The Danger of Wishing for Chaos

Patrick E McSharry, *Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK; Mathematical Institute, University of Oxford, Oxford OX1 3LB, UK; Centre for the Analysis of Time Series, London School of Economics, London, WC2A 2AE, UK*

THE DANGER OF WISHING FOR CHAOS

Correspondence should be addressed to: Patrick E McSharry, Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK, Patrick@mcsharry.net, Tel: +44 1865 273095, Fax: +44 1865 273905

Abstract:

With the discovery of chaos came the hope of finding simple models that would be capable of explaining complex phenomena. Numerous papers claimed to find low-dimensional chaos in a number of areas ranging from the brain to the stockmarket. Years later, many of these claims have been disproved and the fantastic hopes pinned on chaos have been toned down as research with more realistic objectives follows. The difficulty in calculating reliable estimates of the correlation dimension and the maximal Lyapunov exponent, two of the hallmarks of chaos, are explored. Given that nonlinear dynamics is a relatively new and growing field of science, the need for statistical testing is greater than ever. Surrogate data provides one possible approach but great care is needed in generating relevant surrogates and in interpreting the results. Examples of misleading applications and challenges for the future of research in nonlinear dynamics are discussed.

KEYWORDS: Chaos, nonlinear dynamics, correlation dimension, Lyapunov exponent, surrogates

INTRODUCTION

The word chaos has become synonymous with complicated systems that are beyond the realm of understanding. At worst, chaos describes a state of complete disorder and confusion. To a scientist, chaos presents a less pessimistic view of life: behaviour so unpredictable as to appear random because of the inherent sensitivity to small perturbations in the initial conditions. This suggests that many complex systems could possibly be described by low-dimensional deterministic mathematical models. It was the discovery of this intermediate possibility between order and disorder that brought so much hope and excitement in the 1980s, culminating with popular science books such as “Chaos” (Gleick, 1988). While most real-world systems are undoubtedly nonlinear, the quality of data available often favours traditional linear analyses. In practice, a modeling framework that incorporates a mixture of deterministic and stochastic approaches may have the greatest utility.

The tools of the nonlinear dynamicist borrow from a diverse range of disciplines including physics, mathematics, statistics and engineering. While these traditional scientific disciplines offer a strong framework, this new field of research requires the formation of many new methods and techniques. It is interesting to note a division of research at this point: method-driven and application-driven. Firstly, there are those who have focused on the construction and testing of new methods for carrying out “nonlinear time series analysis” - the name given to the investigation of systems and models that may display chaos (Kantz & Schreiber, 2003). On the other hand, there are many researchers interested in exploring whether the ideas and concepts from nonlinear dynamics can provide additional insight into their particular fields of expertise.

This division has and continues to present an unfortunate danger for the credibility of nonlinear time series analysis. The proposition of new nonlinear methods without suitable statistical testing and the subsequent adoption of these methods by practitioners could destroy the integrity of all research, which uses nonlinear techniques. Falling into a state of disrepute is likely to cause a setback to the scientific progress of nonlinear dynamics.

In this paper we will argue that it is extremely difficult to show that any real system is chaotic because of the requirement of collecting a large number of high quality observations from a period when the underlying dynamics are stationary. For this reason, all estimates of system invariants should be accompanied by an expression of the level uncertainty associated with these estimates. Nevertheless, the techniques arising from the field of nonlinear time series analysis can make a substantial contribution to science. We shall look at some of the tools available to those researchers who believe that nonlinear methods may offer a more in-depth understanding of their particular systems of interest. We will focus on the mathematical sanity checks afforded by statistical tests that can help to avoid drawing misleading conclusions. Specifically, we will investigate (i) the difficulty in identifying low-dimension chaos using empirical data, (ii) the challenges in constructing, estimating and evaluating nonlinear models, and (iii) the need for statistical tests that are relevant to the specific application being proposed.

The layout of the paper is as follows. We first review the mathematical concepts underpinning a number of nonlinear methods and the definitions of system invariants such as the correlation dimension and maximal Lyapunov exponent. Then we discuss why it is so difficult to estimate these invariants from empirical data and to conclude that any real system is actually chaotic.

Next we provide some guidelines for justifying the use of nonlinear methods and explore the use of surrogate data for statistical testing. In the penultimate section we discuss some of the challenges facing the nonlinear dynamics research community when proposing the application of new methods and give examples from the field of biomedical research. The final section provides a discussion of approaches for promoting scientific progress in the field of nonlinear dynamics and concludes the papers.

MATHEMATICAL TOOLS

The usual starting point for any nonlinear time series analysis is an attempt at reconstructing the underlying state space, which in thoery provides a unique description of the state of the system. For univariate time series, s_n ($n=1,\ldots,N$), a set of m-dimensional state vectors may be defined using a delay coordinate reconstruction:

$$x_n = (s_{n-(m-1)\tau}, s_{n-(m-2)\tau}, \ldots, s_{n-\tau}, s_n),$$

where τ is known as the time delay or time lag. The mathematical theory of embedding tells us that under certain conditions, it should be possible to use this reconstruction as a faithful representation of the underlying dynamics if m satisfies $m > 2D_F$, where D_F is the box counting dimension (Takens, 1981; Sauer et al., 1991). Unfortunately, this mathematical theory tells us little about what to do when we are faced with noisy data of finite duration. For example, while the choice of τ is irrelevant in the theory, it is extremely important in practice. Suggested values for τ are often based on mutual information (Fraser, 1986) or a geometric interpretation of the reconstruction (Rosenstein et al., 1994). Similarly, the fact that we are unlikely to know the value of D_F beforehand means that we also have to estimate a reasonable value for m from the data. The method of *false nearest neighbours* provides one possibility for estimating a sufficient value for m (Kennel et al., 1992). This is typically achieved by varying the size of m and monitoring the number of false nearest neighbours associated with areas of the reconstructed state space that have self-intersections. Unfortunately, the detection of false nearest neighbours is subject to the choice of an arbitrary constant, which will vary with position in state space.   By testing for consistency between the dynamics imposed by a model and the observational uncertainty while allowing for variation in the local instabilities of the nonlinear dynamics throughout state space, it is possible to determine a robust estimate for the minimum value for m (McSharry & Smith, 2004).

Fractal dimensions

The spate of papers aiming to measure the fractal dimensions of different systems was perhaps motivated by the hope of determining the number of active degrees of freedom. The ability to demonstrate that a particular dataset was generated by a low-dimensional deterministic process implies that, in theory, there exists a model with a few degrees of freedom which might be able to represent the dynamics.   There are many examples of misleading estimates for the dimension of complicated real-world systems. For example, the weather was reported as having a dimension between 3 and 8; see (Lorenz, 1991) for a critique. Similarly in the case of stock market returns, evidence of nonlinearities was found but claims of low-dimensional chaos were not well-justified (Scheinkman & Le Baron, 1989).

Although there is an entire family of dimensions, known as Renyi dimensions, the correlation dimension, D_2, is the easiest to calculate from data (Kantz & Schreiber, 2003). D_2 reflects how the probability that the distance between two randomly chosen points will be less than ε, scales as a function of ε. The correlation integral, which counts the fraction of pairs ($\boldsymbol{x}_i$, $\boldsymbol{x}_j$) whose distance is smaller than ε, is defined by

$$C(\varepsilon, N) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \theta(\varepsilon - \| x_i - x_j \|)$$

where θ is the Heaviside step function, $\theta(x)$=0 if $x \leq 0$ and $\theta(x)$=1 if $x > 0$ (Grassberger and Procaccia, 1983). C is expected to scale like a power law, $C \propto \varepsilon^{D_2}$, for an infinite amount of data where $N \to \infty$ and $\varepsilon \to 0$, and the correlation dimension D_2 is defined by

$$D_2 = \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{d \ln C(\varepsilon, N)}{d \ln \varepsilon}$$

There are two problems when attempting to identify a scaling region: (i) the finite sample size places an upper limit on N, and (ii) the finite accuracy of the data and the sparseness of near neighbours when ε is small.

Unfortunately, even if it were possible to overcome the obstacles above, it is likely that temporal correlations in the data will give rise to underestimates of D_2. It has been shown that infinite dimensional stochastic signals can lead to finite and low-dimensional estimates (Theiler, 1986). The estimates for C can be biased by temporal correlations since small spatial separations between pairs of points can occur because they were observed closely in time. For finite data sets, this effect of points nearby in time can be limited by restricting the sums over i and j so that $|i-j| > W$ for some constant W (Theiler, 1986). Even when i is much greater than j, the distance between $\boldsymbol{x}_{i+1}$ and $\boldsymbol{x}_{j+1}$ is unlikely to be independent of the distance separating $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ leading to further biases in the estimate. While we can never be sure whether there is a sufficient amount of data, a space-time-separation diagram may be employed to ascertain with confidence that we do not have enough data (Provenzale et al., 1992).

Lyapunov exponents

The most familiar signature of chaos is displayed by the unpredictability of the future despite the system obeying deterministic equations of motion. This unpredictability appears as an increasing average forecast error with larger prediction lead times. The phrase *sensitive dependence on initial conditions* is often used to describe the inherent instability of the solutions that cause this unpredictability. Specifically, two nearby initial conditions will, on average, diverge over time. While many linear systems give rise to a slow rate of divergence, it is the exponential divergence demonstrated by some nonlinear systems, that is characteristic of chaotic systems.

Consider a deterministic dynamical system described by the discrete map, $\boldsymbol{x}_{n+1} = F(\boldsymbol{x}_n)$. The evolution of an infinitesimal uncertainty, ε_0, in the initial condition, $\boldsymbol{x}_0$, over a finite number of time steps, k, is given by

$$\varepsilon_K = \boldsymbol{M}(\boldsymbol{x}_0,k)\ \varepsilon_0,$$

where $\boldsymbol{M}(\boldsymbol{x}_0,k)$ is the linear tangent propagator formed by the product of the Jacobians along the k steps of the trajectory $\boldsymbol{M}(\boldsymbol{x}_0,k) = \boldsymbol{J}(\boldsymbol{x}_{k-1})\ \boldsymbol{J}(\boldsymbol{x}_{k-2}) \ldots \boldsymbol{J}(\boldsymbol{x}_0)$.

The linear dynamics of uncertainty growth may be analysed using the singular value decomposition (SVD) of $\boldsymbol{M}$, giving $\boldsymbol{M} = \boldsymbol{U\Sigma V}^T$, where the columns of the orthogonal matrix $\boldsymbol{U}(\boldsymbol{V})$ are the left(right) singular vectors respectively and the entries of the diagonal matrix $\boldsymbol{\Sigma}$ are the singular values, $\sigma_i(\boldsymbol{x}_0,k)$, usually ranked in decreasing order (Strang, 1988). From a geometrical point of view, the linear dynamics have the effect of transforming a circle into an ellipse, where the right singular vectors describe the directions of the axes of the ellipse and the singular values reflect the lengths of these axes. Finite-time Lyapunov exponents, which depend both on the initial position, $\boldsymbol{x}_0$, and the number of time steps, k, are defined by

$$\lambda_i^{(k)}(\boldsymbol{x}_0) = \frac{1}{k}\log_2 \sigma_i(\boldsymbol{x}_0, k)\,.$$

The system invariants known as the Lyapunov exponents are then defined by taking the limit as k goes to infinity,

$$\Lambda_i = \lim_{k\to\infty} \lambda_i^{(k)}(\boldsymbol{x}_0)\,.$$

A system is said to be chaotic if the leading Lyapunov exponent, Λ_1, is positive. Alternatively, a negative Λ_1 indicates the existence of a stable fixed point.

While the expression for $\lambda_i^{(k)}(\boldsymbol{x}_0)$ may be rewritten as $\sigma_i(\boldsymbol{x}_0, k) \propto e^{k\lambda_i^{(k)}(\boldsymbol{x}_0)\ln 2}$, the effective growth rate defined by a positive $\lambda_i^{(k)}(\boldsymbol{x}_0)$ does not imply exponential growth; it is only in the case of Λ_i, that a positive value implies an effective exponential growth (Ziehmann et al., 1999). There is a common misconception that the value of Λ_1 provides a measure of the unpredictability. This is untrue since Λ_1 only applies to infinitesimal uncertainties, and as such an infinitesimal uncertainty can never limit predictability. Furthermore, Λ_1 is an average defined over an infinitely long trajectory and does not reflect short-term dynamics. For some chaotic systems, it can be shown that in the short term all uncertainties may shrink for a finite time (Smith et al., 1999).

ESTIMATES FROM DATA

Unfortunately, the attractiveness of finding chaos in real-world systems has the potential to cloud the better judgement of many researchers. There are three fundamental problems with seeking chaos in such systems. Firstly, the existence of chaos can only be mathematically proven for a handful of simple toy models, such as the logistic map (May, 1976). Secondly, our perspective of the underlying dynamics of the system is obtained from the time series of observations or data that we collect using some measurement process. The quality of such data sets is limited by the observational uncertainty that arises from measurement errors, noise, artefacts and missing values. Thirdly, the underlying dynamical processes may not be stationary implying that no single mathematical model is capable of describing the dynamics throughout the entire duration of the experiment used to record the data. The combination of these problems implies that we are unlikely to obtain data of a high enough quality (resolution and duration) from a stationary epoch of the underlying dynamics to enable us to calculate the correlation dimension, D_2, or the maximal Lyapunov exponent Λ_1. In the following sections, we investigate the difficulties associated with estimating these two system invariants when faced with noisy time series of short duration. We also suggest that despite these estimation difficulties, the resulting nonlinear quantities may still be of value.

Estimating the correlation dimension

When computing D_2 it is important to know what type of error should be expected for a given quantity of data, N. Although it can be shown both numerically and analytically that this error generically scales as $O(1/N^{1/2})$ for $N \to \infty$, this is not always the case (Theiler, 1992). While the quantity of data required to reliably estimate D_2 for a particular system has been explored by a number of investigators, giving rise to a range of differing opinions, it suffices to say that the results are pessimistic for the majority of interesting real-world systems. Various claims for the minimum number of data points, N_{min}, include $N_{\min} = D_2^{42}$ (Smith, 1989), $N_{\min} = 2^{D_2}(D_2+1)^{D_2}$ (Nerenberg and Essex, 1990), $N_{\min} = 10^{\frac{D_2}{2}}$ (Ruelle, 1990), $N_{\min} = 10^{2+\frac{2}{5}D_2}$ (Tsonis, 1992) and $N_{\min} = 10^{1+\frac{D_2}{2}}$ (Kantz & Schreiber, 2003). The variation of N_{min} with D_2 suggested by each of these estimates is illustrated in Fig. 1. The best case scenario implies that N=100,000 points are necessary for estimating D_2 =10. In short, the news is bad if not completely foreboding for any researcher hoping to estimate D_2 from short noisy time series.

Insert Figure 1 About Here

An alternative approach is to use the Takens' estimator, which gives a maximum likelihood estimate of D_2 without estimating the slope of the correlation integral directly (Takens, 1985).

An extension to this approach produces an estimate of D_2, which is consistent with measures at all smaller length scales (Guerrero & Smith, 2003). Furthermore, this approach provides constraints on the accuracy of the estimate. The difficulty in calculating reliable estimates of D_2 suggests that the uncertainty in such estimates should be clearly presented through the use of a distribution or error bars. Without a quantification of the uncertainty, the estimate should certainly be treated with caution.

In practice, it is likely to be easier to build a data-driven nonlinear model than to compute a reliable estimate of the dimension. If such a nonlinear model is useful for fulfilling ones objectives, whether they are prediction, classification or the pursuit of a more in-depth understanding and it can outperform simple linear benchmarks, then this in itself can arguably justify the use of nonlinear methods. Being unable to say for certain, whether or not a system is chaotic, is not justification to give up and resign ourselves to traditional linear techniques, but rather a warning about what is possible and how the results of nonlinear time series analyses should be presented.

Estimating Lyapunov exponents

In practice, when dealing with experimental data, it is usual to attempt to estimate the maximal Lyapunov exponent, Λ_1 by investigating a long but finite trajectory obtained by making k as large as possible. Some initial attempts at calculating Λ_1 gave misleading results since they blindly assumed the existence of exponential growth, and therefore can produce a finite value of Λ_1 for stochastic data where the true value is infinite (see, for example, Wolf, 1985). For high quality experimental data, it may be possible to approximate the underlying dynamics by a model, which provides a means of estimating the entire Lyapunov spectrum (Sano & Sawada, 1985; Eckmann et al., 1986; Brown, 1991). More recently, tests for the exponential growth have been advocated (Rosenstein et al., 1993; Kantz, 1994).

Consider the perfect scenario where the equations of motion are known exactly. For example, we will use the two-dimensional chaotic Ikeda map (Ikeda, 1980). The finite time Lyapunov exponents, $\lambda_i^{(k)}$, may be calculated using a recursive QR decomposition (see Abarbanel et al., 1992). Figure 2 shows the convergence of $\lambda_i^{(k)}$ for simulations using one thousand different initial conditions. Even in this noise-free scenario, we find that a relatively large value of k is required to obtain an accurate estimate for Λ_1. The mean of the distribution converges to the Lyapunov exponent via $\left\langle \lambda_i^{(k)} \right\rangle = \Lambda_i + a_i k^{-\alpha_i}$ whereas the standard deviation scales as $b_i k^{-\beta_i}$. Scaling exponents, α_i and β_i typically range between 0.5 and 1 (Abarbanel et al., 1992). The central limit theorem does not necessarily apply to deterministic systems. Note that the distributions for $k=2^{18}$ are non-normal implying that uncertainty is best presented using percentiles, as shown in Fig. 2.

Insert Figure 2 About Here

For experimental data, it is usually difficult to collect sufficient high quality data from a period during which the underlying dynamics are stationary. With the exception of carefully controlled electronic circuits and lasers, most real-world systems are unlikely to provide suitable datasets. In fact it is often a challenge to obtain enough data to pursue a linear analysis. For example, recommendations for the analysis of heart rate variability suggest using a five-minute window to calculate the power spectrum – thereby providing a balance between the minimum data requirements for the Fast Fourier Transform and the period of time where the underlying process is likely to be stationary (Malik & Camm, 1995). In the case of the brain, the electrical activity measured by the electroencephalogram reflects extremely erratic behaviour, making it exceedingly difficult to find stationary periods. Numerous studies attempting to estimate nonlinear measures have opted for windows ranging between 20 and 40 seconds (see McSharry et al., 2003a and references therein). In both of these examples, any attempt to detect the underlying nonlinear dynamics may be obstructed by the noise and non-stationarity. The use of intra-cranial electrodes as opposed to non-invasive scalp electrodes for the identification or prediction of epileptic seizures may increase the likelihood of detecting and exploiting the nonlinearity, but the applicability is obviously reduced.

The utility of nonlinear measures

Despite the fact that invariants such as D_2 and Λ_1 cannot be reliably estimated for short noisy data sets, the applications of the techniques used for their computation will still provide a measure of a nonlinear quantity. It is important to recognize that this quantity no longer reflects an invariant of the system, but simply a nonlinear measure obtained from a particular segment of data. It is still possible to ask (i) if monitoring the time-dependence of this quantity is useful for prediction or classification and (ii) whether this quantity can outperform benchmarks given by the traditional indicators that are based on classical linear statistics.

Using the words “effective” or “approximate” as pre-fixes emphasises that the quantity being calculated is based on the theory of nonlinear dynamical systems but that this quantity should not be taken as a convergent estimate of an invariant of the underlying system. Following this prescription, a number of techniques have been proposed for applications in the field of biomedical research. An effective correlation dimension was postulated as a method for anticipating epileptic seizures (Lehnertz, 1999). Approximate entropy was proposed as a measure of system complexity (Pincus, 1991) and applied to heart rate time series (Pincus & Goldberger, 1994). The difficulty in estimating the entropy of short noisy time series motivated an alternative approach with improved accuracy, known as sample entropy (Richman & Moorman, 2000). Similarly, it is reasonable to ask whether an effective Lyapunov exponent can enhance our ability to classify or detect dynamical transitions. The convergence and divergence of short-term maximum Lyapunov exponents from adaptively selected electrodes has been used to provide predictions of epileptic seizures (Iasemidis et al., 2003).

An alternative method for detecting nonlinear transitions is to measure changes in the distribution of points falling in different Voronoi partitions of the multi-dimensional state space. This conceptually intuitive method, multi-dimensional probability evolution (MDPE), is capable of detecting changes in the underlying dynamics that are invisible to linear statistics. MDPE was employed to detect epileptic seizures using electroencephalograms from non-invasive scalp electrodes. While both variance and MDPE are able to detect seizures, MDPE gave less false positives (McSharry et al., 2002). MDPE was also employed to identify partial epileptic seizures from heart rate time series and may actually be better at detecting nonlinear dynamical transitions than the effective nonlinear measures based on D_2 or Λ_1 since the latter can have identical values for different multi-dimensional distributions (McSharry, 2004).

The explanations as to why these nonlinear measures should outperform linear statistics range from little more than hand-waving arguments to carefully tested hypotheses. Unfortunately, it is only too easy to dream up explanations for success after the fact, which often leads to a misleading approach to conducting scientific research. We will return to the question of what constitutes a useful diagnostic in the last section and propose a framework for ensuring that novel nonlinear methods are of real value. This requires both an investigation of scientific merit and an exploration of simple benchmark tests in order to justify their benefits for practical applications. First, we will investigate approaches for constructing statistical tests of nonlinear methods.

STATISTICAL TESTING

Nonlinear dynamics provide one possible explanation for the many irregularities found in time series, yet linear stochastic processes also have the ability to generate complicated signals. Given that so much theory and expertise exists in specifying, constructing and evaluating linear models, it is extremely imprudent to blindly assume that a nonlinear approach is best without a thorough scientific justification. In order to motivate the construction of nonlinear models, one should be confident that the data collected contains the characteristics of an underlying nonlinear process, which could not have been produced by a linear process. Both nonlinear determinism and the stochasticity arising from random shocks to the system or variations in the parameters are candidates for the irregularity and complexity apparent in empirical data. A certain degree of confidence can be obtained by performing a statistical significance test using the *method of surrogate data.* Each surrogate data set appears like the original time series but only contains some specific prescribed characteristics, e.g. the observations were generated by a linear stochastic process. The idea is to test whether a given nonlinear measure when applied to the original data gives a result that is different to that obtained by applying it to a collection of these surrogates.

Given a nonlinear measure, γ, which takes on the value γ_0 when computed from the original data set, we wish to know whether this particular value suggests that the underlying dynamical process is nonlinear. Another way to think about this is to ask what distribution, $p(\gamma)$, of values would we expect to obtain from a *similar* linear stochastic process. If the value γ_0 is not consistent with $p(\gamma)$, then the data might be nonlinear. In general we will not have any theory to determine an analytical expression for $p(\gamma)$ and so we use the surrogates to estimate this distribution using a Monte Carlo approach (Theiler et al., 1992). We test against a chosen null-hypothesis by constructing N surrogate data sets which (i) preserve certain characteristics of the original time series and (ii) are also consistent with the specified null-hypothesis. By specifying the probability, α, that we are prepared to reject the null hypothesis although it is true, we obtain a test that is valid at the (1- α) significance level. A rank-based one-sided test with significance (1- α) may be employed by generating $N = 1/\alpha -1$ surrogates in order to test whether γ_0 is smaller than expected for data obeying the null hypothesis. By computing the N values for the nonlinear measure: γ_ι $(i=1,\ldots,N)$, we can reject the null hypothesis whenever γ_0 is smaller than all of the γ_ι. For example $N=19$ surrogates are required for testing at the 95% significance level.

The sophistication of the surrogates employed depends on the nature of the characteristics that we want to preserve. Surrogates that preserve the empirical distribution can be obtained by randomly shuffling the original data set without repetition. These can then be used to test the null hypothesis that the data are independent random numbers sampled from some fixed but unknown distribution. Another null hypothesis is that the data comes from a stationary linear stochastic process with normally distributed inputs. Surrogates with the same power spectrum may be obtained by taking the Fourier transform of the original data, randomizing the phases and transforming back to the time domain. This technique may also be applied to multivariate datasets (Pritchard & Theiler, 1996). A more general null hypothesis is that the data was generated by a stationary linear stochastic process with normally distributed inputs, which was then subjected to a monotonic instantaneous time-independent measurement function. To test this hypothesis, we require surrogates with the same empirical distribution and power spectrum. Amplitude-adjusted surrogates are capable of preserving the empirical distribution but yield a slightly modified power spectrum (Theiler et al., 1992). A technique for polishing these surrogates so that they better replicate the power spectrum is also available (Schreiber & Schmitz, 1996). An approach that aims to preserve both the distribution and the spectrum exactly uses a normal autoregressive process and a monotonic static transform (Kugiumtzis, 2002). Constrained randomization using simulated annealing provides a means of generalizing the approach to specify a wide variety of surrogates including multivariate data (Schreiber, 1998).

The method of surrogate data may also be used for testing new techniques such as medical diagnostic tools for classification or prediction. The possibility of using nonlinear methods for facilitating medical diagnostics calls for a different kind of test. This should test the efficacy of the new method against simple linear benchmarks. To achieve this goal, we suggest the use of *clinically relevant surrogates* (McSharry et al., 2003b). By generating surrogates that are not relevant to the specific clinical framework, it is possible to erroneously identify promising techniques for the wrong reasons. For example, there have been a number of conflicting investigations concerning the predictability of epileptic seizures using nonlinear methods. While there are likely to be nonlinear processes active in the transition leading to an epileptic seizure, the question is whether nonlinear methods provide any additional skill, in detecting this dynamical transition, beyond that offered by traditional linear techniques. Some investigations have suggested that a nonlinear measure based on the correlation integral outperformed a linear measure using the autocorrelation function (Martinerie et al., 1998). Unfortunately, this claim was supported by a surrogate data analysis that could not distinguish between changes in the variance and those in the correlation integral. If a simple benchmark, such as moving variance, is to be outperformed before the medical community seriously adopts a new nonlinear technique, then this must be accounted for in the surrogates. One approach is to use block surrogates, which attempt to preserve the time-dependence of the variance (McSharry et al., 2003a).

CHALLENGES

The principle, known as Occam's Razor, suggests that when attempting to select from within a class of models we should aim to identify the simplest model that is compatible with the observations. While increasing the complexity of a model naturally gives more freedom to provide a better fit to the observations, a model with too many parameters will not distinguish between the generative dynamics that we wish to extract and fluctuations due to factors such as measurement errors, non-stationarity and noise. This problem, known as over-fitting is particularly relevant when attempting to construct nonlinear models since such models have the complexity to adapt to extremely complicated data. We will now consider model complexity, parameter estimation and evaluation with regard to constructing nonlinear models.

There are two obvious routes to take when attempting to identify the optimal size or complexity of a nonlinear model. Given a set of candidate explanatory variables, one approach for constructing a model starts with no terms and then considers the addition of one term at a time (Judd & Mees, 1995). An alternative starts with a general description that includes all possible variables and then successively eliminates terms until arriving at a reduced form of the model. The objective of both approaches is to arrive at an intermediate model, which optimises a specific cost function that measures the goodness of fit of the model to the observations. While the latter approach requires more computational effort, it has the advantage of being less likely to fall into local minima (Hendry & Krolzig, 2003). A number of different reduction paths can be searched simultaneously to prevent the algorithm from removing a relevant variable, while retaining other variables as proxies. In this way the global minimum in the cost function can be approached more reliably. This approach has been successfully employed for constructing models of biochemical reactions (Crampin et al., 2004a).

The additional complexity and extra parameters required to specify nonlinear models also implies that while they may be better suited for reproducing time series of irregular observations, they are also more likely to fit the noise as well as the actual dynamics. This problem, known as *over-fitting* implies that the model will provide good results on in-sample data (that used for specifying and training the model), but is likely to fail to generalise to new data. It is extremely difficult to distinguish between fluctuations due to the underlying dynamics and those due to the observational uncertainty. An obvious sign of an over-fit model is one that performs better on training (in-sample) data than testing (out-of-sample) data as shown in Fig. 3. Splitting the available dataset in two and using one half for training and the other for testing is a good procedure for detecting such problems. An alternative method that may be employed when working with short datasets is to generate leave-one-out models and to predict the omitted point (or points), thereby providing quasi out-of-sample results.

Insert Figure 3 About Here

One approach for reducing over-fitting is to add a penalty term to the cost function that penalises non-parsimonious models. The more complex the model, measured by the number of parameters, the heavier the penalty. In this way, the addition of the penalty attempts to reproduce the effect of evaluating the model on an out-of-sample testing dataset, by creating a minimum in the cost function for an intermediate model complexity, as shown in Fig. 3. Note that such penalty terms are often referred to as regularisation techniques and can also be viewed as the incorporation of priors. The simplest form of penalty function is one that is proportional to the number of parameters (Akaike, 1974). A penalty function that depends on both the number of observations and number of parameters may be determined using a Bayesian approach (Schwarz, 1978). Alternatively, this penalty function may be arrived at by assuming that the optimal model minimises the description length of an encoding of the data (Rissanen, 1980). While the idea of having a cost function that works for all models is attractive, it is more realistic that a universal cost function does not exist and that the appropriate cost function will depend intimately on the structure of the model and its application.

When applying nonlinear models to noisy time series, one should expect the goodness of fit, measured by the discrepancy between the model and the data, to vary with position in state space. While this property of nonlinear models is generally ignored, the contribution of the local stretching factors induced by the nonlinear model can be used to identify model error by checking for the consistency between the model and the observational uncertainty (McSharry & Smith, 2004). This approach allows the user to detect why the model is inadequate and can also suggest how these imperfections can be reduced.

Parameter estimation for nonlinear models is also difficult since many of the traditional linear goodness of fit tests are unsuitable for applications using nonlinear models. For example, many classical techniques assume normal distributions for the measurement errors and the model error term. In contrast, even in the case of normally distributed measurement errors, the application of nonlinear models for generating predictions will convolve these measurement errors so that the effective model error term depends on the local structure of the nonlinear model. This implies that standard parameter estimation methods, such as least squares and total least squares will lead to biased estimates. For this reason it is better to use a parameter estimation technique that directly incorporates the interplay between the nonlinear structure of the model and the observational uncertainty (McSharry & Smith, 1999). In fact, root-mean-square evaluation statistics could actually reject the very nonlinear system that generated the data in the first place.

Meteorologists are familiar with the task of generating forecasts using nonlinear weather models. A method that has recently been employed for justifying the use of nonlinear models for prediction is to generate and evaluate probabilistic forecasts (Palmer, 2000). These forecasts, known as ensemble forecasts, take account of the uncertainty in the initial conditions by sampling from the initial distribution and propagating each trajectory forward in time. In this way, the ensemble forecasts can also quantify the uncertainty in the prediction. One approach for evaluating these ensemble forecasts is based on information theory (Roulston & Smith, 2002). Indeed a true measure of the success of a nonlinear modeling framework should depend on the application of the model. For this reason, it may be necessary to incorporate the users' utility function in order to provide a model with optimal functionality.

DISCUSSION AND CONCLUSION

A growing interest in interdisciplinary research and the resulting interplay between traditional model-driven research and modern data-driven modeling techniques may explain why nonlinear dynamics is suddenly contributing to solving so many practical problems (Stark & Hardy, 2003). Furthermore, the increasing availability of efficient computing resources, both in terms of speed and memory, greatly facilitates the application of nonlinear time series analysis techniques. This is especially relevant in the case of large data sets, which are required if such techniques are to prove themselves useful. The ability to analyse large data sets has given rise to new fields of interdisciplinary research, such as systems biology (Kitano, 2002). New experimental techniques in biochemistry offer an increasing number of empirical datasets and nonlinear models are likely to provide a deeper understanding of the mechanisms underlying complex biochemical reactions (Crampin et al., 2004b).

At present the utility of nonlinear methods is still being debated. A large amount of research is required before these new methods can compete with the vast body of expertise that is available for guiding the application of classical linear statistical techniques. The foremost challenge concerns the statistical testing of new nonlinear methods and the ability to be confident that similar results cannot be achieved by simpler linear approaches. Fortunately, today's researchers can easily carry out extensive Monte Carlo simulations on an average PC. Such simulations are necessary for applying statistical tests in the absence of analytical distributions, a typical challenge in the construction of relevant null hypotheses for evaluating nonlinear time series analysis techniques. The use of clinically relevant statistical tests is particularly important if these nonlinear methods are to be used for constructing new medical diagnostic tools.

In order to make scientific progress it is necessary to have well understood and carefully defined benchmarks. The biomedical signal processing community is accustomed to testing new techniques on public domain databases such as Physiobank (Goldberger et al., 2000; www.physionet.org/physiobank). While the use of such databases is to be applauded, it does not fully address the possibility of over-fitting, whereby a model or technique appears to have high performance statistics but fails to generalize to new data. This failure can be circumvented, to some extent, by holding back some of the data for testing while the remaining data is used for learning. These databases contain real biomedical signals recorded from human subjects under standard clinical settings. For this reason, it is impossible to know a priori which part of the signal corresponds to the underlying dynamics and how much can be attributed to noise. This is complicated further by electrical interference, movement artifacts, muscle artifacts and missing data.

One approach for comparing methods is to employ a database of synthetic signals. A nonlinear dynamical model has been used to generate realistic synthetic electrocardiogram signals (McSharry et al., 2003c). Open-source code for a freely available algorithm provides researchers with the ability to generate electrocardiogram signals with known characteristics, both in the time and frequency domains (www.physionet.org/matlab/ecgsyn). Different realizations of the electrocardiogram signal may be generated by varying the seed of the random number generator. These realizations can then be used to compute the uncertainty inherent in signal processing techniques and to provide error bars for estimates. The ability to easily produce such signals equips many researchers with the data necessary to compare and evaluate their new biomedical signal processing techniques. While this is currently a useful exercise for illustrating the advantages of new techniques, it is hoped that in the future it could form the basis of a gold seal of approval for new techniques. Many journals now insist that databases used for obtaining new scientific evidence are made public-domain as a prerequisite of publication. This has the advantage of allowing method-driven researchers to compare different techniques. Having access to carefully labeled public-domain databases, such as those at Physionet, can reduce the time required to accumulate and investigate biomedical databases and increase the pace of scientific exploration.

By promoting good practice when applying nonlinear time series techniques it will be possible to avoid the many dangerous pitfalls that await the unsuspecting researcher. In summary we recommend that (i) estimates calculated from time series are supplied with confidence intervals,

(ii) the method of surrogates is employed before discarding traditional linear models and (iii) synthetic data is used to explore new techniques and to formulate simple benchmark tests. While the pursuit of chaos may be akin to searching for the Holy Grail, the future of nonlinear time series analysis and its application to empirical data is likely to be a rewarding one.

ACKNOWLEDGMENTS

PEM is a Royal Academy of Engineering Research Fellow and acknowledges support from the Engineering and Physical Sciences Research Council, UK. This paper benefited from numerous discussions with Leonard A. Smith.

REFERENCES

Abarbanel, H. D. I., Brown, R. & Kennel, M. B. (1992). Local Lyapunov Exponents Computed from Observed Data. *Journal of Nonlinear Science*, *2*, 343-365.

Akaike, H., (1974). A new look at the statistical idenification model. *IEEE Trans. Automat. Contr*. *19*, 716–723.

Brown, R., Bryant, P., & Abarbanel, H. D.. I. (1991). Computing the Lyapunov spectrum of a dynamical system from observed time series. *Physical Review A*, *43*, 27-87.

Crampin, E. J., Schnell, S. & McSharry, P. E. (2004a). Mathematical and computational techniques to deduce complex biochemical reaction mechanisms, *Progress in Biophysics and Molecular Biology 86*, 77-112.

Crampin, E. J., McSharry, P. E. & Schnell, S. (2004b). Extracting biochemical reaction kinetics from time series data. In M. Gh. Negoita et al. (Eds.), *Knowledge-Based Intelligent*

Information & Engineering Systems, Lecture Notes in Artificial Intelligence, 3214, 329-336. Springer-Verlag, Berlin Heidelberg.

Eckmann, J.P. Oliffson Kamphorst, S., Ruelle, D. & Ciliberto, S. (1986). Lyapunov exponents from a time series. *Phys .Rev. A, 34* (6), 4971-4979.

Essex, C. and Nerenberg, M. A. H. (1991). Comment on "Deterministic Chaos: The Science and the Fiction" by D. Ruelle. *Proc. R. Soc. Lond. A, 435*, 287-292.

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

Gleick, J. (1988). Chaos: Making a New Science. New York: Penguin Books.

Goldberger A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P.Ch., Mark, R.G., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation 101(23)*, e215-e220.

Grassberger, P. and Procaccia, I. (1983). Measuring the Strangeness of Strange Attractors. Physica D 9, 189-208.

Grassberger, P., & Procaccia, I. (1983). Characterisation of strange attractors. *Physical Review Letters, 50*, 346-369.

Guerrero, A. and Smith, LA. (2003). Towards coherent estimation of correlation dimension.

Physics Letters A 318, 373-379.

Hendry, D. F., Krolzig, H. M. (2003). New developments in automatic general-to-speci.c modelling. In Stigum, B. P. (Ed.), *Econometrics and the Philosophy of Economics*. Princeton University Press, Princeton.

Iasemidis, L. D., Shiau, D. S., Chaovalitwongse, W., Sackellares, J. C., Pardalos, P. M., Principe, J. C., Carney, P. R., Prasad, A., Veeramani, B., Tsakalis, K. (2003). Adaptive epileptic seizure prediction system. *IEEE Trans Biomed Eng,50*(5), 616-627.

Ikeda, K. and Daido, H. (1980). Optical turbulence: chaotic behaviour of transmitted light from a ring cavity. *Physical Review Letters, 45*, 709-712.

Judd, K., Mees, A., (1995). On selecting models for nonlinear time series. *Physica D, 82*, 426–444.

Kantz, H., (1994). A robust method to estimate the maximal Lyapunov exponent of a time series, *Physics Letters A, 185*, 77-87.

Kantz, H. & Schreiber, T. (2003). *Nonlinear time series analysis*. (2nd Edition), Cambridge: Cambridge University Press.

Kennel, M. B. and Brown, R., and Abarbanel, H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Physical Review A 45*, 3403-3411.

Kitano, H. (2002). Systems Biology: a brief overview. *Science, 295*, 1662-1664 (2002)

Kugiumtzis, D. (2002). Statically transformed autoregressive process and surrogate data test for nonlinearity. *Physical Review E, 66*, 025201(R).

Lehnertz, K. and Elger, C. E. (1998) Can Epileptic Seizures be Predicted? Evidence from Nonlinear Time Series Analysis of Brain Electrical Activity, *Physical Review Letters, 80*,

5019–5022.

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

Malik, M. & Camm, A. J. (1995). *Heart rate variability*. Armonk, NY: Futura.

Martinerie, J., Adam, C., Le van Quyen, M., Baulac, M., Renault, B. & Varela, F. J. (1998). Can epileptic crisis be anticipated?, *Nature Medicine, 4*, 1173-1176.

May, R. M. (1976). Simple Mathematical Models with Very Complicated Dynamics. *Nature, 261*, 459-467.

McSharry, P. E. and Smith, L. A. (1999). Better Nonlinear Models from Noisy Data: Attractors with Maximum Likelihood, *Physical Review Letters, 83*, 4285-4288.

McSharry, P. E., He, T., Smith, L. A. & Tarassenko, L. (2002). Linear and nonlinear methods for automatic seizure detection in scalp electroencephalogram recordings. *Medical & Biological Engineering & Computing, 40*, 447-461.

McSharry, P.E., Smith, L. A. and Tarassenko, L. (2003a). Comparison of predictability of epileptic seizures by a linear and a nonlinear method. *IEEE Transactions on Biomedical Engineering 50*(5):628-633.

McSharry, P.E,, Smith, L. A. and Tarassenko, L. (2003b). Prediction of epileptic seizures: are nonlinear methods relevant? *Nature Medicine 9*, 241-242.

McSharry, P. E., Clifford, G., Tarassenko, L. and Smith, L. A. (2003c). A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering, 50*, 289-294.

McSharry, P. E. and Smith, L. A. (2004). Consistent nonlinear dynamics: identifying model inadequacy, *Physica D, 192*, 1-22.

McSharry, P. E. (2004). Detection of dynamical transitions in biomedical signals using nonlinear methods. In M. Gh. Negoita et al., *Knowledge-Based Intelligent Information & Engineering Systems, Lecture Notes Artificial Intelligence 3215*, 483-490, Springer-Verlag, Heidelberg Berlin.

Nerenberg, M. A. H. & Essex, C. (1990). Correlation dimension and systematic geometric effects. *Physical Review A, 42*, 7065-7074.

Palmer, T. N. (2000). Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys*, *63*, 71–116.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA, 88*, 2297–2301.

Pincus, S. M. (1994). Goldberger AL. Physiological time-series analysis: What does regularity quantify? Am J Physio Heart Circ Physiol, 266, H1643-H1656.

Pritchard, D. & Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Physical Review Letters*, *73*, 951-954.

Provenzale, A., Smith, L. A., Vio, R. & Murante, G. (1992). Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica D, 58*, 31-49.

Richman, J. S., Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol 278*, H2039-H2049.

Rissanen, J., 1980. Consistent order estimates of autoregressive processes by shortest description of data. In: Jacobs, O.,et al. (Ed.), Analysis and Optimisation of Stochastic Systems. Academic Press, New York.

Rosenstein, M. T., Collins, J. J., & De Luca, C. J. (1993). A practical method for calculating largest Lyapunov exponents from small data sets, *Physica D, 65*, 117-134.

Rosenstein, M. T., Collins, J. J., & De Luca, C. J. (1994). Reconstruction expansion as a geometry-based framework for choosing proper delay times, *Physica D*, 73, 82-98.

Roulston, M. & Smith, L.A. (2002). Evaluating probabilistic forecasts using information theory. *Monthly Weather Review, 130*,1653-1660.

Ruelle, D. (1990). Deterministic chaos: the science and the fiction. *Proc. R. Soc. Lond. A.*, 427, 241-248.

Sano, M. & Sawada, Y. (1985). Measurement of the Lyapunov spectrum from a chaotic time series. *Physical Review Letters, 55*, 1082-1085.

Sauer, T., Yorke, J. A., & Casdagli, M. (1991). Embedology. *Journal of Statistical Physics,* 65, 579-616.

Scheinkman, J. A. & Le Baron, B. (1989). Nonlinear dynamics and stock returns *The Journal of Business 62*, 311-337.

Schreiber, T. & Schmitz, A. (1996) Improved Surrogate Data for Nonlinearity Tests. *Physical Review Letters, 77*, 635–638.

Schreiber, T. (1998) Constrained Randomization of Time Series Data, *Physical Review Letters, 80*, 2105–2108.

Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Statist. 6*, 461–464.

Smith, L. (1988). Intrinsic limits on dimension calculations. *Physics Letters A, 133*, 283-288.

Smith, L. A., Ziehmann, C. & Fraedrich, K. (1999). Uncertainty Dynamics and Predictability in Chaotic Systems. *Quart. J.R. Meteorological Soc. 125***,** 2855–2886.

Stark, J and Hardy, K. (2003) Chaos: Useful at last? *Science, 301*, 1192-1193.

Strang, G. (1988). *Linear algebra and its applications*. San Diego: Harcourt College Publishers.

Takens, F. (1981) Detecting strange attractors in turbulence. In *Lecture notes in mathematics, 898. Dynamical systems and turbulence*, 366-381. Springer-Verlag, Berlin.

Takens, F. (1985) On the numerical determination of the dimension of an attractor. In Braaksma, B.L.J., Broer, H.W., and Takens, F. (editors), *Dynamical systems and bifurcations, Lecture Notes in Mathematics, 1125,* Springer, Heidelberg-New York.

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

Theiler, J. (1990). Statistical Precision of Dimension Estimators. *Physical Review. A, 41*, 3038-3051.

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

Tsonis, A. A. (1992). *Chaos: from theory to applications*. Plenum Press, New York.

Wolf, A., Swift, J., Swinney, H., and Vastano, J. (1985). Determining lyapunov exponents from a

time series. *PhysicaD*, *16*, 285-317.

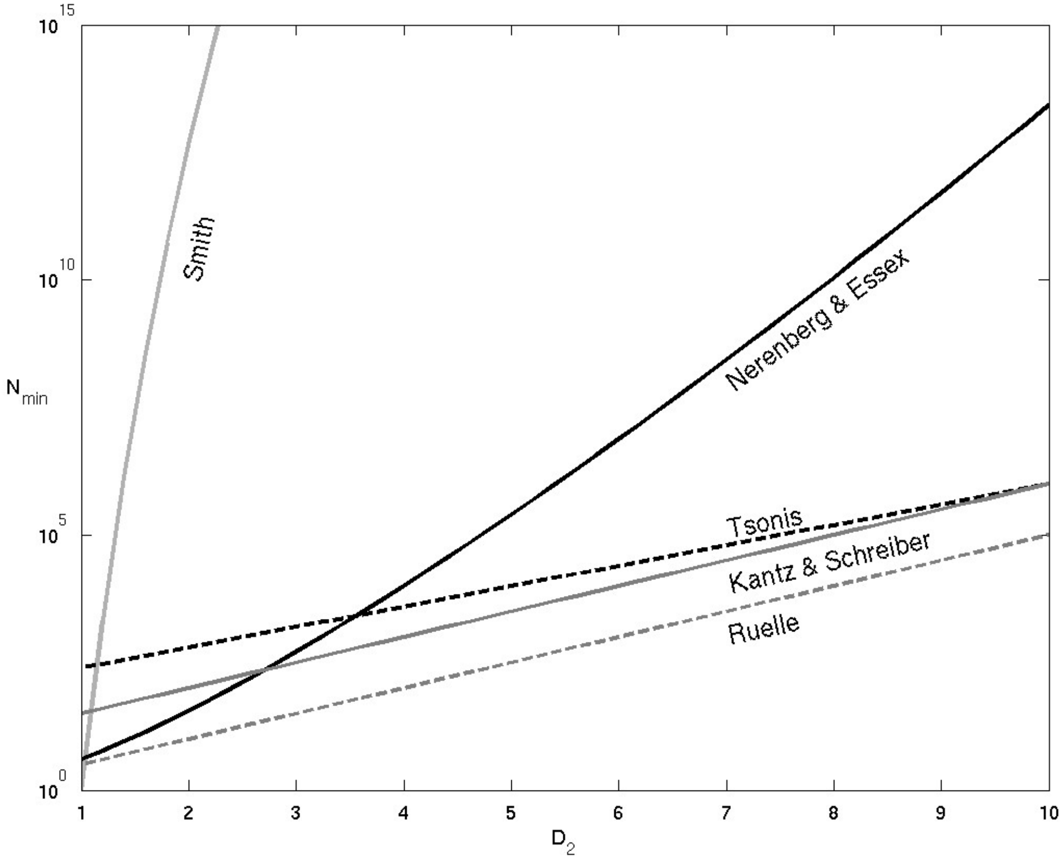
Ziehmann, C., Smith, L. A., & Kurths, J. (1999). The Bootstrap and Lyapunov Exponents in

Deterministic Chaos , *Physica D, 126*, 49-59.

Figure 1. The suggested minimum number of data points, N_{min}, required to provide a reliable estimate of the correlation dimension, D_2.

Figure 2. Finite-time Lyapunov exponents for the two-dimensional chaotic Ikeda map: (a) convergence of $\lambda_1^{(k)}$ with k, (b) convergence of $\lambda_2^{(k)}$ with k, (c) probability density function of $\lambda_1^{(k)}$ of for $k=2^{18}$, and (d) probability density function of $\lambda_2^{(k)}$ of for $k=2^{18}$. In (a) and (c), the dashed lines reflect the 5%, 50%, 95% percentiles and the solid line is the mean.

Figure 3. Analysis of model error versus model complexity as measured by the number of parameters, K, using a training and testing dataset (left). Adding a penalty term to the cost function provides an alternative approach for avoiding over-fitting problems and may enable the identification of a model that generalizes well (right).



10^15
10^10
10^5
10^0
N_min
Smith
Nerenberg & Essex
Tsonis
Kantz & Schreiber
Ruelle
1
2
3
4
5
6
7
8
9
10
D_2

Figure 1

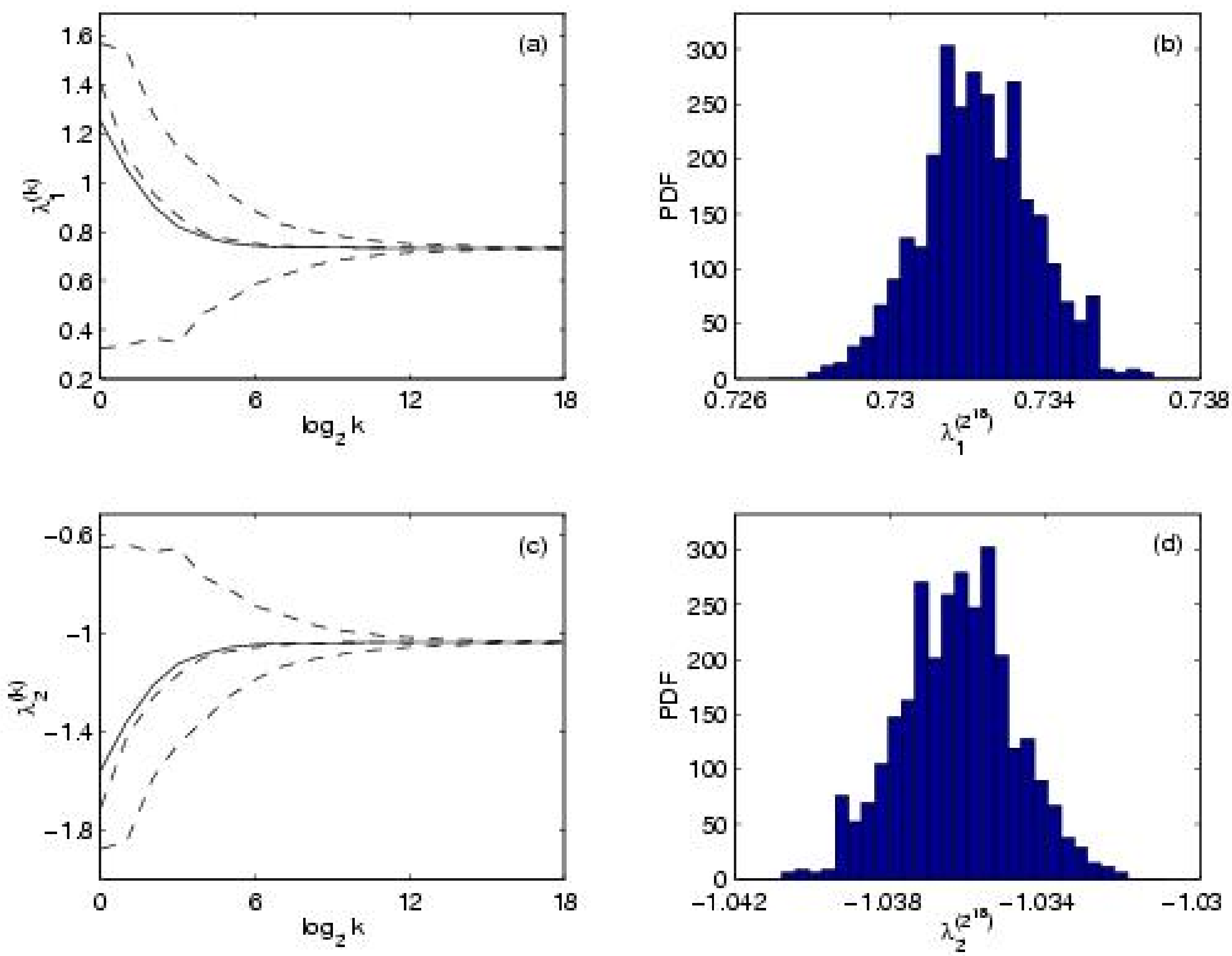

(a)
(b)
(c)
(d)
PDF
PDF
log2 k
log2 k
0.726
0.73
0.734
0.738
−1.042
−1.038
−1.034
−1.03

Figure 2

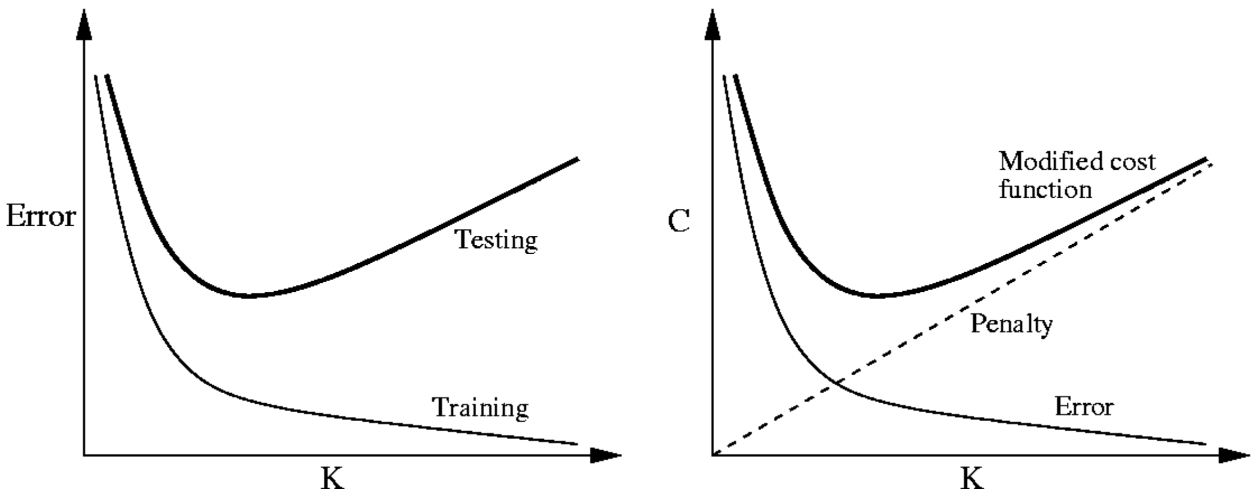

Error
Testing
Training
K
C
Modified cost
function
Penalty
Error
K

Figure 3