Linear filters and nonlinear forecasting

ALUN L. LLOYD AND MICHAEL B. GRAVENOR

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

SUMMARY

We consider the consequences of using linear filters to reduce noise before analysing short time series for low-dimensional chaotic behaviour. We discuss mathematical theory which suggests that certain filters should not affect the results of particular nonlinear analyses. We note that these results have only been proved for purely deterministic systems and need not be true when a stochastic component is present in the time series. In particular, we demonstrate that simple moving average filters can falsely suggest that a white noise data set is chaotic by using a test commonly used by biologists. This incorrect result is not obtained if the method of surrogate data is used together with this test. The results demonstrate the extreme care needed when analysing small data sets by using sophisticated mathematical techniques. The graphical technique we describe may also aid testing for linearity in time series.

1. INTRODUCTION

There are many difficulties to be faced in the analysis of biological time series. Only some of the variables needed to describe the state of the system have been measured. Sampling errors give rise to noise in the data, in addition to any noise due to randomness in the underlying biological process. The time series are often of a short length and may consist of data sampled infrequently. Partial solutions to some of these problems have been provided by recent advances in dynamical systems theory. 'Phase space reconstruction' (Takens 1981) allows the essential part of the dynamics of a multi-dimensional system to be inferred from a single variable sampled at discrete time intervals. Nonlinear forecasting (Farmer & Sidorowich 1987; Sugihara & May 1990) has been suggested as a method which can differentiate between stochastic and various forms of deterministic dynamics. Linear filters can be used to reduce observational noise. However, these mathematical results have only been studied under idealized conditions, for data sets which are far removed from those encountered by biologists in experimental situations. Because of the short length of many biological time series, nonlinear forecasting is often the preferred tool for nonlinear analysis. In this paper we show that nonlinear forecasting can give misleading results when applied to short, noisy time series which have been filtered in an attempt to reduce the level of noise. However, a more sophisticated approach can identify many of the occasions when stochastic dynamics mimic chaotic dynamics. Our results echo those of authors who use different methods, such as dimensional calculations, to test for determinism in time series (Rapp *et al.* 1993).

2. METHODOLOGY

(a) *Nonlinear forecasting*

Given a time series consisting of n values sampled at equal time intervals, we use the 'method of delays' (Takens 1981; Sauer *et al.* 1991; Sauer & Yorke 1993) to reconstruct the dynamics of the original system. The E-dimensional set of points $\boldsymbol{y}_j = (x_j, x_{j-\tau}, \ldots, x_{j-(E-1)\tau})$ is constructed, and we call this reconstruction the phase space. The embedding dimension E must be chosen appropriately; the method we adopt is to choose E so as to optimize the predictions made by nonlinear forecasting (Sugihara & May 1990). Although in theory one can choose any value for the lag time τ, practical problems arise if the data are oversampled (τ small) or undersampled (τ large). In our examples we set τ equal to one.

Nonlinear forecasting can now be used to analyse the data. This technique can be used to differentiate between non-chaotic deterministic dynamics, chaos and deterministic dynamics with additive white noise. The basic idea is to use some part of the data set (in this paper we use the first half of the data set) to produce a set of patterns of behaviour which is then used to make predictions on the remainder of the data.

These predictions are made for several time steps into the future, and their accuracy is assessed by plotting predicted x values against the actual x values and calculating the correlation coefficient (r). The curve obtained by plotting r values for different prediction time intervals T may then give information about the nature of the dynamics. Chaotic deterministic dynamics are characterized by the exponential divergence of neighbouring trajectories, so predictions become less accurate over longer time intervals. This is observed as an r–T curve that has high r values for small T but falls sharply as T increases. If the dynamics are deterministic, but not chaotic, then the curve shows a high r value for all prediction intervals. The effect of additive white noise is to reduce the observed r value by an amount which does not vary much with T. However, it has been realized that autocorrelated noise (also called coloured noise) can give misleading results. This may pose a problem in interpreting the results obtained for biological time series, as most biological populations tend to exhibit short-term autocorrelations.

Figure 1 demonstrates these different forms of the r–T curves using data sets generated from three simple equations: (i) the chaotic logistic map, $x_{j+1} = 4x_j(1-x_j)$; (ii) a sum of

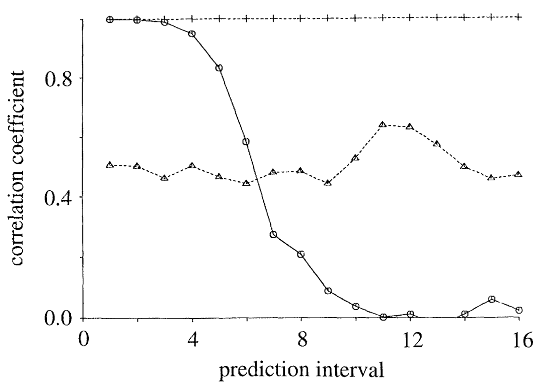

Figure 1. Correlation coefficient against prediction interval (r–T) curves for the chaotic logistic map, $x_{j+1} = 4x_j(1-x_j)$ (solid line and circles), the sum of two sinusoids, $x_j = \sin j + 2\sin 2j$ (broken line and crosses), and the sinusoidal time series with independent normal deviates $\mathcal{N}(0,1)$ added (broken line and triangles). For each of these curves, 1000 data points were used, the embedding dimension E was 3. Normal deviates were generated by the NAG routine G05DDF.

two sinusoids, $x_j = \sin j + 2\sin 2j$; and (iii) the sinusoidal time series, to which we add observational noise, simulated by independent normal deviates $\mathcal{N}(0,1)$.

Many different prediction methods have been proposed (Casdagli 1989). The method used to produce the predictions makes quantitative changes to the r–T curve, but does not change its overall shape (Kennel & Isabelle 1992). We use a zeroth-order predictor, where we just follow the behaviour of the nearest neighbour of the test point.

(*b*) *Predictability and Lyapunov exponents*

We have seen that for chaotic systems the predictability falls off with increasing T. This is a consequence of the sensitive dependence to initial conditions shown by chaotic systems, with nearby trajectories diverging, on average, exponentially at a rate given by the largest Lyapunov exponent (λ) of the system. This divergence will clearly restrict predictability; any errors introduced by the prediction scheme will tend to grow exponentially. Farmer & Sidorowich (1987) and Casdagli (1989) give asymptotic results that connect the prediction error to the Lyapunov exponents. It should be emphasized that Lyapunov exponents are a global property of a chaotic system, measuring divergences averaged over the whole attractor. Local Lyapunov exponents can be defined which may be more useful in quantifying predictability for different parts of an attractor (Gallez & Babloyantz 1991).

(*c*) *Linear filtering*

To remove observational noise from experimental data, some sort of filtering procedure may be used. The most commonly used is a low-pass filter which removes high-frequency components by a smoothing (averaging) process. One example of this is an equal weight m-point moving average

$$z_j = (1/m) \sum_{k=0}^{m-1} x_{j+k}.$$

Moving averages have been used in the analysis of data from measles epidemics (Schaffer & Kot 1985).

Low-frequency components may be removed by using a high-pass filter; the simplest example is taking first differences of the time series which heightens short-term variations in the data.

Both of these filters act in the time domain, but by means of Fourier transforms we can create a linear filter in the frequency domain. Let $\{\tilde{x}_k\}$ denote the n components of the Fourier transform. We can apply any linear map to these components before inverting the transform. For instance, Rapp *et al.* (1993) used a filter defined by multiplying $\{\tilde{x}_k\}$ by F_k, where

$$F_k = \max[0, 1-ck^2], \quad c = 0.37 \times 10^{-6}.$$

When talking about filters we distinguish between finite impulse response (FIR) and infinite impulse response (IIR) filters. These are defined by considering the output sequence z_j obtained from the input sequence which consists of all zeros except at the kth entry where it is one. The filter is an FIR filter if only a finite number of z_j are non-zero, otherwise it is an IIR filter. All three filters defined above are FIR filters, and in this paper we are primarily concerned with FIR filters.

For time series generated by deterministic systems there are mathematical results (Sauer *et al.* 1991; Broomhead *et al.* 1992; Sauer & Yorke 1993) which show that delay coordinate properties are unchanged by using an FIR filter, in particular Lyapunov exponents and the dimension of the attractor are unchanged. These theorems hold for generic filters and observations, and so it is possible to construct examples which appear to contradict these results (Sauer & Yorke 1993). IIR filters introduce new Lyapunov exponents and can change the dimension of the attractor (Badii & Politi 1986; Sauer & Yorke 1993).

3. THE EFFECTS OF SMALL DATA SETS

For short time series, which may be noisy, practical issues become important. Although the theoretical results show that FIR filters leave certain dynamical quantities unchanged, they do not tell us how estimates of them are altered by using filtered data. For instance, Broomhead *et al.* (1992) explain how application of an FIR filter may lead to overestimation of the dimension. It must also be emphasized that the mathematical results were derived for noise-free systems, and that filtering can effect any stochastic component present in the data.

It is often impossible to avoid filtering of data, as it may be introduced by the equipment used to take measurements; for instance, low-pass filters are used by all EEG systems (Rapp 1993). If the filter is an FIR filter then the Lyapunov exponents are invariant. Because these exponents are linked to predictability, it may be argued that filtering should not affect the prediction process. (The new Lyapunov exponents introduced by an IIR filter may have an effect on predictability.)

It is not widely realized that processing the data can introduce short-term autocorrelations. Such autocorrelations can affect the prediction process. Moving averages are commonly used to reduce noise, or to remove seasonal effects, in time series. This procedure can fool the nonlinear forecasting test into seeing a white noise time series as chaotic.

We take x_j to be independent $\mathcal{N}(0,1)$ random variables; numerically we simulate this by using

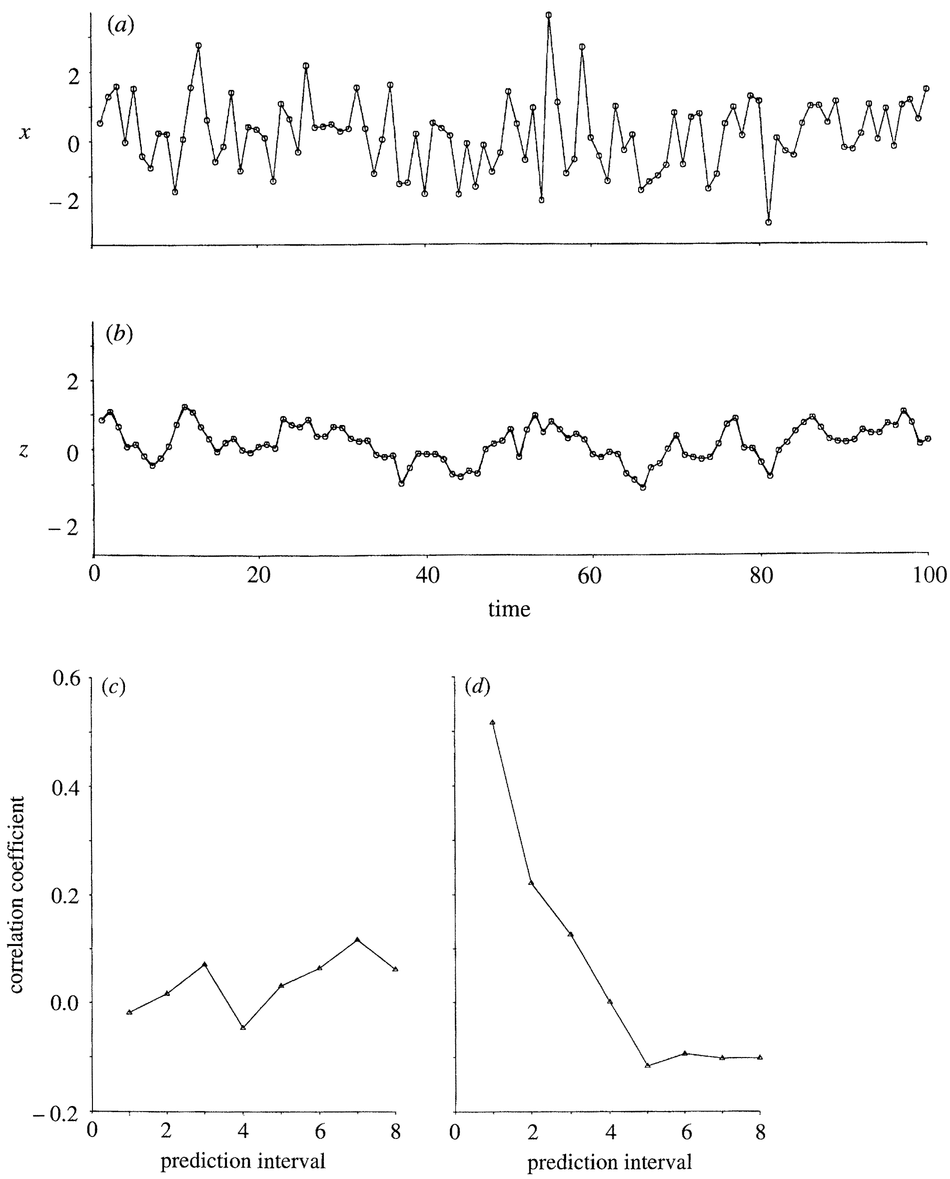

Figure 2. (*a*) The first 100 points from the white noise data set, generated by taking values independently from $\mathcal{N}(0,1)$. (*b*) The first 100 points obtained by using a four-point equal weight moving average on the white noise data set. (*c*) r–T curve for the white noise data set. (*d*) r–T curve for the smoothed white noise data set.

standard routines, such as those provided by the NAG library (NAG Central Office, 7 Banbury Road, Oxford, U.K). We then take a four-point moving average,

$$z_j = (1/4)\,(x_j + x_{j+1} + x_{j+2} + x_{j+3}),$$

and apply the test to this time series. Figure 2 (*a*, *b*) shows the time series x_j and z_j; the smoothing effect of the filter is quite visible. Figure 2 (*c*, *d*) shows the r–T curves for these time series. The former shows that our forecasting scheme cannot produce good predictions for the completely random white noise time series, but the shape of the curve in the latter closely resembles that seen for a chaotic series. This form of filter introduces obvious short-term correlations into time series, but more subtle filtering procedures can also do this. The time domain filter of Rapp *et al.* (1993) is an attempt to simulate the kind of filter found in EEG equipment, and they found that it had a similar smoothing effect on white noise.

4. THE METHOD OF SURROGATE DATA

Autocorrelated noise can give r–T curves similar to those generated by chaotic data sets. The method of surrogate data (Theiler *et al.* 1992) is an attempt to detect false positive results such as this. The basic idea is to take the observed time series and generate series which are random but still preserve certain chosen statistical properties of the original series. These random series are called the surrogates. The forecasting test is applied to both the surrogate and the original series, and one looks to see whether there is a significant difference between the correlation coefficients observed for the randomized and original time series. This idea was used by Stone (1992) when he considered the question of whether the New York measles data set contained a low-dimensional chaotic signal.

This procedure can be placed in a more rigorous statistical framework. We take a null hypothesis, H_0, for instance, H_0: All the structure in the time series is given by the autocorrelation function. If we know the distribution of the test statistic (in our case, the value of the correlation coefficient T steps into the future) under H_0, we can determine whether the result obtained from the actual time series provides significant evidence against the model proposed by the null hypothesis. For a given H_0, and making certain assumptions about the distribution of noise, we may be able to calculate this distribution. Usually, however, we do not know this distribution, so we estimate it by direct Monte Carlo simulation. We calculate the value of the statistic for many surrogate series, and from this set of statistics we can estimate the distribution. Theiler *et al.* (1992) suggest the use of a rank statistic approach, where one looks to see in which percentile of the distribution of the surrogate statistics the observed test statistic falls. An approximate significance level can then be quoted.

The method chosen to produce the surrogates is very important: our surrogate series are produced by using a method which preserves the autocorrelation spectrum (or, equivalently, the Fourier power spectrum). The technique is discussed in detail in Franke & Härdle (1992); this and techniques appropriate to other choices of H_0 are discussed in Theiler *et al.* (1992) and Smith (1992). The Fourier transform of the series is taken, the phases of the (complex) transformed series are randomized, and the inverse Fourier transform is then taken to give the surrogate series. The phases are chosen independently and uniformly between 0 and 2π. To ensure that the surrogate series are real, a symmetry condition must be imposed on the phases: $\phi_{-k} = -\phi_k$. The surrogate series therefore have the same linear autocorrelations as the original series, but any non-linear structure present in the observed series is destroyed. By using these surrogate series we can determine whether the shapes of the r–T curves are due to the linear autocorrelations in the series. This method for generating the surrogates has been used to test whether time series data are consistent with a stationary linear Gaussian process (Theiler *et al.* 1992).

5. RESULTS

This method was applied to surrogate series generated from various data sets. Figure 3*a* shows the results obtained for the simple logistic map. There is a significant difference between the curves for the time series and its surrogates. For predictions up to six time steps into the future, none of the 100 surrogate series has a higher correlation coefficient than the original time series. This is because the autocorrelation function of the logistic map is zero for all times into the future (the iterates are delta-correlated); none of the observed predictability is due to linear autocorrelations. For the white noise data set there is no significant difference between the curves for the data set and the surrogates generated from it (figure 3*b*). The same is true for the filtered white noise time series (figure 3*c*); the decline in predictability is seen to be a consequence of the linear correlations introduced by the filtering. For the filtered logistic data, however, we can still distinguish a significant difference between the curves (figure 3*d*). This data set is chaotic but shows significant linear autocorrelations. As a result there is a smaller difference between the quality of predictions made for the original and surrogate series than was seen for the unfiltered data (figure 3*a*). Figure 3*e* shows r–T curves for data generated by adding observational noise, simulated by independent $\mathcal{N}(0, 0.1)$ deviates, to the logistic map data, and 100 surrogates generated from this series. The results obtained when this data set is filtered are shown in figure 3*f*. Because the iterates of the logistic map lie between 0 and 1, the noise-to-signal ratio is high in this example. Even though the noise has a large effect on the r–T curves for the original data set, we can still observe a difference between the surrogates and the original data.

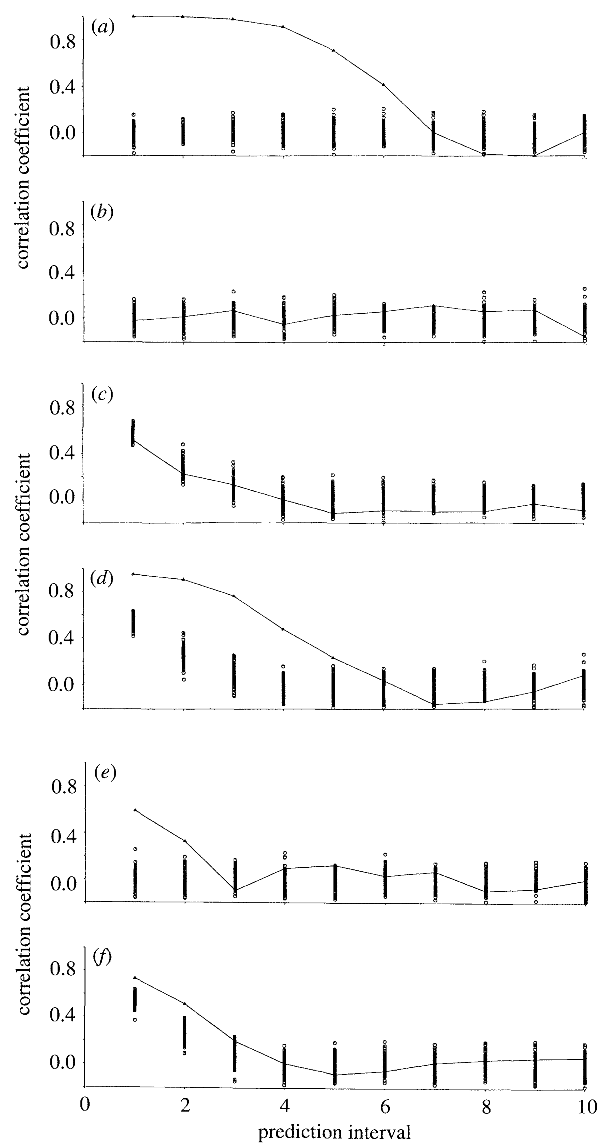

Figure 3. Surrogate data analysis using 500 points from the following data sets: (*a*) chaotic logistic map; (*b*) white noise data set; (*c*) white noise data smoothed by using four-point moving average; (*d*) logistic map data smoothed by using four-point moving average; (*e*) logistic map plus independent $\mathcal{N}(0, 0.1)$ deviates; (*f*) (logistic + noise) smoothed by using four-point moving average; 100 surrogate series were generated as explained in the text; r–T curves for these series are shown by the circles. The r–T curve for the original data is shown as the solid line.

It has been pointed out (H. Tong, personal communication) that figure 3 can be interpreted in the context of testing for linearity as discussed above. Figure 3*a* shows strong evidence against linearity, as would be expected because the iterates of the logistic map are delta-correlated. Figure 3*b*, *c* shows that the data are consistent with linearity. Figure 3*d* shows that, even after introducing linear correlations by using a four-point moving average, the logistic data still exhibit fairly large departures from linearity. Figure 3*e*, *f* shows that the addition of high variance noise masks most (but not all) of the evidence of departure from linearity seen in the noise-free cases.

By using this method we can separate the effects on predictability of linear autocorrelations, such as those introduced by filtering, and other dynamical effects. If we have an unfiltered data set (keeping in mind the caveat that some equipment may use filters), we can apply linear filters in an attempt to reduce noise, and then check that any structure seen in the r–T curve is not a result of filtering.

6. DISCUSSION

The idea that filtering can affect statistical analyses of time series is not new; Cole (1954) drew biologists' attention to the fact that smoothing random data by using a moving average can lead to the introduction of spurious cyclic behaviour. However, the filtering of data is often highly desirable, or even unavoidable. Unless suitable precautions are taken, filtering may bias further analyses applied to the data. The method presented here can identify when this is the case. This not only enables the rejection of spurious results, but also allows filtering procedures to be used with more confidence because we can determine whether structure in the filtered data was present in the original data or was introduced by the filtering.

Surrogate data techniques can be used more generally to test whether a time series may be a realization of various other stochastic models, given appropriate choices for H_0. To test whether the data are consistent with a given H_0, one needs an appropriate method to produce the surrogates. For many interesting choices of null hypotheses (for instance to test noisy limit cycle behaviour), it is not clear how the surrogates can be generated. Of course, given a finite time series we can never be certain that it arises from a deterministic system, but we can discredit a range of specific models. As a result, it is always important that the time series is taken in context. The mathematical techniques do not take any account of what sort of system the data comes from, but to obtain useful information we need to construct relevant null hypotheses.

Other tests for determinism, such as dimension calculations, have been used with surrogate data, and it may be helpful to use more than one test because different tests are sensitive to different features in the data (Theiler *et al.* 1992). However, nonlinear forecasting can be used with relatively short data sets; as a result it may be the only nonlinear analysis method available for short biological data sets. Although this test has known weaknesses when applied to autocorrelated time series which are typical in biology, the ideas discussed in this paper enable conclusions to be drawn more confidently from the test results.

We thank R. M. May and A. R. McLean for their helpful discussions and encouragement. We also thank the referees for their thoughtful comments which clarified several points. A.L.L. is supported by the Wellcome Trust (Biomathematical Scholarship, grant number 036143/Z/92/Z), and M.B.G. by a Natural Environment Research Council Studentship.

REFERENCES

Badii, R. & Politi, A. 1986 On the fractal dimension of filtered chaotic signals. In *Dimensions and entropies in chaotic systems* (ed. G. Mayer-Kress). Berlin: Springer-Verlag.

Broomhead, D. S., Huke, J. P. & Muldoon, M. R. 1992 Linear filters and non-linear systems. *Jl R. statist. Soc.* **54**B, 373–382.

Casdagli, M. 1989 Nonlinear prediction of chaotic time series. *Physica* **35**D, 335–356.

Cole, L. C. 1954 Some features of random population cycles. *J. Wildl. Mgmt* **18**, 2–24.

Farmer, J. D. & Sidorowich, J. J. 1987 Predicting chaotic time series. *Phys. Rev. Lett.* **59**, 845–848.

Franke, J. & Härdle, W. 1992 On bootstrapping kernel spectral estimates. *Annls Statist.* **20**, 121–145.

Gallez, D. & Babloyantz, A. 1991 Lyapunov exponents for nonuniform attractors. *Phys. Lett.* **161**A, 247–254.

Kennel, M. B. & Isabelle, S. 1992 Method to distinguish possible chaos from colored noise and to determine embedding parameters. *Phys. Rev.* **46**A, 3111–3118.

Rapp, P. E. 1993 Chaos in the neurosciences: cautionary tales from the frontier. *Biologist* **40**, 89–94.

Rapp, P. E., Albano, A. M., Schmah, T. I. & Farwell, L. A. 1993 Filtered noise can mimic low-dimensional chaotic attractors. *Phys. Rev.* **47**E, 2289–2297.

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.

Schaffer, W. M. & Kot, M. 1985 Nearly one dimensional dynamics in an epidemic. *J. theor. Biol.* **112**, 403–427.

Smith, L. A. 1992 Identification and prediction of low dimensional dynamics. *Physica* **58**D, 50–76.

Stone, L. 1992 Coloured noise or low-dimensional chaos? *Proc. R. Soc. Lond.* B **250**, 77–81.

Sugihara, G. & May, R. M. 1990 Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature, Lond.* **344**, 734–741.

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

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. 1992 Testing for nonlinearity in time series: the method of surrogate data. *Physica* **58**D, 77–94.

Received 21 January 1994; accepted 28 February 1994