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

# Nonlinear Forecasting for the Classification of Natural Time Series

George Sugihara

*Phil. Trans. R. Soc. Lond. A* 1994 **348**, 477-495
doi: 10.1098/rsta.1994.0106

# Nonlinear forecasting for the classification of natural time series†

By George Sugihara

*Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0202, U.S.A.*

There is a growing trend in the natural sciences to view time series as products of dynamical systems. This viewpoint has proven to be particularly useful in stimulating debate and insight into the nature of the underlying generating mechanisms. Here I review some of the issues concerning the use of forecasting in the detection of nonlinearities and possible chaos, particularly with regard to stochastic chaos. Moreover, it is shown how recent attempts to measure meaningful Lyapunov exponents for ecological data are fundamentally flawed, and that when observational noise is convolved with process noise, computing Lyapunov exponents for the real system will be difficult. Such problems pave the way for more operational definitions of dynamic complexity (cf. Yao & Tong, this volume).

Aside from its use in the characterization of chaos, nonlinear forecasting can be used more broadly in pragmatic classification problems. Here I review a recent example of nonlinear forecasting as it is applied to classify human heart rhythms. In particular, it is shown how forecast nonlinearity can be a good discriminator of the physiological effects of age, and how prediction–decay may discriminate heart-disease. In so doing, I introduce a method for characterizing nonlinearity using 'S-maps' and a method for analysing multiple short time series with composite attractors.

## 1. Introduction

Suppose we are given a time series such as the one shown in figure 1*a*, where we have few preconceptions. These are data for the Scripps Pier diatoms that were extracted from a record that W. E. Allen collected over a 20 year period (Tont 1981). From a dynamical systems viewpoint, three potential sources of complexity here are: (1) observational or measurement error; (2) the dimensionality of the system, or the possible action of many variables (possibly giving rise to process noise); and (3) the multiplicative or nonlinear interaction of these variables. Herein, such nonlinear behaviour, both chaotic and stable, shall be called 'dynamic complexity'.

Real time series such as this one probably contains a mixture of these effects, and in the past decade, there has been a special focus in the natural sciences

† This paper was produced from the author's disk by using the TEX typesetting system.

*Phil. Trans. R. Soc. Lond.* A (1994) **348**, 477–495
*Printed in Great Britain*

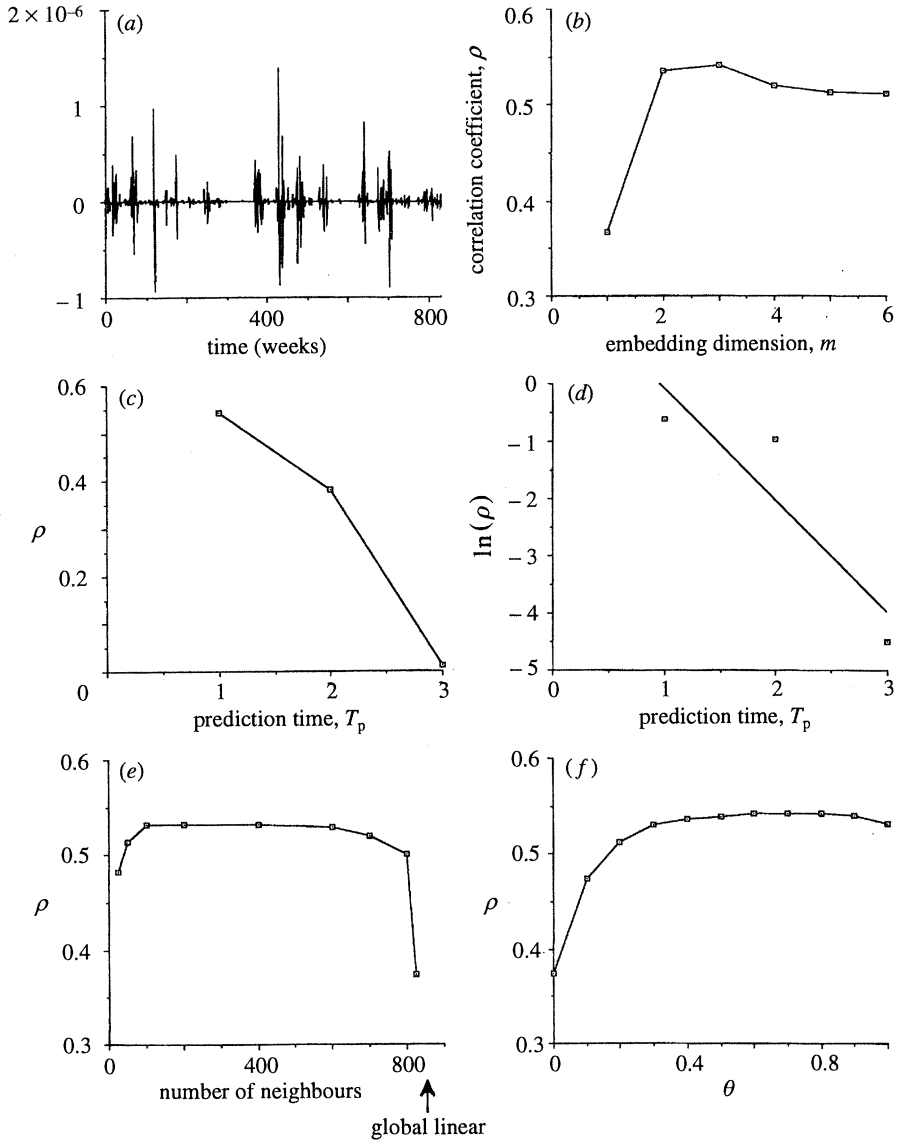


Figure 1. (*a*) Scripps Pier diatom series. Weekly abundances first differenced. (*b*) Predictability (correlation coefficient, $\rho$) versus embedding dimension for the Scripps diatoms. Note the peak in predictability at $m = 3$, followed by a broad decline. The nonlinear predictor used here and in (*c*), (*d*) and (*f*) is an S-map (see § 2) with $\theta = 0.75$. (*c*), (*d*) Prediction–decay plots. (*e*) DVS plot (see § 1 *b* (i)). Note the similarity between this plot and the one obtained in figure 4*b* for stochastic chaos. (*f*) A difference plot (§ 2) to detect nonlinearity in these data. $\theta = 0$ corresponds to the global linear model; the model becomes more nonlinear as $\theta$ is tuned upwards. Notice how markedly predictability is improved as the model becomes more nonlinear. A simple difference test (§ 2*a*) for the $H_0$ : $\rho_{\text{linear}} = \rho_{\text{nonlinear}}$ suggests rejection of $H_0$ at $p < 0.0005$. (After Sugihara & May 1990.)

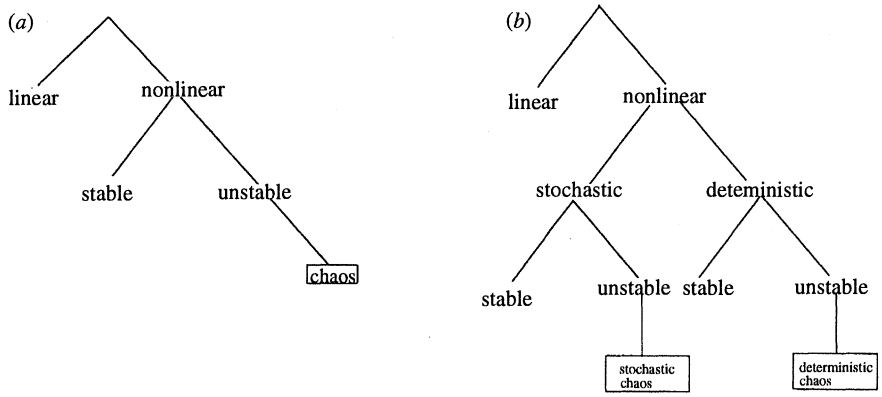


Figure 2. (*a*) Classification of null alternatives under the assumption of observational noise. (*b*) Classification of null alternatives under the assumption of process noise. Process noise is that part of the system which is unexplained and is approximated as IID noise. Strictly speaking with process noise there can be no deterministic chaos, except in the limiting case (no process noise) or in a noise-free skeleton.

to uncover evidence for dynamic complexity (Ascioti *et al.* 1993; Lefebvre *et al.* 1993; Tsonis 1993; Greshenfeld & Weigend 1993). The compelling reasons for this interest are that both nonlinearity and chaos can hold out the promise of producing complex time signatures with relatively simple (low dimensional) underlying dynamics, and the acknowledgment that such nonlinearity may allow improved short-term predictability while at the same time it can place limits on long-term forecastabilty. That is, one would like to learn the limits to forecastabilty, how complex the phenomenon is, and how important nonlinear effects should be in the structure of an explicit model.

I shall now review some of the issues in this search, particularly as they apply to uncovering nonlinearities in biology where the issues differ in a fundamental way from those in the conventional signal-processing literature.

## 2. Observational noise

Figure 2*a* shows the alternatives that need to be considered when one is only considering observational noise. Here, given the noise-free process, $\boldsymbol{S}_t = \boldsymbol{F}(\boldsymbol{S}_{t-1})$, $\boldsymbol{S}_t \in R^m$, the series that is actually observed is

$$X_t = h(\boldsymbol{S}_t) + \epsilon_{\text{obs}}, \tag{2.1}$$

$X_t \in R^1$, $h$ is the observation function that maps points in $R^m$ to $R^1$, and $\epsilon_{\text{obs}}$ is the IID observational error. Most of the early work which involved forecasting focused on distinguishing chaos from measurement noise (Casdagli 1989; Sugihara & May 1990; Casdagli *et al.* 1992). This is an especially important distinction in ecological data where population estimates may vary over several orders of magnitude. Forecast performance is used as the criterion for judging both the sufficiency of an embedding and the nonlinearity of the process. Simple scaling rules have been developed for (2.1) to distinguish the decay in predictability that would be expected from chaotic signals (Wales 1991) versus those from linear and fractal processes contaminated with white and coloured noise (Tsonis & Elsnor

1992). The basic philosophy here is that the model that produces the best out-of-sample forecast is the best dynamical description of the data. Thus a time series would be classified as nonlinear, if a nonlinear model is capable of producing significantly better forecasts (out-of-sample) than an equivalent linear model. And it would be further classified as chaotic if, in addition, there was evidence of rapid (exponential) decay in predictability characteristic of chaotic systems. The obvious appeal is in its directness and in the relatively modest amounts of data it requires (Sugihara & May 1990). The method is easy to apply and, as we shall see, can also be applied to study cases with process noise.

### (*a*) *Process noise and stochastic chaos*

As shown in figure 2*b*, the problem appears to become more complicated in terms of the number of alternatives that need to be considered when the decision is made to explicitly add process noise to the picture. In the exploration of real data, this is ultimately a philosophical decision. Process noise is that part of the system dynamics that remains unexplained. According to this view, the underlying system consists of an explained part (that which is modelled, or the noise-free skeleton) and an unexplained part (approximated by IID noise of some kind, $\epsilon_{\text{process}}$). Notice that process noise will generally be a function of the noise-free map that is chosen. Thus, according to this view, the real $m$-dimensional system dynamics can only be approximated in $k$ dimensions, with the dynamics in the remaining $m - k$ dimensions approximated as noise. Hence, if the true system dynamics are

$$\boldsymbol{S}_t = \boldsymbol{F}(\boldsymbol{S}_{t-1})$$

$\boldsymbol{S}_t \in R^m$. Then in $R^k$, $k \ll m$, we have

$$\boldsymbol{Z}_t = \boldsymbol{J}(\boldsymbol{S}_t),$$

where $\boldsymbol{J}$ behaves like an observation function, mapping $R^m$ to $R^k$. But notice that

$$\boldsymbol{Z}_t \neq \boldsymbol{G}(\boldsymbol{Z}_{t-1}),$$

where $\boldsymbol{G}$ is a diffeomorphism since the trajectories are no longer unique (may cross). Thus, we approximate the full $m$-dimensional dynamics in $k$ dimensions as

$$\boldsymbol{Z}_t \approx \boldsymbol{G}(\boldsymbol{Z}_{t-1}, \epsilon_{\text{process}}), \tag{2.2}$$

where $\boldsymbol{G}(\boldsymbol{Z}_{t-1})$ is the modelled part or noise-free skeleton and the dynamics in the remaining $m-k$ dimensions are approximated as IID noise of some kind. A specific form for incorporating process noise which is commonly studied is one where $\boldsymbol{Z}_t = \boldsymbol{G}(\boldsymbol{Z}_{t-1})$ is the noise-free map, and error-free (in terms of measurement error) observations on the system are approximated (although equality is used) as

$$\boldsymbol{Z}_t = \boldsymbol{G}(\boldsymbol{Z}_{t-1} + \epsilon_{\text{process}}). \tag{2.3}$$

In this example, process noise is added and iterated through the dynamics of a noise-free map. The invariant measure for the system is induced by the process noise.

An interesting phenomenon that emerges with process noise is stochastic chaos (Tong 1990; Chan & Tong 1994; Rand & Wilson 1991; Sidorowich 1992), which is likely the most ubiquitous form of chaos in nature. This is illustrated by the




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

noise-driven logistic map with a two-point cycle shown in figure 3. Here, $X_t = aX_{t-1}(1 - X_{t-1}) + \epsilon_{\text{process}}$, $a = 3.4$, $\epsilon_{\text{process}}$ is gaussian $(0, \sigma)$ with the domain boundaries (0,1) treated as reflecting barriers, to keep the function restricted to the interval (0,1). Again, the invariant measure, or the specific sequence of points visited on the noise-free map is not determined by the map itself, but by the interaction of the noise on the map. Thus, if the Lyapunov exponent, $L$, for the system is calculated from the geometric mean of the jacobian elements, $J$, the invariant measure must be induced by the noise:

$$L = \lim_{n\to\infty} \frac{1}{n} \sum \ln(|J|).$$

If the noise level is increased from zero to very low levels, there is a tendency to drift slightly away from the 2-point equilibria, initially spending more time visiting flatter portions of the map (net contraction) to yield more stable dynamics (figure 3*b*, *d*). As the noise level is increased a bit further, steeper, unstable portions of the map are visited more frequently (regions of net divergence), and the Lyapunov exponent can become positive. Thus, the system can become stochastically chaotic, although the skeleton is stable (figure 3*c*, *d*), and such behaviour can be achieved at very low noise levels ($\sigma = 0.05$). In this example, the Lyapunov exponent appears to be a smooth function of noise level (figure 3*d*).

(i) *Identifying stochastic chaos*

Strictly speaking, deterministic chaos (the right-hand limb of figure 2*b*) does not exist for a system with process noise; only stochastic chaos is possible. It may exist as an abstraction for a noise-free skeleton, and as will be discussed, this may be a reasonable thing to consider in certain circumstances such as noise-filtering in engineering applications (Abarbanel *et al.* 1993); however, for a natural system with process noise, the noise-free skeleton may have no meaning at all. Here the difficult alternatives that one wants to distinguish are stable versus unstable nonlinear stochasticity (stochastic chaos).

(ii) *DVS plots*

Figure 4*b* shows an attempt to use Casdagli's DVS criterion to make this distinction (Casdagli 1992) for the noise-driven logistic example. This forecasting test, which shall be explained in more detail in § 2, involves constructing a series of local linear maps from different numbers of neighbours. The maps constructed from very small neighbourhoods are essentially nonlinear models, and the map constructed from all neighbours is the global linear model. The main idea illustrated in figure 4*a* is that if the data are from a chaotic process then the local linear models with few neighbours should outperform those with large or intermediate numbers of neighbours; if the data are from a nonlinear stochastic process, the peak in performance should be realized at intermediate numbers of neighbours; whereas, if the data are from a linear process, the global linear model should perform best. This elegant criterion has been shown to work well at distinguishing those alternatives for which it was designed, but as seen in 4*b*, it does not distinguish stable versus unstable nonlinear stochasticity in any obvious way.




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

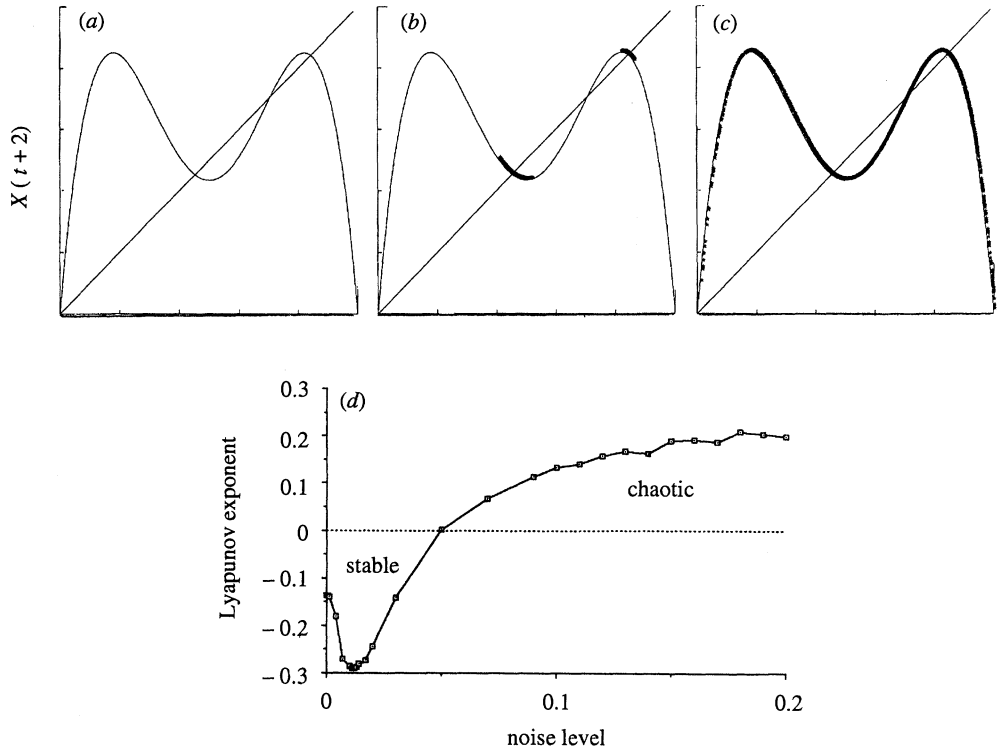


Figure 3. Stochastic chaos induced by the action of gaussian noise on a stable period-2 quadratic map ($X_{t+1} = 3.2X_t(1-X_t)+\epsilon$). (*a*) The noise-free ($\epsilon = 0$) time-2 map. (*b*) Shows what portions of the noise free map are hit as a very small level of noise ($\sigma = 0.005$) is introduced. Notice that because the system spends more time in the flatter more stable portions of the map, it is actually more stable than the noise-free map. (*c*) Has a slightly higher noise level ($\sigma = 0.1$) and shows how the system spends more time in the steeper unstable portions of the map (repellor regions where the tangent |slope| $> 1$). (*d*) Summarizes these effects showing how the Lyapunov exponent for the system (invariant measure is induced by the noise) is a smooth function of noise level. Thus, process noise can induce chaos even if the noise-free skeleton is chosen to be stable.

### (iii) *Prediction–decay curves*

Prediction–decay curves are obtained by plotting the correlation coefficient (Sugihara & May 1990; Wales 1991; Tsonis & Elsnor 1992) or r.m.s. error (Farmer & Sidorowich 1987) (both either logged or unlogged) against prediction time ($T_p$). As can be seen in figure 4*c*, stable systems will decay, however, at a rate which is considerably slower than the exponentially decaying unstable cases. Stable systems decay because the gaussian noise which is small on average, will occasionally produce a value from the tails of the distribution that will push the system into a repellor region (unstable or local divergence). Uniform noise from a distribution whose width is smaller than the smallest domain of attraction around the time-2 fixed points, will not decay. Thus, in general, one should expect steeper prediction–decay curves when the system is chaotic and shallower ones when it is stable. As can be seen in figure 4*c*, this qualitative distinction may be useful at identifying the extremes, but there is a smooth (albeit sharp with respect to increasing noise level) continuum in between, where the distinction will be blurred

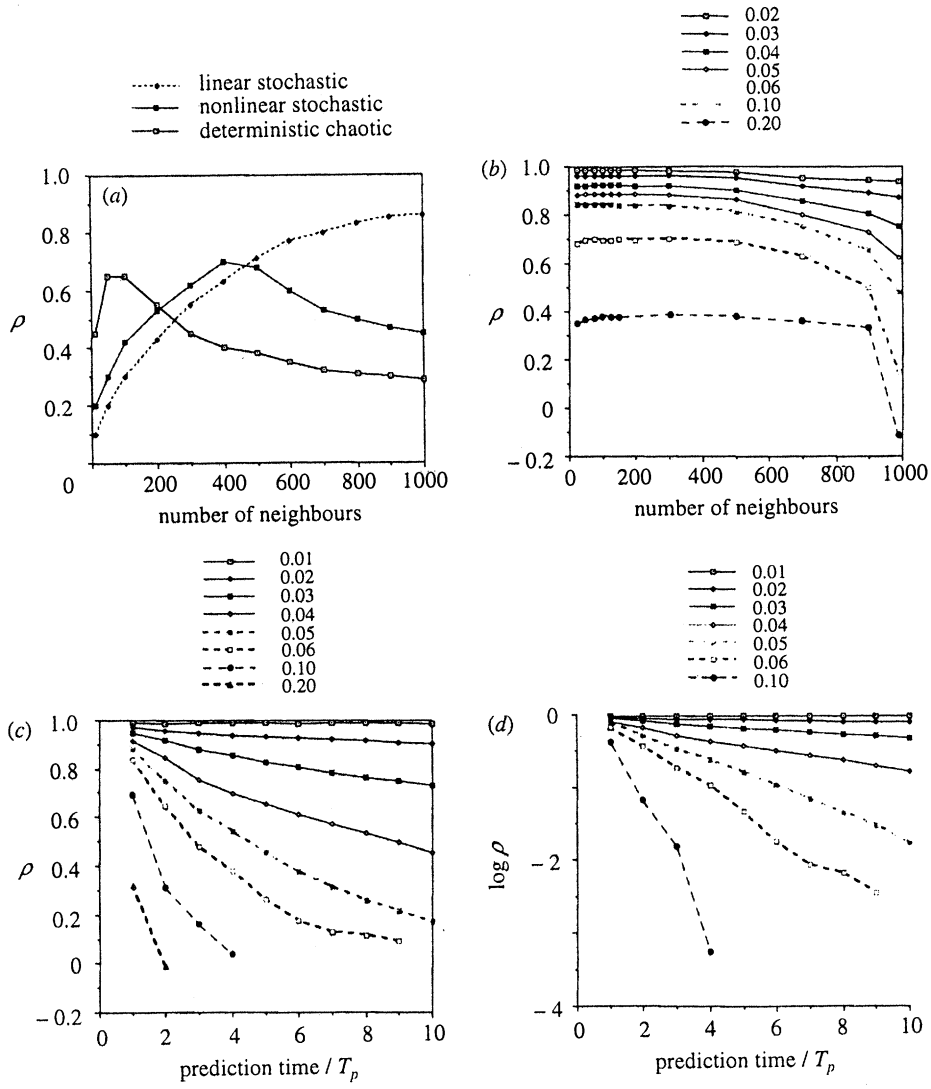


Figure 4. The use of forecast methods to detect stochastic chaos. (*a*) Hypothetical example of Casdagli's (1992) DVS plot to distinguish between chaotic, nonlinear stochastic, and linear stochastic alternatives. (*b*) DVS plots for the data in figure 3, to distinguish stable (solid lines) from unstable (dashed lines) nonlinear stochasticity. (*c*), (*d*) Prediction–decay curves to distinguish stable (solid lines) from unstable (dashed lines) nonlinear stochasticity. As expected (Sugihara & May 1990; Wales 1991), the stable cases exhibit slow decay while the unstable (chaotic) cases exhibit rapid exponential decay. With uniform noise instead of gaussian, the stable cases will show no decay.

in any practical sense. None the less, applied with care, prediction–decay curves can be a useful tool for distinguishing between stable and chaotic stochasticity.

### (iv) *Lyapunov exponents and reality*

Another way to make this distinction is to approximate the noise-free map with a smooth fitted function (Crutchfield & MacNamara 1987; Tong 1990; Nychka *et*

*al.* 1992) and then repeat the process used to compute the Lyapunov exponents in figure 3*d*. This idea has recently been applied, however incorrectly, in ecology (Ellner 1991; Turchin & Taylor 1992; Hastings *et al.* 1993; Hanski *et al.* 1993) where Lyapunov exponents have been computed for the abstracted noise-free maps, rather than for the system (in some cases fitting $R^5$ surfaces to time-series containing as few as 30 points!). In the fitting procedure the process noise is assumed to have the specific additive form of (2.3). However, it is then ignored in the invariant measure for the system. Computing Lyapunov exponents for a noise-free skeleton may make sense in engineering applications for noise-reduction, or if the skeleton is a classical or a well-defined fundamental model. Thus, if one has a transistor which is fed noisy inputs, where the noise-free skeleton is the action of the transistor, it is meaningful here to try to filter out the noise and understand the transistor. However, when one is trying to understand a natural system such as the Scripps diatom data in figure 1, and the map is obtained by phenomenological fitting so that one really does not know what it represents, it can be a meaningless exercise (e.g. are the coordinates all biological or could physical forcing be represented? What are the subsystem boundaries?). Moreover, whether the skeleton is linear or nonlinear or stable or unstable can depend to an uncomfortable extent on the method of fitting that one chooses to apply to the small noisy data sets that biologists must typically work with. Thus, although one can in principle compute Lyapunov exponents for a noise-free skeleton and determine whether it is deterministically chaotic by the rigorous definition, it is uncertain what this determination actually means, when the real dynamics are assumed to contain noise.

Although with model data from our noisy logistic example it may be possible to get reasonable estimates of the Lyapunov exponent, with real data, where observational noise may be involved, this may not be possible (Yao & Tong, this volume). With observational noise convolved on top of process noise we are looking at data of the form

$$Y_t = h(\boldsymbol{Z}_t) + \epsilon_{\text{obs}},$$

where again, $\boldsymbol{Z}_t = \boldsymbol{G}(\boldsymbol{Z}_{t-1} + \epsilon_{\text{process}})$, $\boldsymbol{Z}_t \in R^k$, $k \ll m$, and $h$ maps $R^k$ to $R^1$. The difficult problem is not in estimating some noise-free skeleton (Nychka *et al.* 1992; Turchin & Taylor 1992; Hastings *et al.* 1993). Rather, the difficulty, which may not have a practical solution (e.g. in terms of data requirements), is in disentangling the process noise from the observational noise to reconstruct the invariant measure. One can no longer simply compute products of the jacobians from the time series itself because the invariant measure is contaminated with observational noise. Unless there is prior knowledge of the observational noise or the process noise, it is difficult to see how an invariant measure for the system can be deduced from the data. More work needs to be done here.

### (*b*) *Should we care about chaos?*

The above conundrum has prompted Yao & Tong (this volume) to propose an alternative definition of sensitive dependence that appears to be more practical than Lyapunov exponents. One of the central reasons for wanting to know whether a system is chaotic is the possibility of low-dimensional determinism. This motivation is somewhat diminished by the fact and probable ubiquity of stochastic chaos, where ultimately the driving force for prediction decay might




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

be the interaction of some lower dimensional map with the remaining high dimensional process noise.

Regardless of the fact that we still cannot compute meaningful Lyapunov exponents, the various other methods that have been developed for nonlinear analysis of time series remain useful tools for understanding nature and for constructing arguments for chaotic dynamics. These include the new criterion of Yao & Tong (this volume), nonlinear forecasting (Kravtsov & Etkin 1984; Sugihara & May 1990; Casdagli 1989, 1992), and the use of null models (Brock 1986; Osborne & Provenzale 1987; Theiler *et al.* 1991). Forecast methods may be especially useful since they regard any process noise as may be present as convolved in the system. The three-way forecast criterion of Sugihara & May requires: (1) finding an optimal low-dimensional model (optimal in terms of a peak in predictability versus embedding); (2) demonstration of strong nonlinearity (nonlinear model performs significantly better than linear model); and (3) steeply declining prediction–decay curves. The emphasis is on nonlinearity and forecast decay. Operational methods such as this are easy to implement and may provide good evidence (albeit not 'proof') for the operation of chaotic dynamics narrowly defined. Alternatively, one could take the three-way criterion as an 'operational definition' of chaos.

## 3. Sequential locally weighted global linear maps: SLWGLM (or 'S-map')

The idea of locally weighted maps has been discussed in various contexts in the statistics literature (Cleveland 1979; Hardle 1990; Tong 1990), by connectionists (Bottou 1993), and in the literature on dynamical systems (Crutchfield 1979; Kravtsov & Etkin 1984; Crutchfield & MacNamara 1987; Farmer & Sidorowich 1987; Casdagli 1989, 1992; Casdagli *et al.* 1991, 1992; Diebold & Nason 1990; Stokbro & Umberger 1992). Such ideas have also been discussed in relation to the geometry of the attractor (Sugihara & May 1990; Mees 1991). I will now describe a predictor which is similar to several of the above techniques, but which gives more skilful forecasts when the time signature contains a mixture of linear and nonlinear components. Moreover, like Casdagli's DVS model, this model can span the spectrum of models from global linear to local nonlinear, and as described in the next section, it can provide a simple test for nonlinearity in a time series (see also the elegant BDS criterion of Brock *et al.* (1987)).

Suppose we have an embedding of an observed time series $\boldsymbol{X}_t \in R^{m+1}$, where $\boldsymbol{X}_t(0) \equiv 1$ (for the constant term in the solution of (3.2) below). Let the time series value $T_\mathrm{p}$ time steps forward be $\boldsymbol{X}_{t+T_\mathrm{p}}(1) = Y_t$. Then the forecast at $T_\mathrm{p}$ is

$$\hat{Y}_t = \sum_{j=0}^{m} \boldsymbol{C}_t(j)\boldsymbol{X}_t(j). \tag{3.1}$$

For each predictee, $\boldsymbol{X}_t$, one finds the SVD solution for $\boldsymbol{C}$ by using historical points from the fitting set or library set (referred to below by the subscript '$i$') as follows:

$$\boldsymbol{B} = \boldsymbol{A}\boldsymbol{C}, \tag{3.2}$$

where

$$B_i = w(||X_i - X_t||)Y_i, \quad A_{ij} = w(||X_i - X_t||)\boldsymbol{X}_i(j)$$

and

$$w(d) = \mathrm{e}^{-\theta d/\bar{d}}, \tag{3.3}$$

$\theta \geqslant 0$, $d$ is the distance between the predictee and the neighbour vector, and the scale factor, $\bar{d}$, is the average distance between neighbours. Note that $\boldsymbol{A}$ has dimensions $n_i \times (m+1)$, where $n_i$ is size of the library. To maintain independence in the out-of-sample solution for each fitted map, I eliminate all vectors from the library whose coordinates include any of the coordinates of the predictee $X_t$ in the time series. Thus for each prediction, I construct a different exponentially weighted global linear map, where the degree of the local weighting is controlled by $\theta$. If $\theta = 0$, we have the simple global linear solution, as $\theta$ is tuned to higher positive values, the solutions become more local and hence nonlinear.

Notice that if instead of (3.3), the weighting function is $w(d) = 1$ for $d < k$ and $w(d) = 0$ otherwise, one recovers the local linear maps described by Casdagli (1992) and Diebold & Nason (1990). When there is a mixture of linear and nonlinear signal in the data (e.g. nonlinear response to regular forcing) this weighting function can penalize distant points too severely (zero weight); thus the exponential but global fit was chosen for this particular S-map.

### (*a*) *The difference test: Scripps diatoms*

S-maps can provide a useful way to characterize the degree of nonlinearity in a stationary time series in terms of the difference in forecast skill achieved with a nonlinear versus a linear model (see also the bi-spectrum of Subba Rao & Gabor (1984) and Subba Rao (1991) and the BDS test of Brock *et al.* (1987, 1991) and Brock & Potter (1993). In figure 1*f* it is applied to the Scripps diatom data, where the difference in linear versus nonlinear predictability is shown as a function of $\theta$. Comparisons of forecast skill achieved with nonlinear versus linear models shall be called 'difference tests'. As described in Sugihara & May (1990), one can ask whether the difference between linear and nonlinear models is significant simply by comparing the two correlation coefficients. This can be done by computing a $z$-statistic (Fisher's $z$-transform for the product moment correlation coefficient) to test the null hypothesis that $\rho_{\mathrm{nonlinear}} = \rho_{\mathrm{linear}}$. The availability of this standard test was one of the reasons that Sugihara & May (1990) used the correlation coefficient to measure predictability.

As can be seen, this difference is highly significant in the diatom data ($p < 0.0005$). This strong nonlinearity, combined with the steep prediction–decay curve (figure 1*c*, *d*) is evidence for the operation of chaotic dynamics. Notice further, how similar this DVS plot looks (figure 1*d*) to those obtained in figure 4*b* for stochastic chaos; both show a sharp decline in predictability as the global linear map is approached.

### (*b*) *The difference test: surrogate measles*

Figure 5 contains another example of a difference test as applied to the New York measles data (Schaffer 1985*a*, *b*; Olsen *et al.* 1988) and data from a null hypothesis (Stone 1992) which questioned the characterization of these data as an example of chaotic dynamics (Schaffer *et al.* 1988, 1990; Olsen *et al.* 1988; Sugihara & May 1990; Sugihara *et al.* 1990). This rather complicated null hypothesis incorporated much of the natural detail of the epidemic. Indeed there was some question as to its neutrality as a null hypothesis (Stone 1992). It was




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

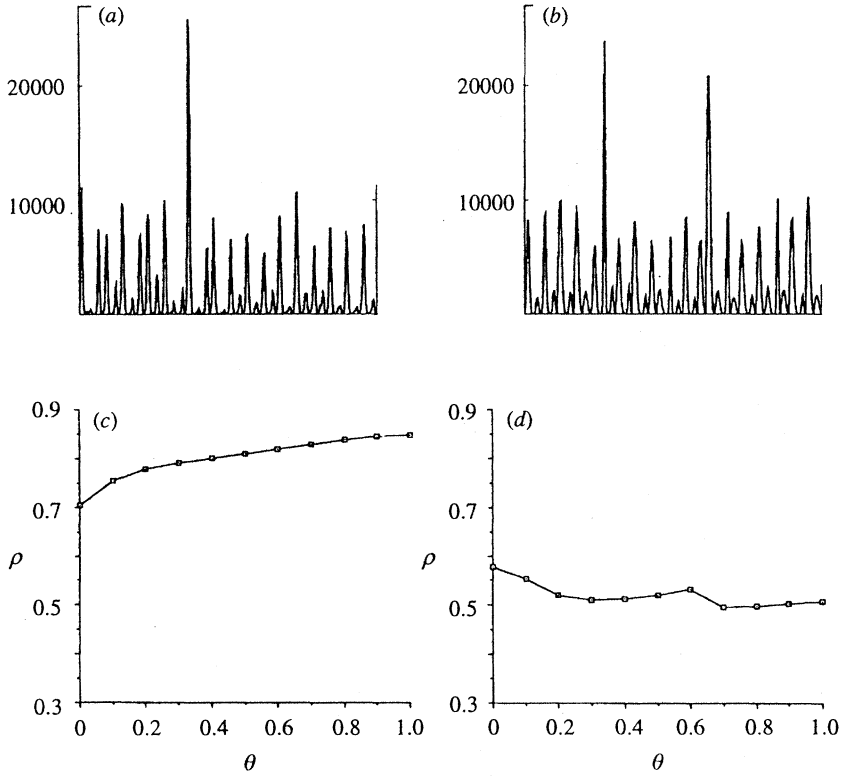


Figure 5. Test of Stone's surrogate null hypothesis for measles using a difference test and an S-map. (*a*) The real NY measles epidemic series. (*b*) A typical surrogate series constructed by stitching together empirically fit gaussian wave forms. (*c*) The difference plot for NY measles showing how strongly nonlinear the real data are ($p < 0.005$). (*c*) The difference plot for surrogate data is not nonlinear. When the correlation coefficients (S-map, $\theta = 1$) for surrogate measles and the real data are compared using a $z$-statistic, one can reject this null hypothesis ($p < 0.0005$).

constructed by stitching together gaussian wave forms taken from four empirically fit distributions sampled in a probabilistic order that would mimic the pattern observed. As seen in figure 5*b*, it very effectively mimics the general shape and features of the original measles data (figure 5*a*). These surrogates also had a degree of predictability which was similar to the real data (using a nonlinear forecast model) and which had similarly steep prediction–decay curves, yet were not from a chaotic process. This prompted Stone to accept the null hypothesis. However, as seen in figure 5*c*, *d*, they do not have the same apparent degree of nonlinearity as the natural data. One can safely reject this null hypothesis at the $p < 0.0001$ level. The difference between linear and nonlinear predictability is an important characteristic to consider when constructing null models to study chaos (see figure 2), and is useful in distinguishing chaos from coloured noise (Sugihara & May 1990).

## 4. Broader applications of nonlinear forecasting in medicine

The aim of this next section is to show how, apart from addressing the more narrow question of detecting chaos, nonlinear methods can have other important

applications in the biological sciences. In particular, it is shown how the key characteristics, nonlinearity and prediction–decay, can be useful attributes for the pragmatic classification of medical time series.

### (*a*) *Human heart rhythms and forecasting for practical diagnosis*

An area of application where nonlinear methods and forecasting will likely have a great deal to offer is in the study of human heart rhythms and in particular the characterization and diagnosis of pathologies. This is an area where there has been some very strong theoretical work (e.g. Glass & Mackey 1988), however, less has been done in the way of applying modern time series methods to analyse empirical EKG data (Lefebvre *et al.* 1993; Lipshitz & Goldberger 1992; Gough 1992; Skinner *et al.* 1990; Goldberger *et al.* 1990, 1984).

Two important results from the empirical analysis of human heart rhythms have to do with differences in RR-intervals as a function of age, and those differences in heart rhythms that may be associated with heart disease. With regard to age, Kaplan *et al.* (1991) found that young healthy hearts contain a high degree of variability (measured in terms of standard deviation) which becomes diminished as a normal part of the aging process. Lipshitz & Goldberger (1992) found a decrease in the correlation dimension with age. As regards health, Kleiger *et al.* (1987), showed how after an acute myocardial infarction, low heart rate variability (low variance in RR-intervals) is associated with increased mortality. Lefebvre *et al.* (1993) found significant differences in predictability with nonlinear forecast methods, between older patients who had suffered a heart attack and older patients who had not. They also found suggestive differences (though not significant) in the rate of prediction–decay of diseased versus healthy patients.

Here, I shall apply the S-map given in (3.1) to EKG data from 29 patients kindly provided by J. Lefebvre, M. Kamath, D. Goodings and M. Fallen (see Lefebvre *et al.* (1993) for details). The aim of this exercise is to determine whether the degree of nonlinearity (difference test) and the rate of prediction–decay can discriminate age and health. I shall do this in two ways: by modelling each patient separately and by creating composite mappings for each group.

In all cases, I shall use first-differenced data embedded with a standard lag of 1 and an embedding dimension $m = 7$. This embedding convention was used by Lefebvre *et al.* (1993) with the simplex predictor (Sugihara & May 1990), and $m = 7$ is in the plateau of optimal predictability for all cases. Moreover, as Lefebvre *et al.* (1993) noted, a vector having seven components is just long enough to span a complete respiratory cycle with respiration likely being an important controlling factor.

#### (i) *Modelling individual attractors*

Figure 6 shows illustrative examples of prediction–decay curves and difference plots constructed by comparing changes in absolute forecast error for a young versus an old healthy patient (6*c*, *d*) and for a healthy versus a heart-diseased patient (6*a*, *b*). For reasons which will be explained in the next section, absolute forecast error is a better discriminator of heart disease than the correlation coefficient. Table 1 presents the group means and the significance levels obtained for these two comparisons. As can be seen, both nonlinearity (as measured by the difference in absolute forecast error) and decay are highly significant discriminators of heart disease, however, only nonlinearity is a good discriminator of age.





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

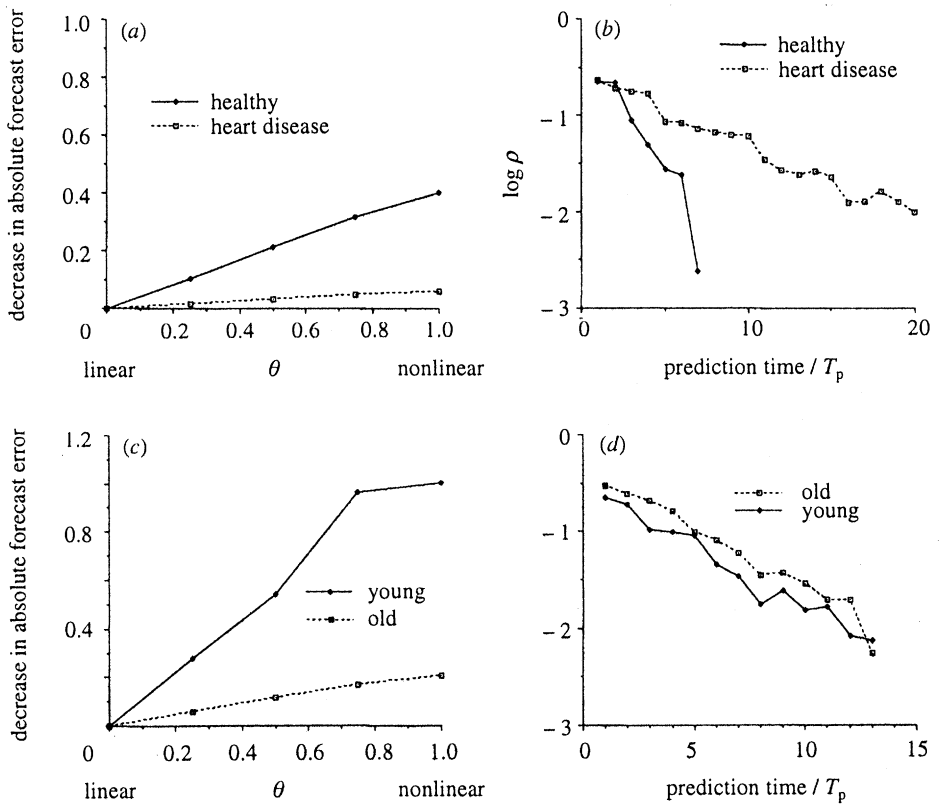


Figure 6. Illustrative example of the classification of human heart rates by age and health using nonlinearity (decrease in absolute forecast error) and prediction–decay slope. (*a*) Difference plot showing how this healthy patient (patient OH1 in Lefebvre *et al.* 1993) has a greater decrease in absolute forecast error with increasing nonlinearity (hence, is more nonlinear) than a patient who has heart disease (patient NB1). (*b*) Prediction–decay curves for the same patients, showing how healthy patients have steeper prediction–decay curves than heart-diseased patients. (*c*) Difference plot showing how a young healthy patient (patient YH1 in Lefebvre *et al.* 1993) has a greater decrease in absolute forecast error with increasing nonlinearity (hence, is more nonlinear) than an older healthy patient (patient OH3). (*d*) Prediction–decay curves for the same subjects, showing how prediction–decay slope does not distinguish age (see table 1 for ensemble statistics).

The more rapid prediction decay in healthy versus diseased patients is consistent with the suggestive result reported by Lefebvre *et al.* (1993).

Figure 7 shows a weak positive relationship between nonlinearity and decay. Note that this relationship only becomes significant when the axes are log transformed. Although the evidence is not compelling in this specific example, such 'difference-decay' plots may be a useful tool when exploring for chaos in real data. With chaotic dynamics one might expect a tendency toward a positive relationship between nonlinearity and decay.

(ii) *Modelling composite attractors*

When individual time series are short, one can imagine constructing a composite attractor by piecing together segments from individuals (samples) belonging to a particular homogeneous group. Thus one could construct composite maps that are intended to capture the dynamical mean of a group. This might be most

Table 1. *EKG study – summary of results for individual maps: mean values and significance levels*

(Mean values and significance levels (t-test) for the difference in absolute forecast error realized with a nonlinear model ($\theta = 1$) against a global linear model ($\theta = 0$), and for prediction–decay slopes.)

| | nonlinearity | | pred.–decay | |
|---|---|---|---|---|
| | diff. (error) | $p$ | slope | $p$ |
| healthy $n = 14$ | 1.065 | $p < 0.005$ | 0.163 | $p < 0.005$ |
| disease $n = 15$ | 0.314 | | 0.102 | |
| young $n = 6$ | 1.670 | $p < 0.005$ | 0.162 | $p < 0.4$ (n.s.) |
| old $n = 8$ | 0.609 | | 0.164 | |

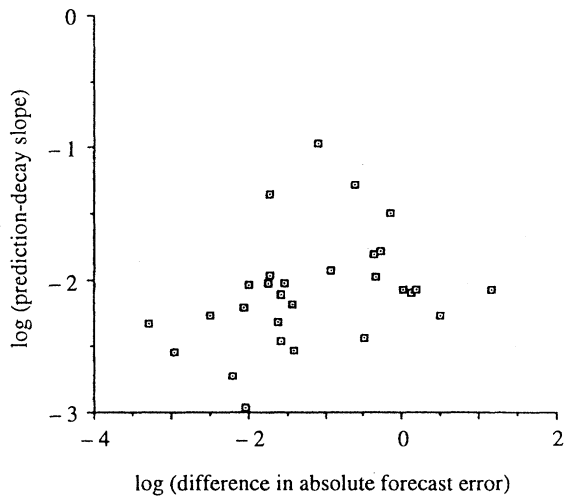


Figure 7. Difference–decay plot showing a weak relationship ($\rho = 0.333$, $0.25 < p < 0.05$) between prediction–decay slope and nonlinearity for the 29 patients taken as a group. Note that the axes are log-transformed.

useful when individual time series are short or for data-intensive calculations such as measuring the correlation dimension.

Here I have created composite attractors for each of the four groups. Heteroscedasticity is minimized by normalizing the means (= 0) and variances (= 1) for each group before concatenating the time series. To avoid the problems created by the discontinuities at the seams between time series, I eliminate $m \times \tau + \max(T_p)$ time series values before and after the joints from all calculations

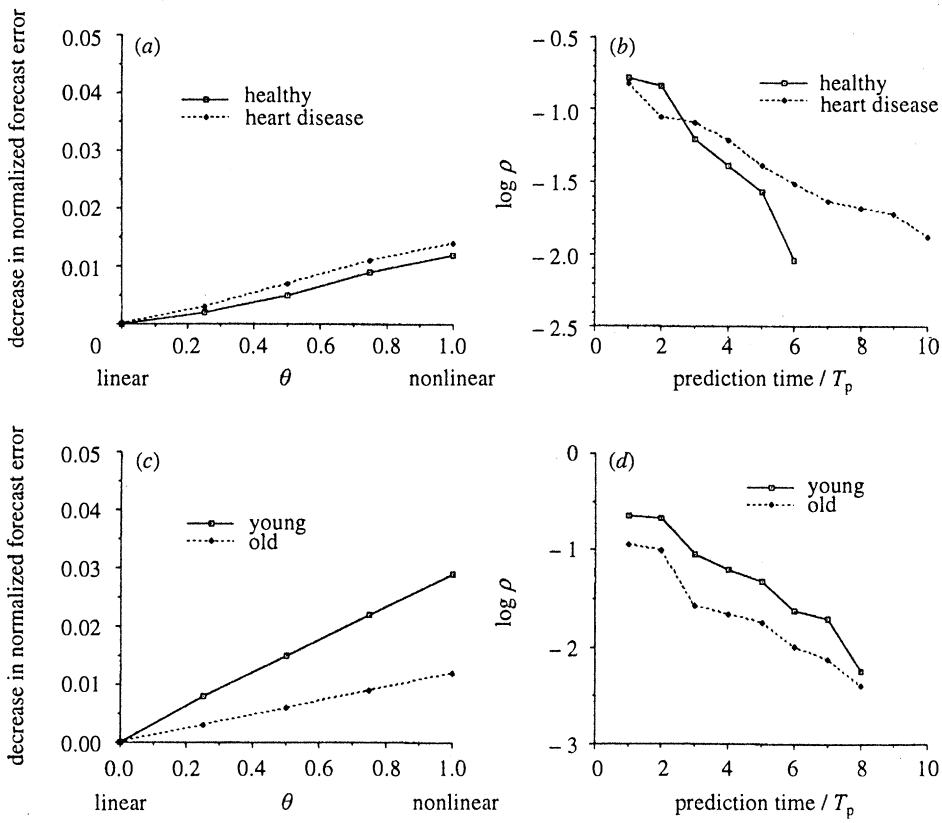


Figure 8. Use of normalized composite maps for the classification of human heart rates (see table 2 for summary statistics).

Table 2. *EKG study – summary of results from normalized composite maps* ($N = 5000$ *in all cases*)

| | nonlinearity | | predictability | | pred.–decay | |
|---|---|---|---|---|---|---|
| | diff. (error) | $p$ | $\rho$ | $p$ | slope | $p$ |
| healthy | 0.015 | $p < 0.2$ (n.s.) | 0.420 | $p < 0.0005$ | 0.309 | $p < 0.025$ |
| disease | 0.016 | | 0.351 | | 0.111 | |
| young | 0.023 | $p < 0.0001$ | 0.522 | $p < 0.0001$ | 0.266 | $p < 0.5$ (n.s.) |
| old | 0.011 | | 0.378 | | 0.260 | |

(libraries and predictions). To speed calculations, and to normalize comparisons, each composite S-map was constructed from 5000 randomly sampled vectors.

Figure 8 shows the prediction–decay curves and the difference plots obtained by comparing changes in absolute forecast error for the young attractor versus the old healthy map (8*c*, *d*) and for the healthy versus heart-diseased map (8*a*, *b*).




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

Table 2 summarizes the comparisons. One should take special note that by normalizing the variances of all groups we have effectively factored out variance as a discriminating variable. Thus, any effect that is seen here is over and above the effect of variance in classifying heart rhythms. As with the individual maps, with the composite maps, prediction decay slope is not a significant discriminator of age but is a significant discriminator of health ($p < 0.01$). The healthy attractor had a more precipitous decay (reaching the noise floor at $T_p = 6$ beats, roughly the time for one breathing cycle) than the diseased attractor (hit the noise floor at $T_p = 11$ beats). As with the individual maps, nonlinearity in the normalized composite maps was a good discriminator of age however, unlike the individual maps it was not a good discriminator of health. This difference is due to the normalization of variances in the composite maps. Just as mean and variance scale together, improvements in absolute forecast error seen in the individual case should scale with the magnitude of the excursions (standard deviation) in the time series. That is, where the absolute magnitude of the excursions is large, the absolute magnitude of the improvement in predictability is large. Therefore, in terms of diagnosing heart-disease, nonlinearity per se does not contain significantly more information than is already contained in the variance.

As shown in table 2, predictability with the nonlinear model ($\theta = 1$) is much higher in the normalized map for the young subjects than for the old subjects. It is interesting that the heart rates for the young subjects which appeared more variable, are in fact more predictable when the data are normalized. This is likely due to there being more homogeneity in the young group than in the other groups (Sugihara 1994). The strong trends in predictablity shown here are not found in the unnormalized data (Lefebvre *et al.* 1993), where differences in the variances of the RR-intervals will confound this result.

To summarize, these preliminary results show that over and above variance, prediction–decay is a good discriminator of health; healthy patients have a much steeper decay curve than those with heart disease. And nonlinearity is a good discriminator of age; younger adults have more nonlinearity to their heart rhythms than older ones. In addition, nonlinear predictability appears to be a good classifier of age in the normalized maps.

### (*b*) *Infant heart rhythms*

As a suggestive element to add to the results above, I shall briefly mention work in progress with W. Allan and D. Sobel on the development of heart rhythms in infants that shows a very different pattern.

Allan & Sobel (1992) hypothesize that because other autonomic nervous system functions such as pupilary reflex seem to mature at about 30 weeks, that somewhere between 25 and 35 weeks sympathetic and parasympathetic balance evolves and contributes to irregularity in heart rate.

Figure 9 shows an illustrative example of this possibility for an infant aged 25 weeks and one aged 40 weeks. Here the S-map was applied to heart rate data taken by chest leads, $n = 1000$. I emphasize that these are very preliminary results. Although one cannot generalize from a sample of 1, it is intriguing nonetheless to hypothesize that heart rhythms might follow an ontogenetic sequence, starting out simple with high short-term linear autocorrelation, maturing to nonlinearity at birth; then from young adulthood onwards simplifying with age to a signal that is once again more strongly dominated by linear characteristics.




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

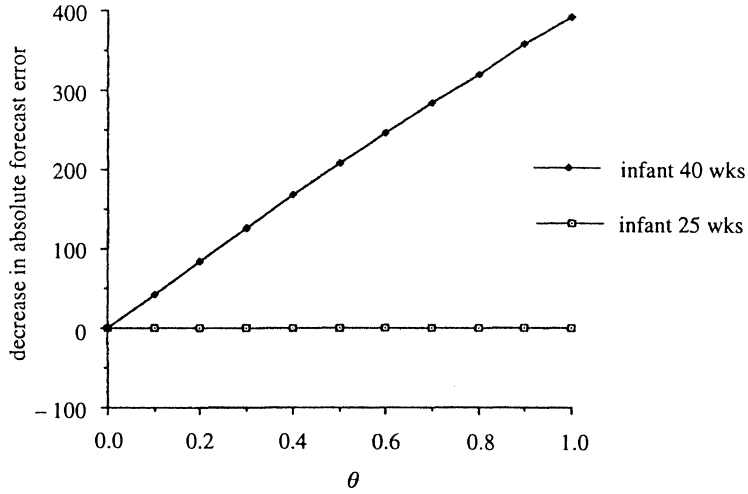


Figure 9. An anecdotal comparison of heart rate nonlinearity in an infant 25 weeks and an infant 40 weeks. The younger infant did not have the nonlinearity that was present in the older infant.

## 5. Summary

In summary, despite concerns over the feasibility of a rigorous test for chaos in real data (from the calculation of meaningful Lyapunov exponents) there are methods available that allow one to gather qualitative evidence for the operation of such dynamics. Moreover, the narrow question of chaos aside, as I have tried to suggest with the EKG study, there are useful and important ways that nonlinear methods can be applied in the natural sciences.

I thank Martin Casdagli, Jim Crutchfield, Dennis Grace, Cleridy Lenhert, Richard Penner and John Sidorowich for helpful and challenging discussions. I am indebted to Julie Lefebvre, Mark Kamath and D. G. Goodings for their kindness and generosity in providing me with their data on adult heart rates. I also thank Walter Allan and Daniel Sobel for letting me use some of their data on infant heart rates for figure 9. Finally, I to thank Howell Tong and Bryan Grenfell for their thorough and insightful comments on the final manuscript. This work was funded in part by ONR:N00014-92-J-4068, and endowment income from the John Dove Issacs Chair in Natural Philosophy.

## References

Abarbanel, H. D. I., Brown, R., Sidorowich, J. J. & Tsimrin, L. Sh. 1993 The analysis of observed chaotic data in physical systems. *Rev. mod. Phys.* **64**, 1331–1392.

Allan, W. C. & Sobel, D. B. 1992 The sans of time. *J. Am. med. Ass.* **268**, 984.

Ascioti, F. A., Beltrami, E., Caroll, T. O. & Wirick, C. 1993 Is there chaos in plankton dynamics? *J. Plank. Res.* **15**(6), 603–617.

Bottou, L. 1993 Local learning for optical character recognition. *Neural Information Processing Systems* **5**.

Brock, W. 1986 Distinguishing random and deterministic systems, Abridged version. *J. Econ. Theory* **40**, 168–195.

Brock, W. & Potter, S. M. 1993 Nonlinear time series and macroeconometrics. In *Handbook of statistics* (ed. G. S. Maddala, C. R. Rao & H. D. Vinod), vol. 11, pp. 195–229. Elsevier.

Brock, W., Hsieh, D. & LeBaron, B. 1991 *Nonlinear dynamics, chaos, and instability, statistical theory and economic evidence.* MIT Press.





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

Brock, W., Dechert, W. & Scheinkman, J. 1987 A test for independence based upon the correlation dimension. Department of Economics, University of Wisconsin, University of Houston, and the University of Chicago.

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

Casdagli, M. 1991 A dynamical systems approach to modeling input–output systems. In *Nonlinear modeling and forecasting*. Proceedings Santa Fe Institute Studies in the Sciences of Complexity (ed. M. Casdagli & S. Eubank), vol. 12, pp. 265–281. New York: Addison-Wesley.

Casdagli, M. 1992 Chaos and deterministic versus stochastic modeling. *Jl R. statist. Soc.* B, pp. 303–328.

Casdagli, M., Eubank, S., Farmer, J. D. & Gibson, J. 1991 State space reconstruction in the presence of noise. *Physica* D **51**, 52–98.

Casdagli, M., Jardins, D. D., Eubank, S., Farmer, J. D., Gibson, J., Hunter, N. & Theiler, J. 1992 Nonlinear modeling of chaotic time series, theory and applications. Los Alamos National Laboratory Technical Report LA-UR-91-1637.

Chan, K. S. & Tong, H. 1994 A note on noisy chaos. *Jl R. statist. Soc.* B **56**, 301–311.

Cleveland, W. S. 1979 Robust locally weighted regression and smoothing scatterplots. *J. Am. statist. Ass.* **74**, 829–836.

Crutchfield, J. P. 1979 Prediction and stability in classical mechanics. Bachelor's Thesis, University of California, Santa Cruz.

Crutchfield, J. P. & MacNamara, B. S. 1987 Equations of motion from a data series. *Complex Systems* **1**, 417–452.

Diebold, F. X. & Nason, J. A. 1990 Nonparametric exchange rate prediction? *J. Int. Economics* **28**, 312–332.

Ellner, S. 1991 Detecting low-dimensional chaos in population dynamics data, a critical review. In *Chaos and insect ecology* (ed. J. A. Logan & F. P. Hain), pp. 65–92. Blacksburg, Virginia: Virginia Agricultural Experiment Station, Virginia Polytechnic Institute and State University.

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

Gershenfeld, N. A. & Weigend, A. S. 1993 The future of time series, learning and understanding. In *Time series prediction, forecasting the future and understanding the past* (ed. A. S. Weigend & N. A. Gershenfeld), pp. 1–70. Reading, MA: Addison–Wesley.

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

Goldberger, A. L., Rigney, D. R. & West, B. J. 1990 Chaos and fractals in human physiology. *Sci. Am.* **262**, 43–49.

Goldberger, A. L., Findley, L. J., Blackburn, M. R. & Mandell, A. J. 1984 Nonlinear dynamics in heart failure, implications of long-wavelength cardiopulmonary oscillations. *Am. Heart J.* **107**, 612–615.

Gough, N. A. J. 1992 Fractals, chaos and fetal heart rate. *Lancet* **339**, 182–183.

Hardle, W. 1990 *Applied nonparametric regression* Cambridge University Press.

Hastings, A., Hom, C. L., Ellner, S., Turchin, P. & Godfrey, C. 1993 Chaos in ecology, is mother nature a strange attractor? *A. Rev. Ecol. Syst.* **24**, 1–33.

Kaplan, D. T., Furman, M. I., Pincus, S. M., Ryan, S. M., Lipshitz, L. A. & Goldberger, A. L. 1991 Aging and the complexity of cardiovascular dynamics. *Biophys. J.* **59**, 945–949.

Kleiger, R. E., Miller, J. P., Bigger, J. T., Moss, A. J. & the Multicenter Post-Infarction Research Group 1987 Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am. J. Cardiol.* **59**, 256–262.

Kravtsov, Y. A. & Etkin, V. S. 1984 Dynamic correlation degree and the problem of determining the dynamic nature of random processes . *Radiotekhnika i electrotechnika* **12**, 2358–2364.

Lefebvre, J. H., Goodings, D. A., Kamath, M. V. & Fallen, E. L. 1993 Predictability of normal heart rhythms and deterministic chaos. *Chaos* **3**, 267–276.

Lipshitz, L. A. & Goldberger, A. L. 1992 Loss of 'complexity' and aging, potential applications





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

of fractals and chaos theory to senescence. *J. Am. med. Ass.* **267**, 1806–1809.

Nychka, D., Ellner, S., Gallant, A. R. & McCaffrey, D. 1992 Finding chaos in noisy systems. *Jl R. statist. Soc.* B **54**, 399–426.

Olsen, L. F., Truty, G. L. & Schaffer, W. M. 1988 Oscillation and chaos in epidemics, a nonlinear dynamic study of six childhood diseases in Copenhagen, Denmark. *Theor. Pop. Biol.* **33**, 344–70.

Osborne, A. R. & Provenzale, A. 1989 Finite correlation dimension for stochastic systems with power-law spectra. *Physica* D **35**, 357–381.

Rand, D. A. & Wilson, H. B. 1991 Chaotic stochasticity, a ubiquitous source of unpredictability in epidemics. *Proc. R. Soc. Lond.* B **246**, 179–184.

Schaffer, W. M. 1985*a* Can nonlinear dynamics elucidate mechanisms in ecology and epidemiology? *IMA J. Math. appl. Med. Biol.* **2**, 221–252.

Schaffer, W. M. 1985*b* Order and chaos in ecological systems. *Ecology* **66**, 93–106.

Schaffer, W. M., Olsen, L. F., Truty, G. L., Fulmer, S. L. & Graser, D. J. 1988 Periodic and chaotic dynamics in childhood infections. In *From chemical to biological organization* (ed. M. Markus, S. C. Muller & G. Nicolis), pp. 331–347. New York: Springer-Verlag.

Schaffer, W. M., Olsen, L. F., Truty, G. L. & Fulmer, S. L. 1990 The case for chaos in childhood epidemics. In *The ubiquity of chaos* (ed. S. Krasner), pp. 138–166. Washington, D.C.: AAAS.

Sidorowich, J. J. 1992 Repellors attract attention. *Nature, Lond.* **355**, 584.

Skinner, J. E., Goldberger, A. L., Mayer-Kress, G. & Ideker, R. E. 1990 Chaos in the heart, implications for clinical cardiology. *BioTechnology* **8**, 1018–1024.

Stokbro, K. & Umberger, D. 1992 Forecasting with weighted maps. In *Nonlinear modeling and forecasting* (ed. M. Casdagli & S. Eubank). Redwood City: Addison-Wesley.

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

Subba Rao, T. & Gabor, M. N. 1984 An introduction to bispectral analysis and bilinear time series models. In *Lecture notes in statistics.* New York: Springer-Verlag.

Subba Rao, T. 1991 Analysis of nonlinear time series (and chaos) by bispectral methods. In *Nonlinear modeling and forecasting* (ed. M. Casdagli & S. Eubank). Reading, MA: Addison-Wesley.

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

Sugihara, G., Grenfell, B. & May, R. 1990 Distinguishing error from chaos in ecological time series. *Phil. Trans. R. Soc. Lond.* B **330**, 235–251.

Theiler, J., Galdrikian, B., Longtin, A., Eubank, S. & Farmer, J. D. 1991 Using surrogate data to detect nonlinearity in time series. In *Nonlinear modeling and forecasting* (ed. M. Casdagli & S. Eubank). Reading, MA: Addison-Wesley.

Tong, H. 1990 *Nonlinear time series analysis, a dynamical systems approach.* Oxford University Press.

Tont, S. A. 1981 *J. mar. Res.* **39**, 191–201.

Tsonis, A. A. & Elsnor, J. B. 1992 Nonlinear prediction as a way of distinguishing chaos from random fractal sequences. *Nature, Lond.* **358**, 217–220.

Wales, D. J. 1991 Calculating the rate of loss of information from chaotic time series by forecasting. *Nature, Lond.* **350**, 485–488.





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