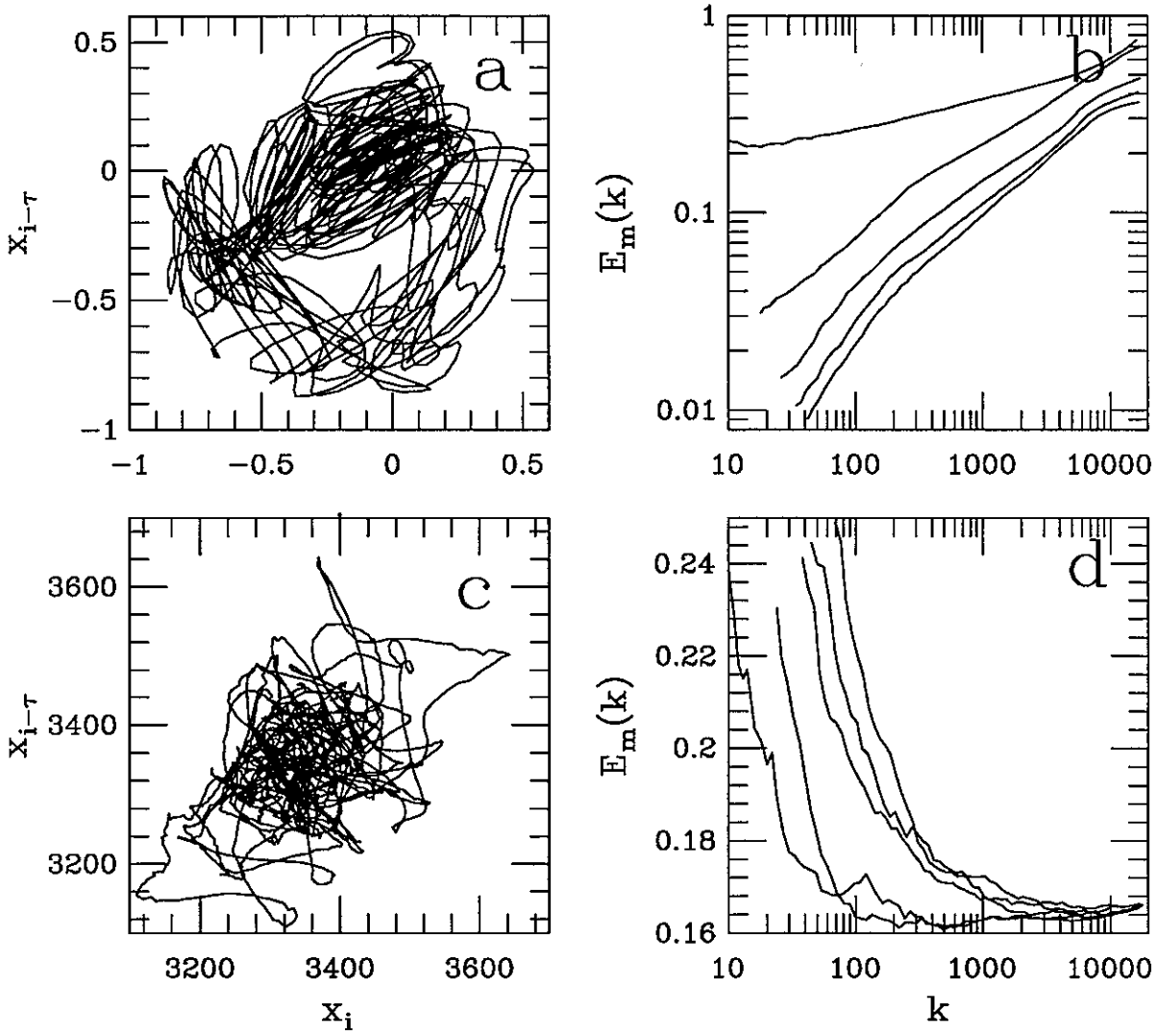
# Figure 1



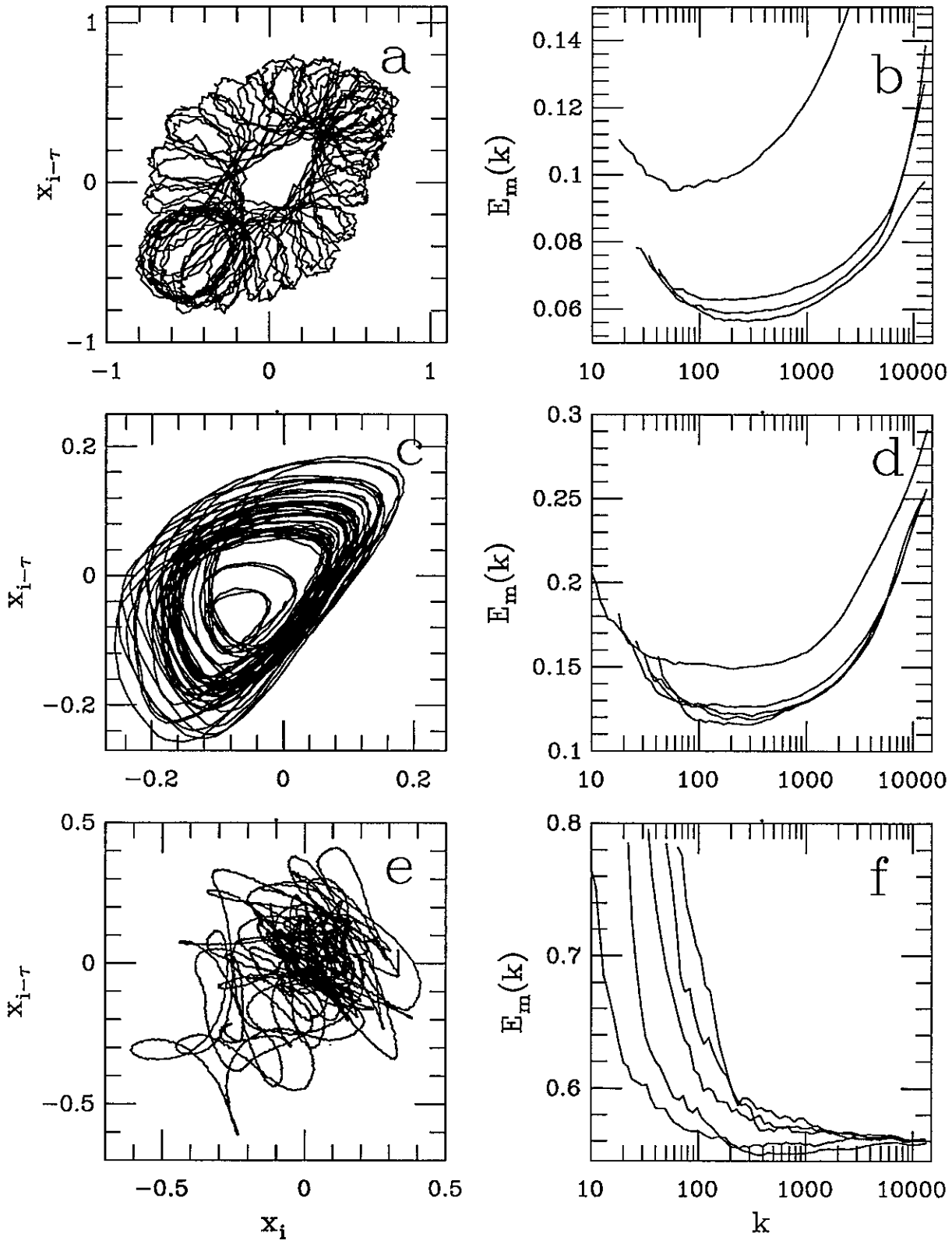
# Figure 2


a
b
c
d
e
f
$x_{i-\tau}$
$E_m(k)$
$x_i$
k
3000
2000
1000
0
1
0.5
0.2
0.1
0.05
0.9
0.8
0.7
0.6
10
100
1000
10000

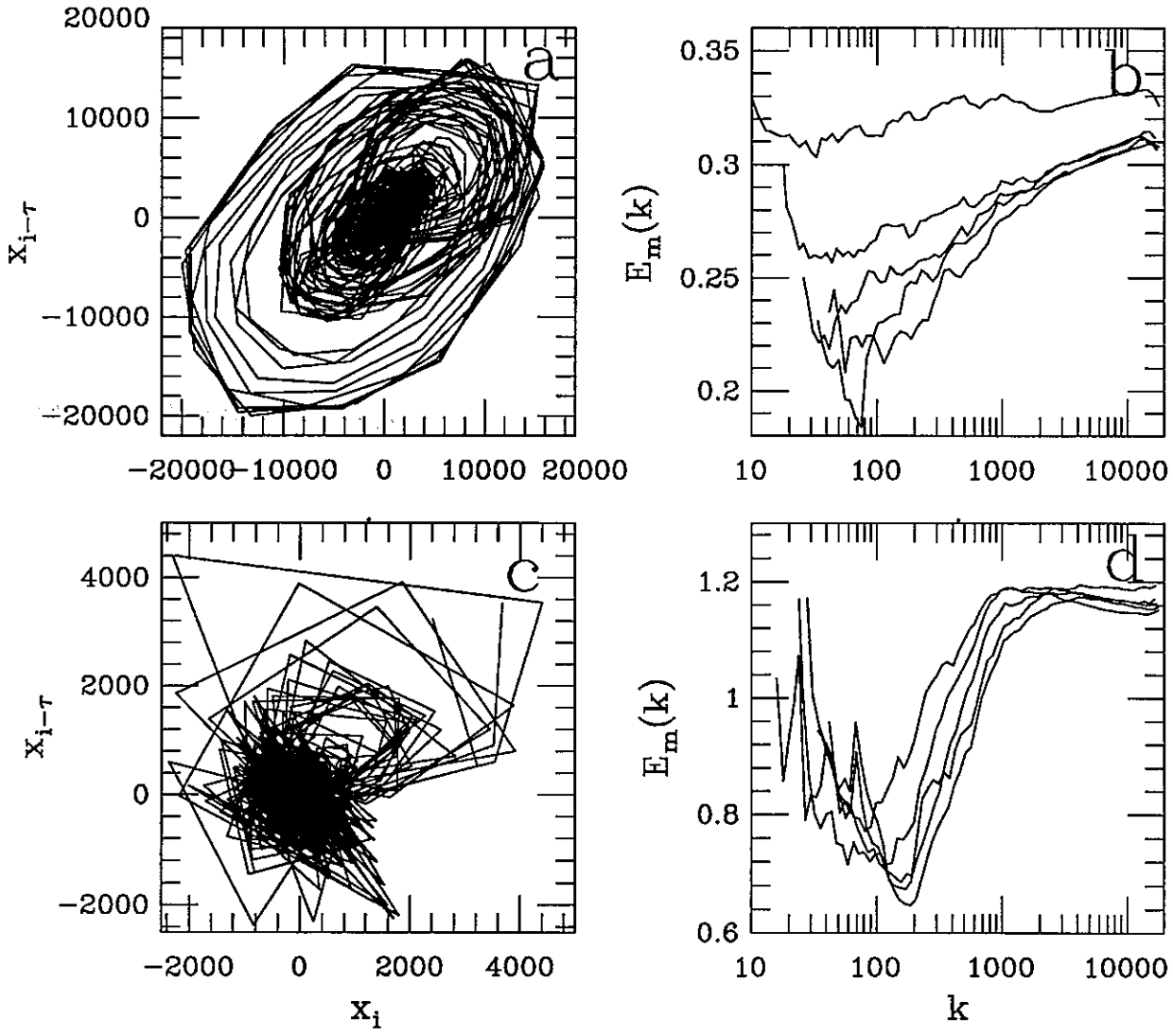
# Figure 3


a
b
c
d
0.5
0
−0.5
−1
$x_{i-\tau}$
−1
−0.5
0
0.5
1
0.1
0.01
$E_m(k)$
10
100
1000
10000
3600
3400
3200
$x_{i-\tau}$
3200
3400
3600
$x_i$
0.24
0.22
0.2
0.18
0.16
$E_m(k)$
10
100
1000
10000
k

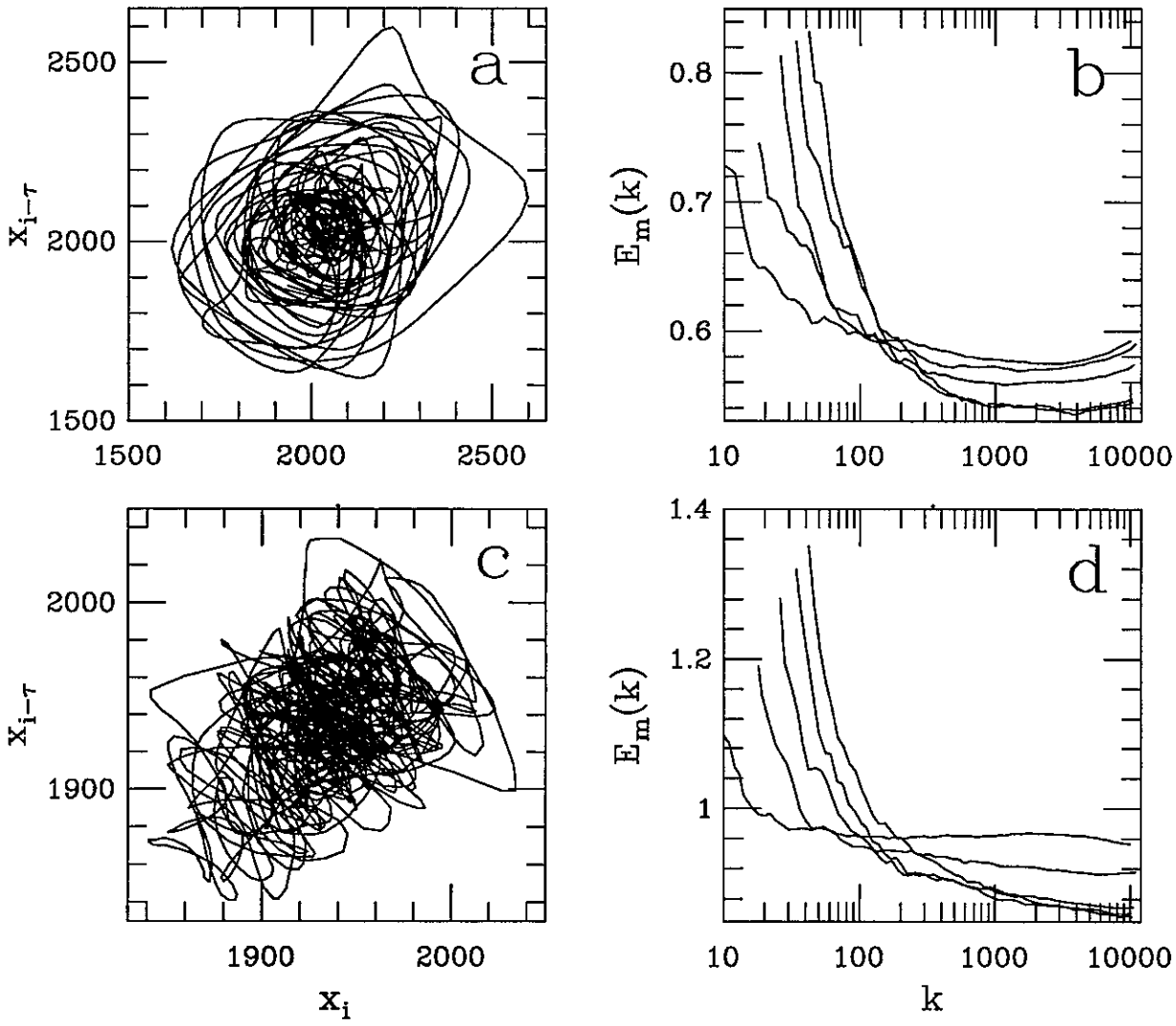
# Figure 4


a
b
c
d
e
f
$x_{i-\tau}$
$x_i$
$E_m(k)$
k
1
0
−1
0.14
0.12
0.1
0.08
0.06
10
100
1000
10000
0.2
−0.2
0.3
0.25
0.2
0.15
0.1
0.5
−0.5
0.8
0.7
0.6

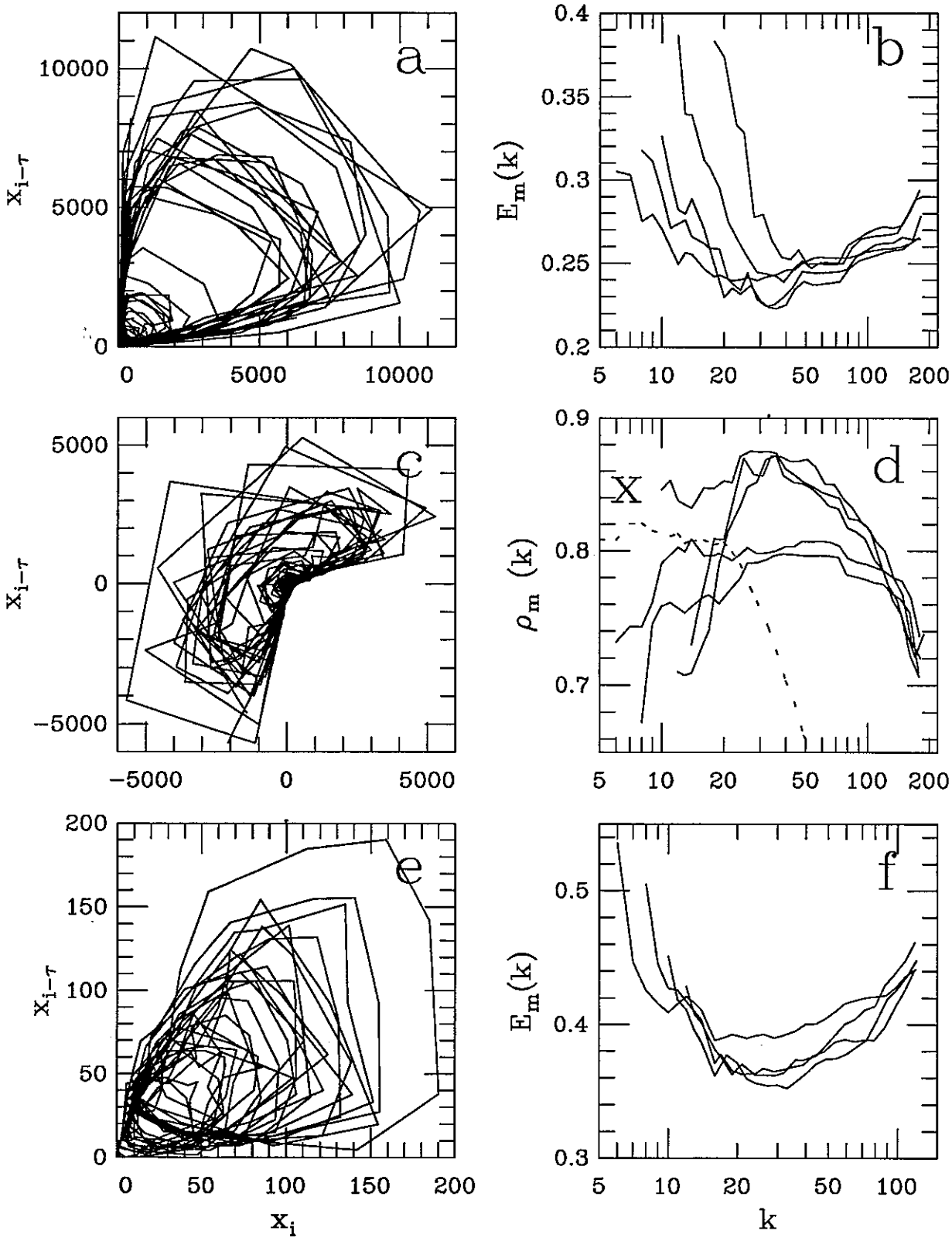
# Figure 5


a
b
c
d
20000
10000
0
−10000
−20000
$x_{i-\tau}$
−20000 −10000 0 10000 20000
0.35
0.3
0.25
0.2
$E_m(k)$
10 100 1000 10000
4000
2000
0
−2000
−2000 0 2000 4000
$x_i$
1.2
1
0.8
0.6
k

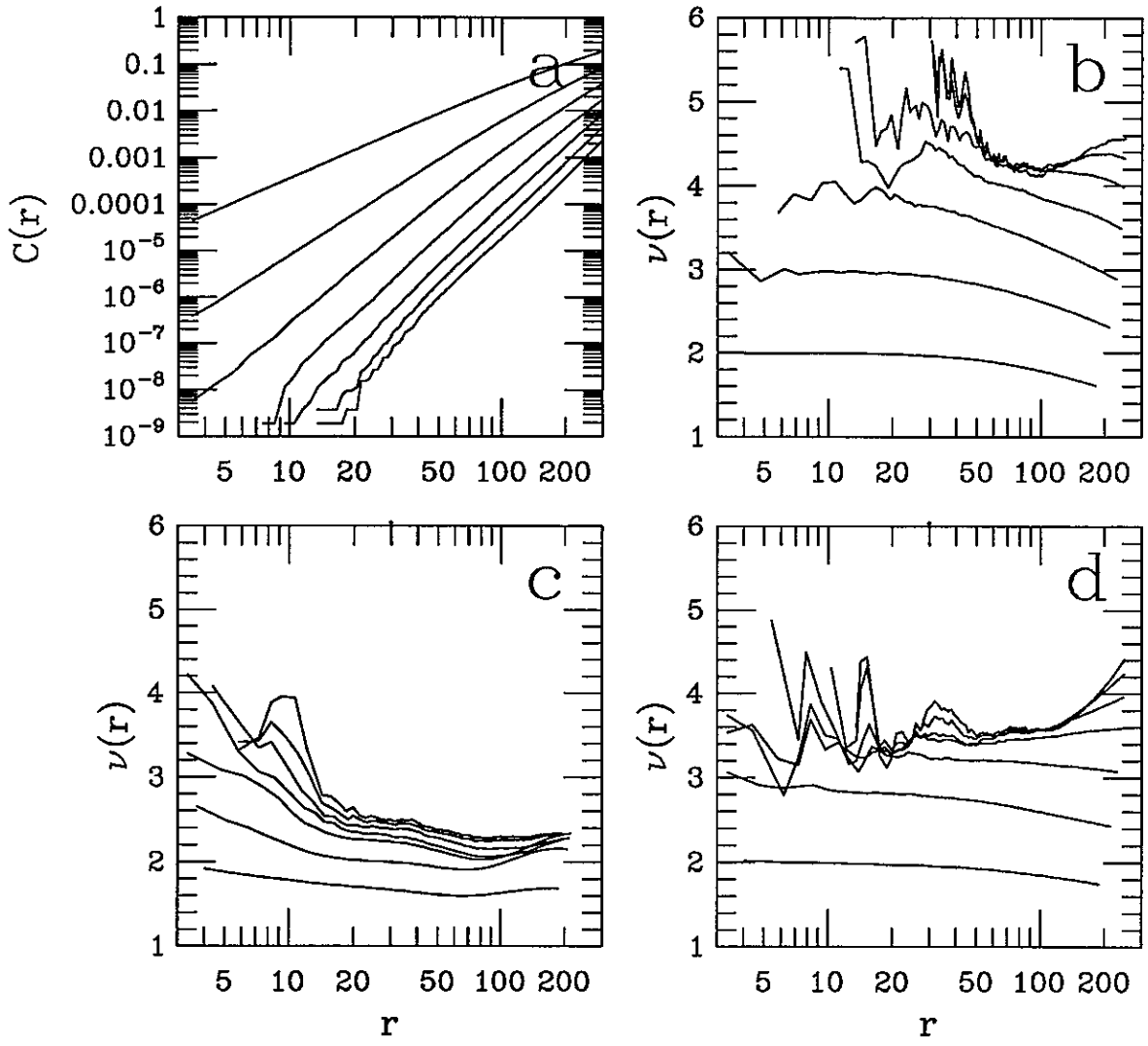
# Figure 6


a
b
c
d
2500
2000
1500
$x_{i-\tau}$
$x_i$
$E_m(k)$
k
0.8
0.7
0.6
1.4
1.2
1
1900
10
100
1000
10000

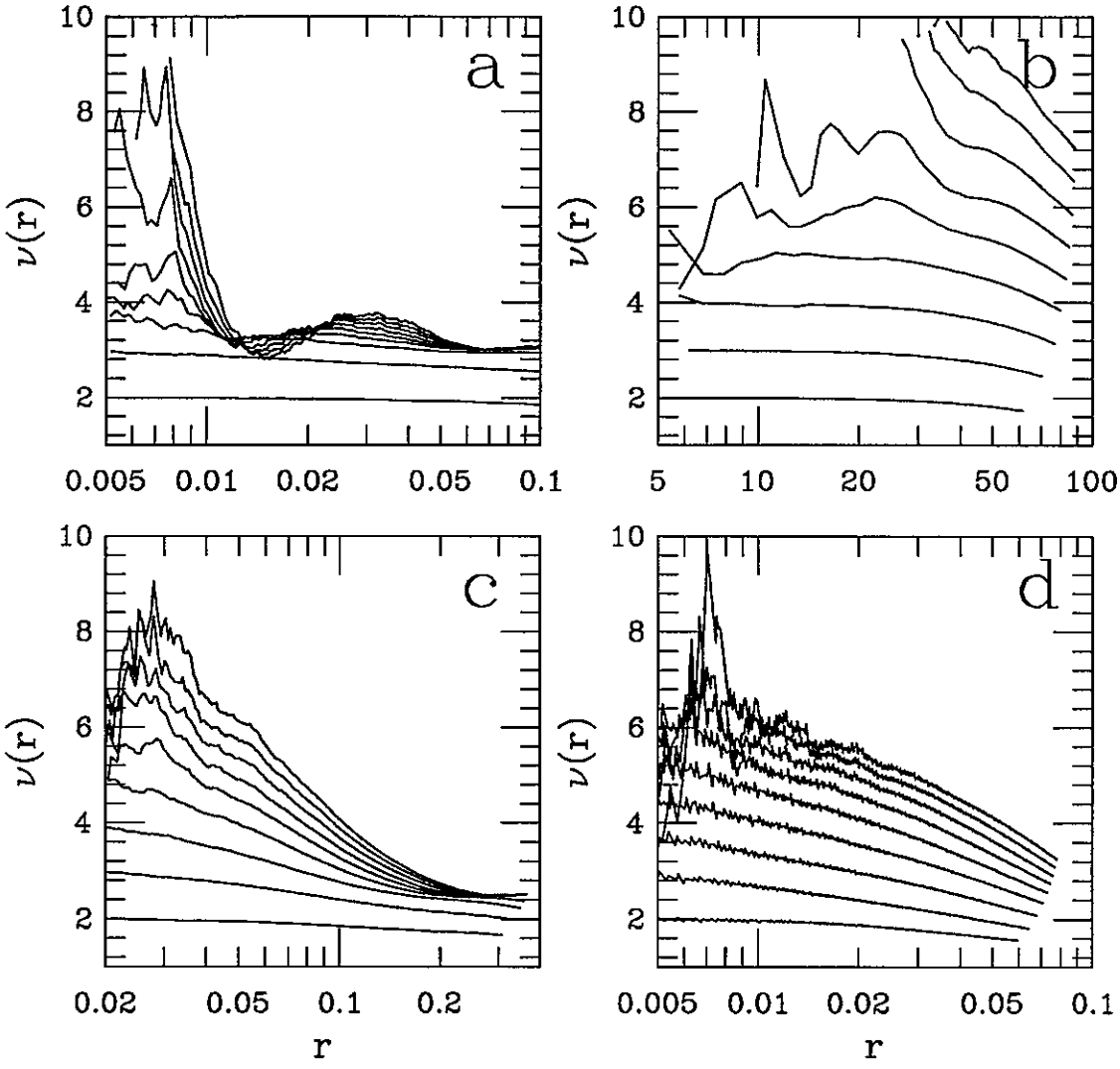
# Figure 7


a
b
c
d
e
f
$x_{i-\tau}$
$x_i$
$E_m(k)$
$\rho_m$ (k)
k
10000
5000
0
0.4
0.35
0.3
0.25
0.2
5
10
20
50
100
200
−5000
0.9
0.8
0.7
150
0.5

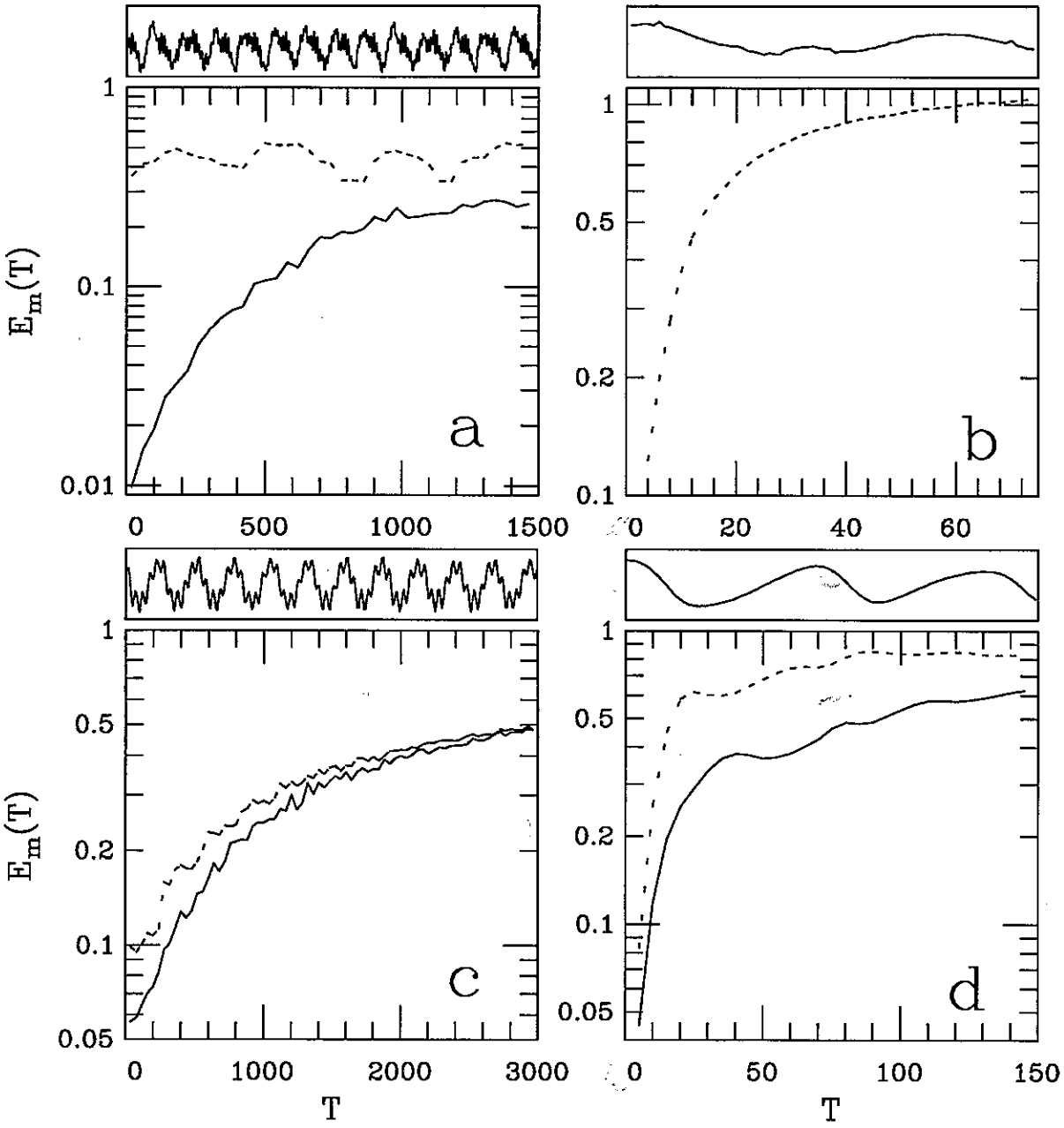
# Figure 8


a
b
c
d
C(r)
ν(r)
r
1
0.1
0.01
0.001
0.0001
10^-5
10^-6
10^-7
10^-8
10^-9
5 10 20 50 100 200
1 2 3 4 5 6

# Figure 9



# Figure 10


a
b
c
d
$E_m(T)$
$E_m(T)$
T
T
1
0.1
0.01
0
500
1000
1500
1
0.5
0.2
0.1
0
20
40
60
1
0.5
0.2
0.1
0.05
0
1000
2000
3000
1
0.5
0.2
0.1
0.05
0
50
100
150