Forecasting high waters at Venice Lagoon using chaotic time series analysis and nonlinear neural networks

J. M. Zaldívar, E. Gutiérrez, I. M. Galván, F. Strozzi and A. Tomasin

ABSTRACT

Time series analysis using nonlinear dynamics systems theory and multilayer neural networks models have been applied to the time sequence of water level data recorded every hour at 'Punta della Salute' from Venice Lagoon during the years 1980–1994. The first method is based on the reconstruction of the state space attractor using time delay embedding vectors and on the characterisation of invariant properties which define its dynamics. The results suggest the existence of a low dimensional chaotic attractor with a Lyapunov dimension, D_L, of around 6.6 and a predictability between 8 and 13 hours ahead. Furthermore, once the attractor has been reconstructed it is possible to make predictions by mapping local-neighbourhood to local-neighbourhood in the reconstructed phase space. To compare the prediction results with another nonlinear method, two nonlinear autoregressive models (NAR) based on multilayer feedforward neural networks have been developed.

From the study, it can be observed that nonlinear forecasting produces adequate results for the 'normal' dynamic behaviour of the water level of Venice Lagoon, outperforming linear algorithms, however, both methods fail to forecast the 'high water' phenomenon more than 2–3 hours ahead.

Key words | forecasting, nonlinear neural networks, time series analysis

J. M. Zaldívar (Corresponding author)
E. Gutiérrez
European Commission,
Joint Research Centre,
Institute for Systems, Informatics and Safety,
TP 250, 21020—Ispra (VA),
Italy

I. M. Galván
Department of Computer Science,
Carlos III University of Madrid,
C/ Butarque 15,
28911—Leganés,
Madrid,
Spain

F. Strozzi
Libero Istituto Universitario Carlo Cattaneo,
Engineering Department,
Quantitative Methods Group,
21053—Castellanza (VA),
Italy

A. Tomasin
CNR-ISDGM
Università Ca'Foscari
Dorso Duro 2137,
30123—Venice,
Italy

INTRODUCTION

Unusually high tides, or sea surges, result from a combination of chaotic climatic elements in conjunction with the more normal, periodic, tidal systems associated with a particular area. The prediction of such events has always been the subject of intense interest to mankind, not only from a human point of view, but also from an economic one. The most famous example of flooding in the Venice Lagoon occurred in November 1966 when, driven by strong winds, the Venice Lagoon rose by nearly 2 m above the normal water level. The damage to the city's homes, churches and museums ran into hundred of millions of Euros.

The complex behaviour of tides is still not properly understood. The exact prediction of the oceanic response to the forcing functions of astronomic and atmospheric agents has so far proved intractable. While the former (astronomic) is a periodic predictable phenomenon caused by the relative motion of the earth, the moon and the sun, the latter is a complex phenomenon usually treated as stochastic. In fact, the problem has been approached either through numerical models or statistical methods. The first approach consists of solving the hydrodynamic equations by means of finite-difference techniques to obtain the variations in sea level induced by the sea tides, and by wind and atmospheric pressure fields acting over the waters of the sea. One and two-dimensional numerical models have been developed to forecast the 'high waters' phenomenon in Venice (Accerboni *et al*. 1971; Accerboni & Manca 1973; Tomasin 1973). However, these models require the computation of the meteorological forcing functions at each point of the finite difference grid and, hence, they are computationally expensive; (see Vieira *et al.* 1993 for a description of the model and the obtained results). On the other hand, linear stochastic models are suitable for online forecasting since they are simple and their computational burden is low. ARMA (autoregressive moving average) models using data

of pressure and level at Venice, or using data of the Adriatic sea level at different places, are currently used to forecast the water level at Venice Lagoon (see Tomasin 1972, Michelato *et al.* 1983, and Moretti & Tomasin 1984, among others).

The advent of nonlinear time series analysis and the mathematical theorems associated with chaotic dynamics, in the late 1980s, are now making it possible not only to qualify, but also to quantify the behaviour of some complex systems. The techniques, which consist of representing complex system dynamics in multi-dimensional phase-space as a geometrical object, have had some success at predicting chaotic behaviour.

A preliminary study on the application of nonlinear time series analysis using delay coordinate embedding on the tidal data from the Venice Lagoon from 1980 to 1984, was carried out by Vittori (1992 1993). The results obtained show that the a-periodic component, which is super-imposed on the periodic (astronomical) one, possesses a chaotic-deterministic character. The time delay was empirically calculated and a value of 200 hours was selected for most of the calculations. Thereafter, the embedding dimension was calculated using the correlation dimension (Grassberger & Procaccia 1983) as the invariant property and a value of 6 was obtained. With these two values, the largest Lyapunov exponent and, hence, the predictability window, was computed using the algorithm developed by Wolf *et al.* (1985) which gave a value of approximately 40 hours. However, there were some open questions: a time delay of 200 hours seems too long to assume that some correlation between the data exists, and the predictability window given as 40 hours, using only water level data, disagrees with the actual estimation methods. Furthermore, there were no standard errors in the water level forecasting so it is not possible to make a comparison with the previous linear stochastic techniques.

In a recent work, Bergamasco *et al.* (1995) studied wind-driven surface wave data on an offshore platform about 20 km from Venice in the northern Adriatic Sea. Even though they found a finite value for the correlation dimension (approximately 7) and a positive value for the largest Lyapunov exponent (approximately 0.0015 bit/sec), Bergamasco *et al.* (1995) concluded that the correct interpretation was that the data are essentially stochastic, and that the correlation dimension and Lyapunov exponents result from the anomalous statistical behaviour of certain near-Gaussian random processes.

In this work we have tried to develop further this nonlinear time series analysis using delay coordinate embedding by applying recently developed numerical algorithms (Abarbanel 1996; Kantz & Schreiber 1999) in an effort to understand the nonlinear dynamic behaviour of the system and to provide an early warning prediction of unusually high water.

We are also interested in comparing the results obtained from applying nonlinear dynamic systems theory with generalised nonlinear autoregressive models, also named NAR models. In this case, the time series behaviour can be captured by expressing the value $x(k+1)$ as a nonlinear function of the n previous values of the time series, $x(k), \ldots, x(k-n)$, that is:

$$x(k+1) = \boldsymbol{F}[x(k), \ldots, x(k-n)], \tag{1}$$

where k is the time variable and $\boldsymbol{F}$ is some function defining a very large and general class of time series. The explicit form of the function $\boldsymbol{F}$ is usually unknown and it must be determined using approximation techniques. The present study deals with long-term or multi-step prediction, i.e. how to achieve predictions several steps ahead into the future, $x(k+1)$, $x(k+2)$, $\ldots$, $x(k+h)$, starting from information at time k. Hence the goal is to approximate the function $\boldsymbol{F}$ such that the model given by Equation (1) can be used as a multistep prediction scheme.

The development of mathematical analysis has led to the discovery of important classes of approximation functions which can be used to obtain $\boldsymbol{F}$. These include polynomials, trigonometric series, orthogonal functions, splines, etc. Other additional families of functions are artificial neural networks. Different authors (Cybenko 1989; Hornik *et al.* 1989) have shown independently that multilayer neural networks, with as few as one hidden layer and with an arbitrarily large number of neurons in the hidden layer are capable of approximating any non-linear continuous function. In this work, the class of functions associated with a multilayer perceptron and

with one hidden layer has been considered to approximate the functional $\boldsymbol{F}$ in Equation (1).

In the following section the experimental data are presented and briefly analysed. The data can be seen as the sum of the periodic and predictable astronomic effects with the weather effects superimposed. The periodic component may be removed from the time series. In the next section we discuss the idea of reconstructing the phase space of the system by the use of time delays of observed data. We discuss methods of determining the appropriate time delay to use in practical reconstruction of phase space, as well as the dimension of the phase space in which we must work. We apply the techniques to the data from Venice Lagoon. Once the state space has been reconstructed, methods to calculate the invariant quantities preserved by phase space reconstruction are discussed. Principally, we are interested in the calculation of the Lyapunov exponents since they will give us an idea of the predictability window and the type of dynamics of the system. Furthermore, the dynamic degrees of freedom of the water level at Venice Lagoon are also obtained as a first step to develop a model able to reproduce the observed signal. After we have established ways of classifying the physical system leading to the observations, we move on to a discussion of model building to make numerical predictions. Concerning the nonlinear autoregressive models, we first introduce the two neural models based on multilayer feedforward networks used during this study. Thereafter, we discuss how it is possible to reduce the neural network complexity during the training step and, finally, we compare the forecasting results produced by both methods.

The results show that both nonlinear predictors give accurate results for short- and long-term predictions during 'normal' periods, but fail in predicting more than 4 hours ahead the 'high waters' phenomenon.

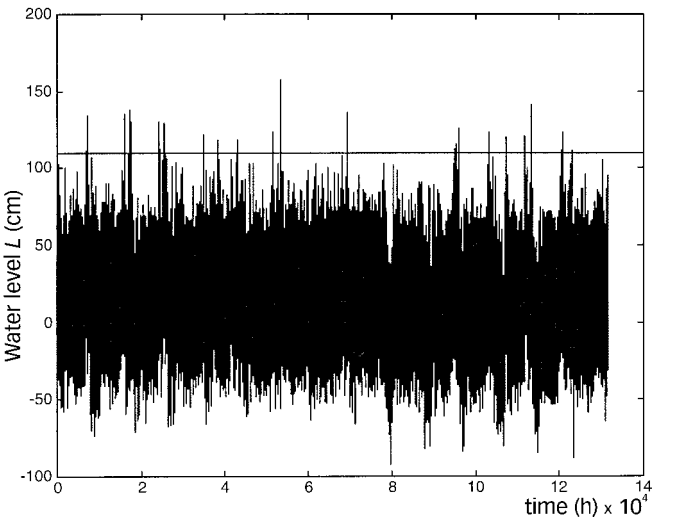

Figure 1 | Water level at the Venice Lagoon from 1980 to 1994.

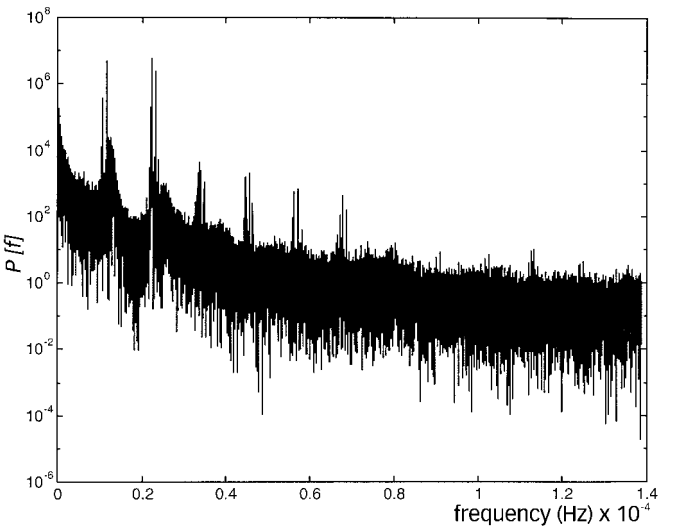

Figure 2 | The Fourier power spectrum for the data from Figure 1.

WATER LEVEL DATA AT VENICE LAGOON

Figure 1 shows the water level of Venice Lagoon between 1980 and 1994 sampled every hour, the high-water level, i.e. no less than 110 cm, is also shown. The high water phenomenon has a characteristic behaviour during the year, November and December being the months in which the phenomenon is more pronounced whereas during the summer it has never been observed (Moretti & Tomasin 1984). The analysis of the power spectrum of these data, see Figure 2, indicates the existence of periodicities related to the diurnal and semidiurnal tides with a period of 12 and 24 hours,

respectively and a broadband spectrum typical of noise or chaotic systems.

Assuming that the total sea level is simply the linear superposition of tidal oscillations and weather-induced oscillations (Michelato *et al.* 1983), then the deviation of the sea level $L(t)$ from the mean sea level value can be expressed as:

$$L(t) = T(t) + S(t) + N(t), \tag{2}$$

where $T(t)$ is the tidal complex, $S(t)$ is the surge, i.e. the meteorological induced perturbation, and $N(t)$ is the noise term which is the residual variation not explicitly accountable and with zero mean value. The residual $R(t)$ can now be defined as:

$$R(t) = L(t) - T(t) = S(t) + N(t). \tag{3}$$

If one supposes that surge-tide interactions in the northern Adriatic basin can be neglected, the term $R(t)$ depends essentially on meteorological parameters. The tidal oscillations produced, which are mainly due to the relative motion of the earth, moon and sun, can be decomposed in a sum of sine functions, each one with its own periodicity. The sum of these sine functions will produce $T(t)$ that can be subtracted from the original signal to obtain $R(t)$. Figure 3 shows this term for the Venice Lagoon between 1980 and 1994 whereas in Figure 4 its power spectrum is shown. As can be seen the periodicities of 12 and 24 hours have practically disappeared and the power spectrum is broadband. Tidal prediction can be performed accurately by several well established methods and, hence, the prediction of the term $R(t)$ is the critical point of the problem.

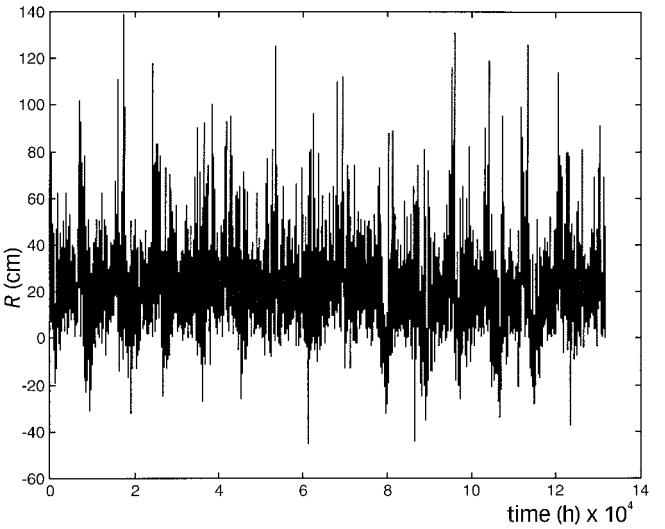

Figure 3 | Residual term, *R*, at the Venice Lagoon from 1980 to 1994.

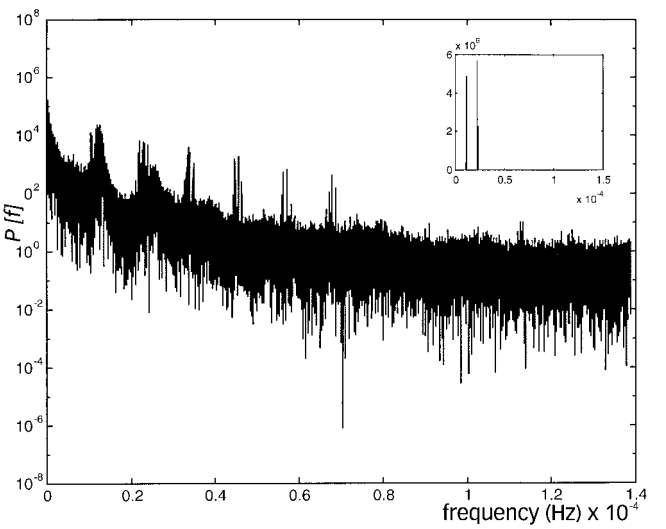

Figure 4 | The Fourier power spectrum for the data from Figure 3 (small figure: power spectrum of the subtracted tidal oscillations).

Hurst coefficient of time series

The Hurst exponent is a measure of the long-time correlations in a time series and was originally used to characterise flow in rivers and dams (Hurst 1951). The Hurst exponent has the characteristic that allows classification of time series since it is able to distinguish the existence of long-range correlations from random noise. In this method, also called rescaled range analysis (R/S analysis), the span of a random process is divided by its variance, resulting in a new variable that depends on the length of the data record. Let us define the time average of the time series $L(t)$ over the interval of time τ:

$$\langle L \rangle_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} L(t)\,. \tag{4}$$

Let us also define $A(t)$, the accumulated departure of $L(t)$ from the mean as:

$$A(t,\tau) = \sum_{u=1}^{t} (L(u) - \langle L \rangle_\tau) \tag{5}$$

so that the span of the process is defined by:

$$S(\tau) = \max_{1\le t\le\tau} A(t,\tau) - \min_{1\le t\le\tau} A(t,\tau)\ . \tag{6}$$

Let us also introduce the standard expression for the variance:

$$V(\tau) = \sqrt{\frac{1}{\tau}\sum_{t=1}^{\tau} (L(t) - \langle L \rangle_\tau)^2}\ . \tag{7}$$

The rescaled Hurst analysis consists in studying the properties of the ratio:

$$R(\tau) = \frac{S(\tau)}{V(\tau)}\ . \tag{8}$$

The dependence of $R(t)$ on the number of data points follows an empirical power law described as $R(t) = R_0\tau^H$ obtained over a wide range of time lengths τ, where H is the Hurst exponent. The Hurst exponent, $0 \le H \le 1$, is equal to 0.5 for random, white noise series, <0.5 for rough anticorrelated series, and >0.5 for positively correlated series.

The deficiencies in time-series analysis for identifying, describing and modelling long-range correlations was pointed out by Mandelbrot & Van Ness (1968). Mandelbrot (1983) using the theory of fractional Brownian motion (fBm), showed that fractional Brownian motion could provide an explicit statistical realisation of the power law scaling, supporting the interpretation of natural phenomena in terms of fractal functions.

Estimating the Hurst coefficient of time series

Several methods are available for estimating the Hurst coefficient of a one-dimensional time series: scaled windowed variance, dispersional analysis, Hurst rescaled range analysis, autocorrelation measures, and power spectral analysis. Bassingthwaighte & Raymond (1994) have demonstrated that the last three methods for estimating H are highly biased and variable. For example, for a series of 512 points, a 95% confidence interval for H, based upon a re-scaled range estimate of $H = 0.5$ will include every H from 0.2 to 0.9. Autocorrelation analysis estimates are highly biased towards $H = 0.5$. Fourier spectral analysis based on the periodogram also has a high variance in its estimates.

In this work we have used the scaled windowed variance method (Cannon *et al.* 1997). The scaled windowed variance methods measure variability at different scales in order to estimate H. A signal is repeatedly divided into windows, but instead of computing the standard deviation of the means within the windows, the means of the standard deviations within the windows are used to obtain an estimate of H (Cannon *et al.* 1997). Short data sets are difficult to characterise. Noise present in a real time series can mask long-range correlations among the signal elements. In the case of white noise, it would induce a bias toward $H = 0.5$. In this case it is possible to detect this by excluding small window sizes when computing the linear regression of log ($R(t)$) versus log (t). In case of coloured noise, it is not known presently how to distinguish between a simple fractal signal and a signal that is the sum of two simple fractal signals.

We have estimated the Hurst coefficient, using the Venice Lagoon level data points ($H = 0.83$) and the residual terms ($H = 0.90$). We have used the bridge detrended (BD) scaled windowed method which has been recommended for series longer than 2^{12} data points, the algorithm is described in Cannon *et al.* (1997). As can be seen from Figures 5 and 6 the analysed time series show long memory effects, $H > 0.5$, even when the tidal component has been removed.

Noise, nonstationarity and nonlinearity in Venice Lagoon data

It is known that some types of noise, that are clearly not associated with low dimensional chaotic systems, are able to fool the algorithms that characterise the structure of chaotic attractors (see Tsonis & Elsner 1992). For this reason, the standard time delay embedding techniques and

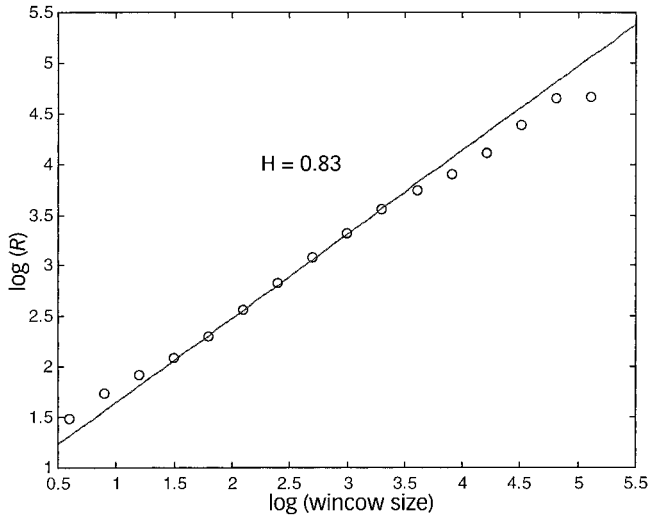

Figure 5 | Estimation of the Hurst coefficient for the level data from Figure 1. Selected points have been excluded from the calculation (see Cannon *et al*. 1997, Table 2).

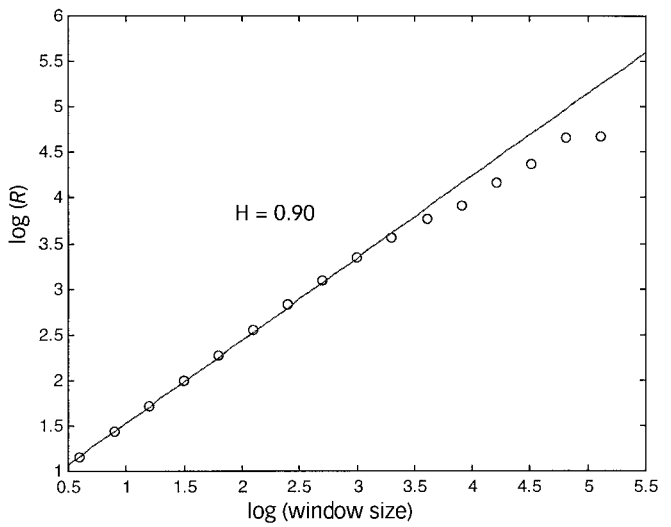

Figure 6 | Estimation of the Hurst coefficient for the residual term data from Figure 3. Selected points have been excluded from the calculation (see Cannon *et al*. 1997, Table 2).

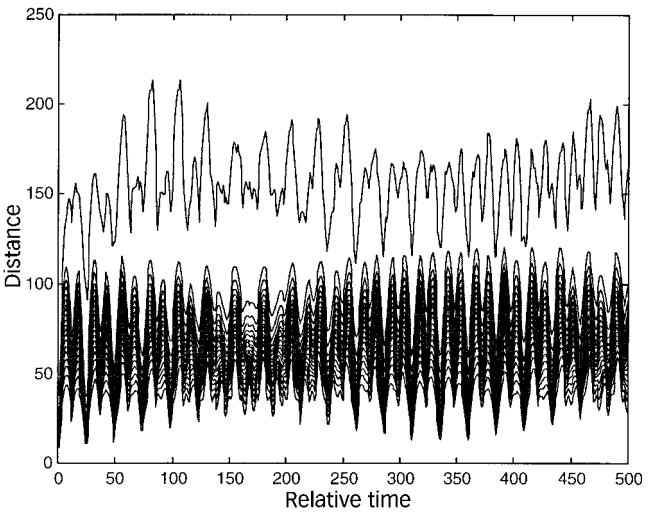

Figure 7 | Space-time separation plot of the Venice Lagoon level data. Lines of constant probability density of a point to be e-neighbour of the current plot if its temporal distance is dt (relative time) are shown. Probability densities are 1/20 to 1 with increments of 1/20 from bottom to top. Clear correlations are visible (calculations were carried out using the stp routine of the TISEAN package, *http://www.mpipks-dresden.mpg.de/~tisean*).

invariants calculation should not be used without a careful evaluation of the conditions for their applicability, and an examination of the consistency of the results obtained. This is more evident in the case of complicated time series from natural systems as in the case of the Venice Lagoon data. The first test to apply under such circumstances is to repeat the analysis increasing the length (number of points) of the signal used. Non-consistency in the results should warn of misleading conclusions.

Another useful technique to distinguish between low-dimensional dynamics and randomness is the space-time separation plot introduced by Provenzale *et al.* (1992). In this technique one usually draws lines of constant probability per time unit of a point to be a neighbour (distance less than ε) in the reconstructed phase space of the current point, when its time distance is δt. This helps in identifying temporal correlations inside the time series. In the case of power-law noises the only points with small spatial separation are dynamically near neighbours, i.e. the series is non-recurrent in phase space. As can be seen in Figure 7 there are clear correlations in the Venice Lagoon data. However, even if this preliminary analysis is encouraging, it should be emphasised that there is no simple test for indicating automatically and unequivocally the presence or absence of chaotic dynamics.

Many techniques for analysing time series assume in their application that the series under investigation is stationary. In the case of chaotic time series this means that the system has reached the attractor. However, one of the most difficult problems in nonlinear time series

analysis is the problem of nonstationarity. Unfortunately, although a precise asymptotic definition of stationarity exists, there is no clear and unambiguous method for determining stationarity in real finite time series (see for example Isliker & Kurths 1993; Manuca & Savit 1996 and Schreiber 1999, for some recent works).

Normally, the methods proposed in the literature are based on the estimation of a certain parameter, for example variance, power spectrum, etc., using different parts of the time series and studying whether the observed variations are found to be significant, i.e. outside the expected statistical fluctuations. The Venice Lagoon level time series is not stationary since the mean sea level from 1923 to 1998 has been continuously increasing, which is probably because of the fact that Venice is slowly moving down. However, we can assume that during the studied period the mean sea level is practically constant.

Related to noise is the problem of nonlinearity. Most of the methods discussed here are most appropriate where the data show strong and consistent nonlinear deterministic signatures. If a moderate amount of noise is present then it is possible that predictability will be limited. To test for nonlinearity we have applied the concept of surrogate data (Theiler *et al*. 1992). This method consists of generating an ensemble of 'surrogate' data sets, which are similar to the original time series but consistent with the null hypothesis which in our case is that the data has been created by a stationary Gaussian linear process, and of computing a discriminating statistic, which in our case is based on the forecasting error, for the original and for each of the surrogate data sets. The results for the Venice Lagoon data set show that the null hypothesis may be rejected at the 95% level of significance, since the prediction error of the data is found to be smaller than that of the surrogate data sets. However, the differences are significantly small and, hence, it seems that there is a considerable amount of noise present in the signal, as one may expect from such a system.

FINDING THE PHASE SPACE

State space reconstruction is the first step in nonlinear time series analysis of data from chaotic systems. Let us consider a system of d ordinary differential equations:

$$\frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{F}(\boldsymbol{x}(t)), \tag{9}$$

where $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_d(t)]$ in R^d and $\boldsymbol{F} = [F_1, F_2, \ldots, F_d]$ is a smooth (i.e. C^1) nonlinear function of $\boldsymbol{x}$.

A time series is a list of numbers, which are assumed to be measurements of an observable quantity over time, which, in the absence of noise, is related to the dynamic system by

$$s(t) = h(\boldsymbol{x}(t)). \tag{10}$$

The system on which the observable quantity is being measured is evolving with time. The phase space reconstruction problem is that of recreating states when the only information available is contained in a time series, i.e. how to go from scalar or univariate observations to the multivariable state space or phase space which is required to study the system? Typically, $\boldsymbol{F}$ and h are both unknown, so we cannot hope to reconstruct states in their original form. However, we may be able to reconstruct a state space that is equivalent to the original in the sense that differential properties are preserved.

Work by Takens (1981) and improvements by Sauer *et al*. (1991) have shown that if the dynamics are based on a d-dimensional Euclidean space, an embedding of the system can be obtained with a $2d+1$-dimensional reconstructed state space using derivatives or delay coordinates. The basic idea of this reconstruction is that if one has an orbit seen projected onto a single axis $s(t)$, then the orbit, which we presume came from an autonomous set of equations, may, by virtue of the projection, overlap with itself in the variables $s(t)$. There is no overlap of the orbit with itself in the true set of state variables according to the uniqueness theorems about the solution of autonomous differential equations. If we can unfold the orbit by providing independent coordinates for a multidimensional space made out of the observations, then we can undo the overlaps coming from the projection and recover orbits which are not ambiguous.

The currently used possibilities for state space reconstructions include (Breeden & Packard 1994), among others, delay coordinates, $\{s(t), s(t-T), s(t-2T), \ldots, s(t-(d_E-1)T\}$, derivative coordinates, $\{s(t), \dot{s}(t), \ddot{s}(t), \ldots\}$,

and global principal value decomposition. The method of reconstruction can make a big difference in the quality of the resulting coordinates, but in general it is not clear which method is the best (Casdagli *et al*. 1991). The lack of a unique solution for all purposes is due in part to the presence of noise and the finite length of the data set.

In this work delay coordinates have been used. A delay coordinate, often referred to as a lag, is simply the observed variable some time T in the past. Delay coordinates are easy to work with and can be effective for very high dimensional cases where it may not be practical to calculate the required number of derivatives. Most of the research on the state-space reconstruction problem has centred on the problems of choosing the time delay, T, and the embedding dimension, d_E, for delay coordinates.

Finding the time delay

The first step in phase space reconstruction is to choose an optimum delay parameter T. Different prescriptions have appeared in the literature to choose T but they are all empirical in nature and do not necessarily provide appropriate estimates:

First passes through zero of the autocorrelation function

In earlier works (Mees *et al*. 1987) it was suggested to use the value of T for which the autocorrelation function

$$C(T) = \sum_{n} [s(n) - \bar{s}][s(n+T) - \bar{s}] \tag{11}$$

first passes through zero which is equivalent to requiring linear independence.

First minimum of the average mutual information

Fraser & Swinney (1986) suggested using the average mutual information (AMI) function, $I(T)$, as a kind of nonlinear correlation function to determine when the values of $s(n)$ and $s(n+T)$ are sufficiently independent of each other to be useful as coordinates in a time delay vector, but not so independent as to have no connection which each other at all. For a discrete time series, $I(T)$ can be calculated as,

$$I(T) = \sum_{n,n+T} P(s(n),s(n+T)) \log_2 \left[\frac{P(s(n),s(n+T))}{P(s(n))P(s(n+T))} \right], \tag{12}$$

where $P(s(n))$ refers to individual probability and $P(s(n),s(n+T))$ is the joint probability density. Following the method developed by Abarbanel (1996), to determine $P(s(n))$ we simply project the values taken from $s(n)$ versus n back onto the $s(n)$ axis and form a histogram of the values. Once normalised, this gives us $P(s(n))$. For the joint distribution of $s(n)$ and $s(n+T)$ we form the two-dimensional histogram in the same way.

In general, the time lag provided by $I(T)$ is normally lower than the one calculated with $C(T)$, $T_{\mathrm{AMI}} \leq T_{\mathrm{correl}}$, and provides the appropriate characteristic time scales for the motion. Even though $C(T)$ is the optimum linear choice from the point of view of predictability in a least squares sense of $s(n+T)$ from knowledge of $s(n)$, it is not clear why it should work for nonlinear systems and it has been shown that in some cases it does not work at all.

Choosing the embedding dimension

The dimension, where a time delay reconstruction of the system phase space provides a necessary number of coordinates to unfold the attractor from overlaps on itself caused by projection, is called the embedding dimension, d_E. This is a global dimension to unfold the dynamics which can be different from the real dimension. Furthermore, this dimension depends on the time series measurement, and hence, if we measure two different quantities from some system, there is no guarantee that the d_E from time delay reconstruction will be the same from each of them.

The usual method for choosing the minimum embedding dimension is to compute some invariant of the attractor. By increasing the embedding dimension used for the computations, one notes when the value of the invariant stops changing. Since these invariants are geometric properties of the attractor, they become independent of d for $d \geq d_E$, i.e. after the geometry is unfolded.

In this work, we have used two methods: False Nearest Neighbours and the E1 & E2 method.

False Nearest Neighbours

The method of False Nearest Neighbours (FNN) was developed by Kennel *et al.* (1992). In this case, the condition of no self-intersection states that if the attractor is to be reconstructed successfully in R^d, then all the neighbour points in R^d should also be neighbours in R^{d+1}. The method checks the neighbours in successively higher embedding dimensions until it finds only a negligible number of false neighbours when increasing the dimension from d to $d+1$. This d is chosen as the embedding dimension.

It was found by Kennel *et al.* (1992) that if the data set is clean from noise, the percentage of false nearest neighbours will drop from nearly 100% in dimension one to strictly zero when d_E is reached. Further, it will remain zero from then on since the attractor is unfolded. If the signal is contaminated with noise (infinite dimension signal) we may not see the percentage of false nearest neighbours drop to near zero in any dimension. In this case, depending on the signal to noise ratio, the determination of d_E will degrade. In the case of a random number generator, the larger the number of data used, the sooner the percentage of false nearest neighbours rises to nearly 100% as we increase d.

E1 & E2 method

The method of FNN has some subjectivity for saying that a neighbour is false since two parameters have to be defined (Kennel *et al.* 1992). To improve this situation, Cao (1997) developed a similar method, which is based on evaluating the mean value of the distance between time-delay vectors, E1, and on another quantity, E2, that checks whether future values are independent of past values. If $E2 = 1$ for any embedding dimension then the signal is a stochastic signal (white or coloured noise). The E1 & E2 method depends only on the time delay, and the embedding dimension is calculated, as in the other methods, when the values of E1 and E2 reach saturation. Cao (1997) showed that the method does not strongly depend on how many points are available, provided there are enough; it can clearly distinguish between deterministic and stochastic signals; and it works well for time series from high-dimensional attractors.

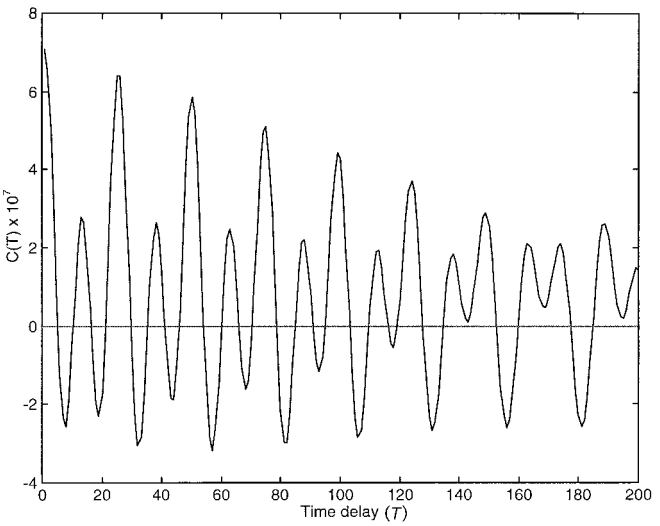

Figure 8 | Autocorrelation function for the Venice Lagoon data from 1980–1989. $C(T)$ has its first zero at $T=6$ hours, and this tells us when the measurements $s(n)$ and $s(n+T)$ are linearly independent.

Reconstructing phase space for Venice Lagoon data

In this work the 15-year time series of the Venice Lagoon level, as well as the residual term, has been divided into two parts: the first one consists of 10 years, from 1980 to 1989, and the second contains the past 5 years. These time series have been analysed independently to check the validity and coherence of the results for the reasons previously discussed. Furthermore, some tests have been performed on the complete time series.

Finding the time delay

The water level data at the Venice Lagoon from 1980 to 1989 was used to calculate the autocorrelation function, Equation (11), and the average mutual information function, Equation (12). As can be seen in Figure 8 the first zero of the autocorrelation function for the level data occurs at $T=6$ hours, whereas for the residual level data (see Figure 9), the intersection through zero occurs at 609 hours (25.4 days). A time delay reconstruction of the

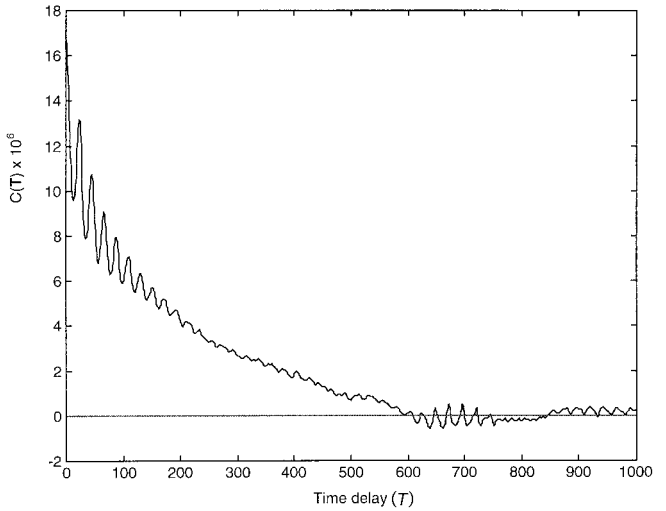

Figure 9 | Autocorrelation function for the residual data from 1980–1989. $C(T)$ crosses through zero at 609 hours (>25 days). This is one of the cases where the autocorrelation function fails to find an adequate time delay.

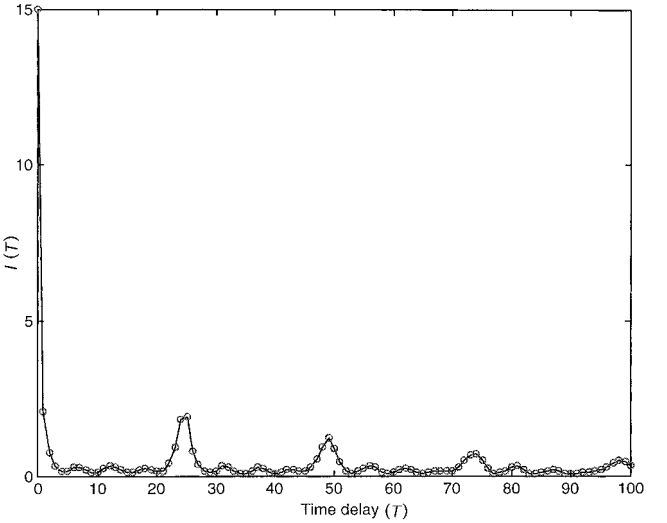

Figure 10 | Average mutual information function for the Venice Lagoon data from 1980–1989. $I(T)$ has its first minimum at $T=4$ hours.

phase space using this time as the lag would reveal little since this is clearly too long for any dynamic correlations in this system to persist. An identical result was reported by Abarbanel (1996) analysing the chaotic volume fluctuations of the Great Salt Lake.

Similar results were obtained using the data from 1990–1994. In this case $T=6$ hours for the level data and there is no intersection through zero for the residual data after as long as 1000 hours (41.7 days).

Figures 10 and 11 show the average mutual information for the above mentioned data set. In this case $T=4$ and 8 hours, respectively. Similar results were obtained using the data from 1990–1994 and using all the time series data, i.e. 1980–1994. In the first case, using the AMI function, $T=5$ and 6 hours for the level and residual data, respectively, whereas with the complete time series $T=4$ and 10 hours, respectively.

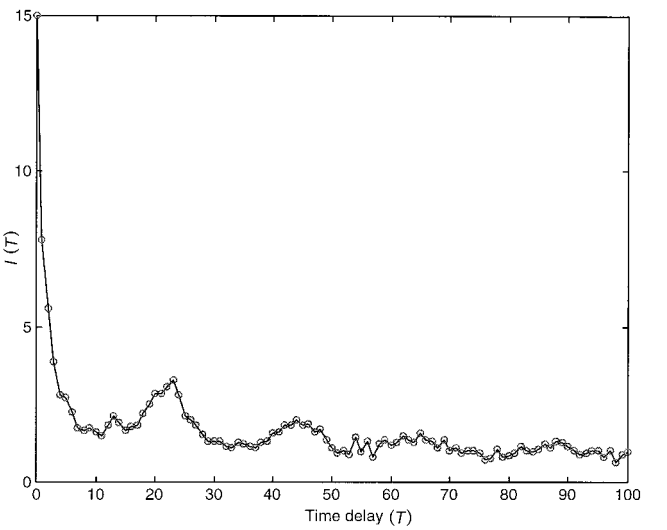

Figure 11 | Average mutual information function for the residual Venice Lagoon data from 1980–1989. $I(T)$ has its first minimum at $T=8$ hours.

Choosing the embedding dimension

Using a time lag $T=6$ hours, the percentage of false nearest neighbours for the water level data from 1980–1989 shows a sharp drop close to zero at $d_E=7$, after which the percentage of false neighbours remains approximately constant, see Figure 12. This provides the evidence that we are dealing with a low dimensional system. The strength of this conclusion is enhanced when similar results are obtained using the complete time series data, i.e. 1980–1994, or the second part, i.e. 1990–1994. In contrast, when carrying out the same procedure for the residual level data, see Figures 13 and 14, it is possible to

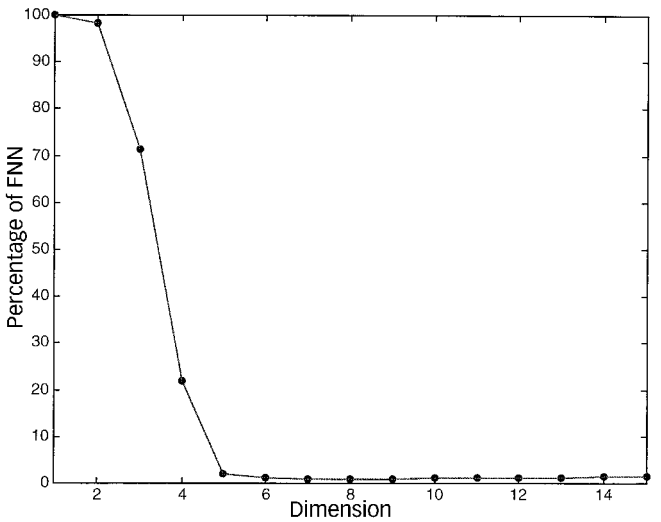

Figure 12 | False nearest neighbours for the water level data at Venice Lagoon 1980–1989, d_E=7.

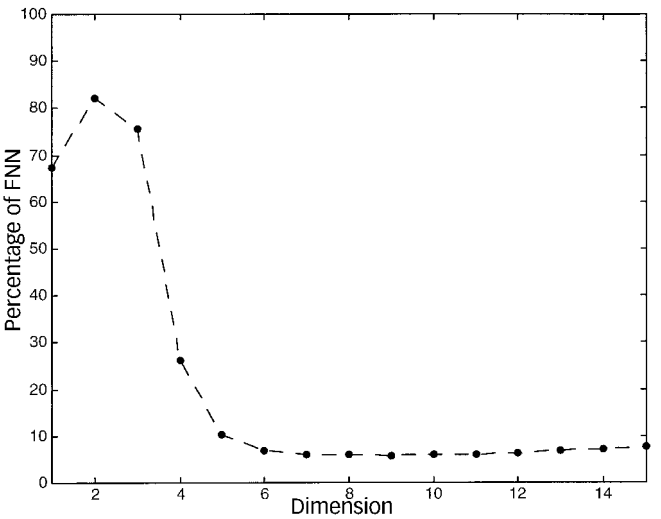

Figure 13 | False nearest neighbours for the residual data at Venice Lagoon 1980–1989.

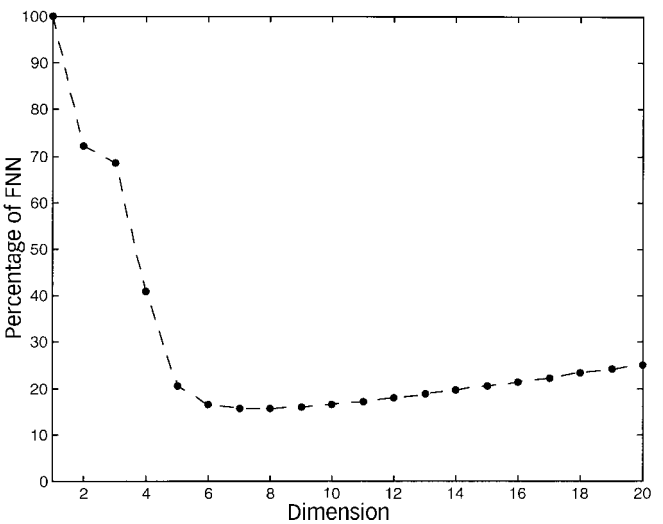

Figure 14 | False nearest neighbours for the residual data at Venice Lagoon 1980–1994.

observe the characteristic behaviour of noisy data with the percentage of false nearest neighbours increasing with the dimension. This fact is accentuated when more data points are included; this is also a generic characteristic of noisy time series data (see Figure 14).

The method E1 & E2 has been also applied to the water level and residual data, respectively, using the same time delays as in the FNN method. The results are shown in Figures 15 and 16. In the first case, E1 and E2 stabilise around an embedding dimension of 8, whereas in the case of the residual time series $d_E \geq 13$. In this case, the results suggest that the residual time series is not stochastic, as it seems from the FNN method. Furthermore, in principle, this method is also able to detect between coloured noise and chaotic signals. Similar results were obtained using all the time series data.

It has long been known that the filtering of chaotic time series data may change the dynamic properties of interest. In fact, Badii *et al.* (1988) discovered that an infinite impulse response (IIR) filter can affect the estimation of the dimension of the original attractor and of invariant measures such as fractal dimension and Lyapunov exponents. Badii *et al.* (1988) found that when the contraction rates associated with the filter were too slow then the reconstructed attractor can have a higher fractal dimension than the original. This can be seen in terms of an increase in the dimension of the system due to the additional filter dynamics. However, for FIR filters mathematical theorems exist which say that the delay coordinate properties are unchanged by the filter (see Ott *et al.* 1994). Of course, the practical use of this information must be tempered with the knowledge that an IIR filter is a limiting case of a FIR filter. In a recent work Davies

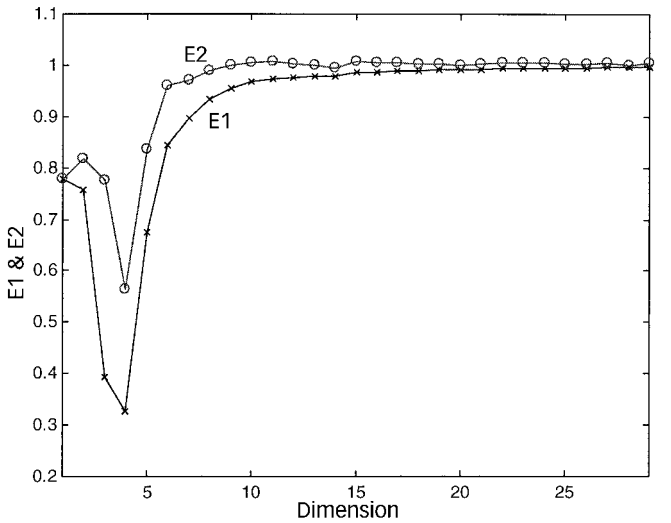

Figure 15 | E1 & E2 method for the level data at Venice Lagoon 1980–1994.

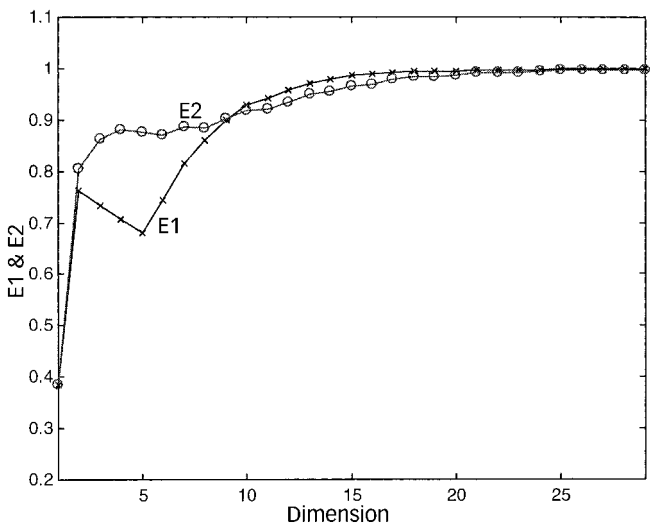

Figure 16 | E1 & E2 method for the residual data at Venice Lagoon 1980–1994.

(1997) proved that, in theory, one may obtain the same Lyapunov exponents from the filtered dynamics.

In the case of Venice Lagoon data, the residual level is obtained by subtracting a sum of sine functions, each one with its own periodicity, which correspond to the relative motion of the earth, moon and sun. This is equivalent to performing multiple IIR filtering, and hence the residual level data may have different properties from the original attractor under study. Furthermore, this treatment may increase the noise:signal ratio and hence it may make it difficult to obtain reliable invariants.

Choosing the dynamic dimension for the Venice Lagoon data

Once one has determined the global number of dimensions required to unfold the attractor, there remains the problem of the number of dynamic degrees of freedom, d_L, which are active in determining the evolution of the system as it moves around the attractor. To calculate this dynamic dimension we have used the method proposed by Abarbanel & Kennel (1993) which consists of evaluating the percentage of local false nearest neighbours.

Using the same idea as the FNN method, Abarbanel & Kennel (1993) proposed a method to study the local structure of the phase space to see if locally one requires fewer dimensions than d_E to capture the evolution of the orbits as they move on the attractor. Their approach was to work in a dimension, $d_L \leqslant d_E$, large enough to ensure that the attractor has been unfolded. In this space, they studied for some data point $\boldsymbol{y}(k)$ what subspace of dimension $d_L \leqslant d_E$ allows the construction of accurate local neighbourhood to neighbourhood maps of the data on the attractor. In fact, for a specified number of neighbours, N_B, of $\boldsymbol{y}(k)$, they provided a local rule for calculating how these points evolve in one time step into the same N_B points near $\boldsymbol{y}(k+1)$. When the percentage of bad predictions becomes independent of d_L and is also insensitive to the number of neighbours N_B, it is possible to say that the correct local dimension for the active degrees of freedom has been identified.

In the case of Venice Lagoon level data, the percentage of bad predictions seen in Figure 17 becomes independent of the number of neighbours N_B and of the local dimension at $d_L = 8$, telling us that this attractor may be adequately described by eight degrees of freedom. Similar results are obtained using the second portion of the time series, i.e. 1990–94, or the whole data, i.e. 1980–1994. This means that models for simulating the dynamic behaviour of water level at Venice Lagoon should have local eighth-dimensional dynamics regardless of the dimensions of the overall space in which the model is built.

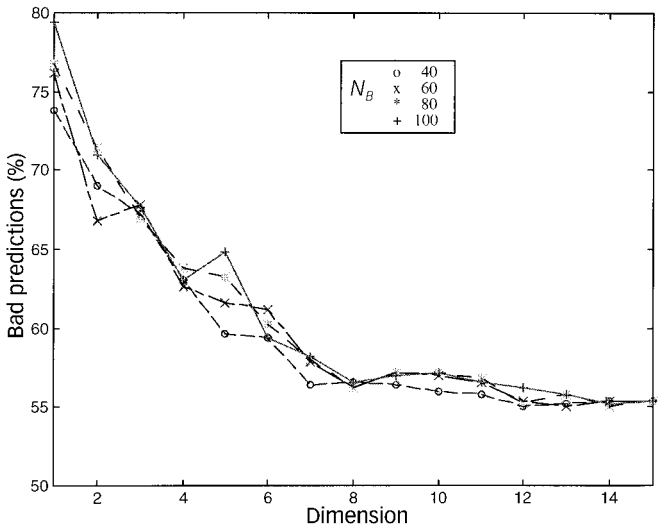

Figure 17 | Local false nearest neigbours for water level data at Venice Lagoon from 1980–1989. From this view d_L=8 might be chosen (recall d_E=8).

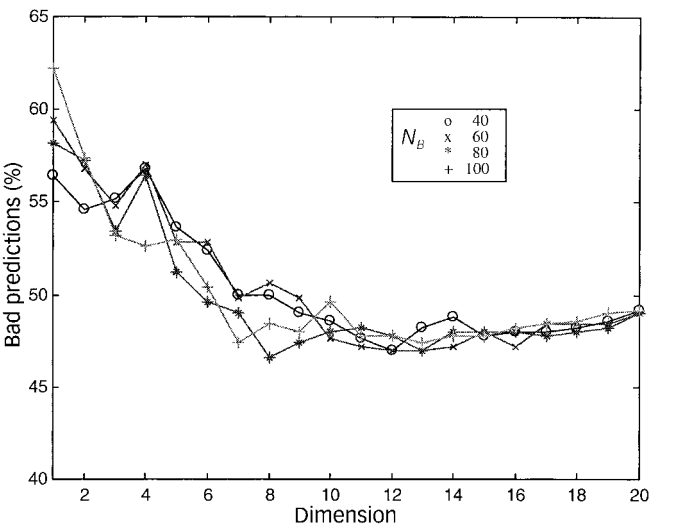

Figure 18 | Local false nearest neighbours for residual level data at Venice Lagoon from 1980–1994. From this view d_L=10 might be chosen.

For the residual data we found the percentage of bad predictions becomes independent of the number of neighbours N_B and of the local dimension at $d_L = 10$, see Figure 18.

INVARIANT CHARACTERISTICS OF THE DYNAMICS

Three classes of tools now exist for the analysis of data generated by chaotic dynamic systems. These involve metric, dynamic and topological invariants. Metric methods depend on the computation of various fractal dimensions or scaling functions (Grassberger & Procaccia 1983). Dynamic methods rely on the estimation of local and global Lyapunov exponents and a Lyapunov dimension, as well as on entropy (Eckmann & Ruelle 1985). Finally, topological methods involve determination of specific topological invariants of the attractor as relative rotation rates for the unstable periodic orbits embedded in the attractor (Gilmore 1998).

In this work we have used the dynamic invariants to characterise the Venice Lagoon data, i.e. Lyapunov exponents and a Lyapunov dimension. Lyapunov exponents quantify how orbits on the attractor move apart (or together) under the evolution of the dynamics. They are invariant under the evolution operator of the system and thus are independent of changes in the initial conditions of the orbit, and they are independent of the coordinate system in which the attractor is observed. This means that it is possible to evaluate them reliably in the reconstructed phase space made of time delay vectors. Thus it is possible to evaluate them from experimental data.

Global Lyapunov exponents

Lyapunov exponents describe the action of the dynamics defining the evolution of trajectories. Given a continuous dynamic system in d-dimensional phase space, i.e. Equation (9), it is possible to monitor the evolution of an infinitesimal d-sphere of initial conditions. This d-sphere will become a d-ellipsoid due to the locally deforming nature of the flow, see Figure 19. The jth one-dimensional Lyapunov exponent, λ_j, is then defined in terms of the length of the ellipsoidal principle axes at time t, $p_j(t)$, as (Wolf *et al*. 1985):

$$\lambda_j = \lim_{t\to\infty} \frac{1}{t} \log_2 \frac{p_j(t)}{p_j(0)} \,, j = 1, \ldots, d \,. \tag{13}$$

The Lyapunov exponent monitors the behaviour of two closely neighbouring points in a direction of the phase

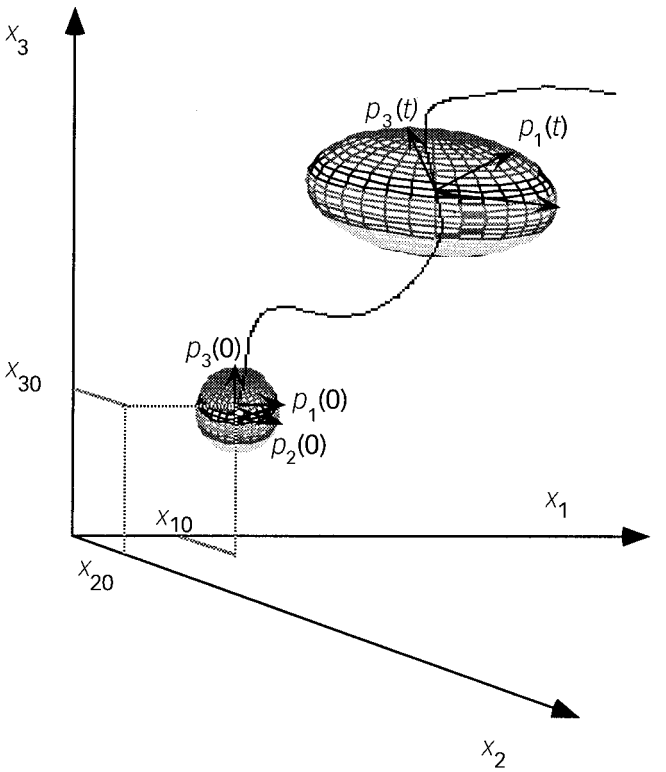

Figure 19 | A schematic representation of the evolution of a dynamic system in the phase space.

space as a function of time. If the points expand away from each other, the Lyapunov exponent will be positive, if they converge, the exponent becomes negative, if the two points stay the same distance apart, the exponent stays near zero. Since the orientation of the ellipsoid changes continuously, the directions associated with a given exponent vary too. Hence, there is not a fixed direction associated with a given exponent. Normally λ_j are ordered with respect to their magnitude, i.e. $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_d$ and the set of all λ_j are called Lyapunov spectra. In general it is possible to use other bases of the logarithm. If base 2 is used, the exponents are measured in bits of information for second (continuous system) or for iteration (discrete system).

The Lyapunov exponents give us a sense of dimension. When the system evolves, the linear extent of the ellipsoid grows (reduces) as $2^{\lambda_1(t)\cdot t}$; the area defined by the first two principle axes grows (reduces) as $2^{(\lambda_1(t)+\lambda_2(t))\cdot t}$; the volume defined by the first three principle axes grows (reduces) as $2^{(\lambda_1(t)+\lambda_2(t)+\lambda_3(t))\cdot t}$; and so on. In general, in a phase space of a higher dimension, the calculation of the volume will be:

$$V(t) = 2^{(\lambda_1(t)+\lambda_2(t)+\ldots+\lambda_d(t))t}\, V(0)\,. \tag{14}$$

Lyapunov dimension

As shown above, for dissipative systems the sum of all exponents is negative, so somewhere there must be a combination of exponents which can be associated with a volume in phase space which neither grows nor shrinks. Kaplan & Yorke (1979) suggested that this be used to define a Lyapunov dimension as:

$$D_L = k + \frac{\sum_{i=1}^{k} \lambda_i}{|\lambda_{k+1}|}, \tag{15}$$

where $\sum_{i=1}^{k} \lambda_i > 0$ and $\sum_{i=1}^{k+1} \lambda_i < 0$.

Invariant characteristics for the Venice Lagoon level data

The local false neighbours for the water and residual level data become nearly independent of both parameters (N_B and d) near $d_L = 8$ and $d_L = 10$, respectively. After that they start to fluctuate, see Figures 17 and 18. This tells us that we should use eight differential equations to model the dynamics of the level of the Venice Lagoon for prediction, and also that this should be the number of true Lyapunov exponents we can expect from these types of systems. Given that the maximum Lyapunov exponent is related to the predictability of the system it also gives an indication of the maximum predictability we can expect from any model we may make.

Table 1 shows the eight computed local Lyapunov exponents for the water level at Venice Lagoon. The Table shows that three Lyapunov exponents are positive, one is close to zero and the others are negative. The Lyapunov exponents tell us, on average, around the attractor, how

Table 1 | Calculated Lyapunov exponents for the Venice Lagoon level and residual data.

	Level data		Residual data	
Lyapunov exponents	Forwards	Backwards	Forwards	Backwards
1	0.1023	0.0788	0.4959	0.4827
2	0.0710	0.0397	0.3942	0.3736
3	0.0336	0.0157	0.2776	0.2751
4	0.0062	– 0.0071	0.1866	0.1649
5	– 0.0262	– 0.0277	0.0488	0.0364
6	– 0.0916	– 0.0808	– 0.1083	– 0.1371
7	– 0.1962	– 0.1920	– 0.3739	– 0.3784
8	– 0.5930	– 0.5267	– 0.9656	– 1.0460
Sum	– 0.5930	– 0.5267	– 0.0446	– 0.2288
D_L	6.58	6.01	7.95	7.78

well we can predict the evolution of the system m steps ahead of wherever we are. Secondly, if the data we are analysing comes from a set of differential equations, then one of the exponents must be zero (Eckmann & Ruelle 1985). Thirdly, the values of λ_i, at their limit, become invariant and they provide us with a tool to characterise our system. Similar results have been obtained using the two parts of the time series non-sequentially and the complete set. Table 1 also gives the total sum of the Lyapunov exponents, which is negative, as expected for a dissipative system. Furthermore, the Lyapunov dimension, Equation (15), is calculated in the last row which gives values between 6.01 and 6.58.

The determination of the local dimension of the dynamics by the local false neighbours test, or forward-backward local Lyapunov exponents, tells us how many dimensions we should use to model the dynamics for purposes of prediction and control. It also tells us how many true Lyapunov exponents we should evaluate for the system. Since Lyapunov exponents are invariants of the attractor, they serve to characterise the system as well as give us an indication of the predictability of any model we might make. The largest Lyapunov exponent gives an indication of how far into the future reliable predictions can be made, and the dynamic dimension gives an indication of how complex a model for making predictions must be. Since the predictability time is about τ_s/λ_1, τ_s being the sampling time, this means that models for the Venice Lagoon level data could allow predictions that lie between 8 and 13 hours ahead. Hence we can expect that for any given prediction time in exceedence of this range, the intrinsic instabilities of the system will make predictions highly unreliable.

We have tried to recover the same invariants (Davies 1997), i.e. Lyapunov exponents, from the residual data but, as can be seen in Table 1, there are differences between both time series and, hence, the filtering has affected the structure of the attractor. In this case the Lyapunov exponents are higher and the predictability is reduced to 2 hours.

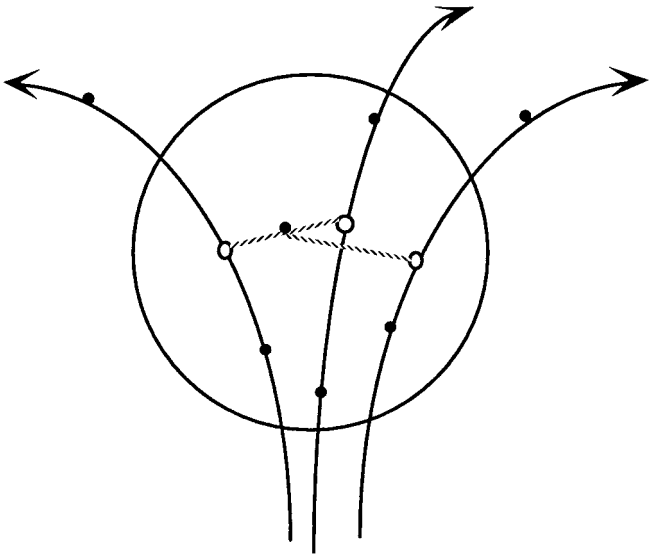

Figure 20 | The point for which we want to predict its dynamic behaviour is shown together with its nearest neighbours in reconstructed state space. By interpolating the more appropriate neighbours it is possible to predict the trajectory and then reconstruct the time series.

FORECASTING HIGH WATERS AT VENICE LAGOON USING CHAOS THEORY TECHNIQUES

Having characterised the dynamic behaviour of the water level at Venice Lagoon as well as the predictability limits, the idea is to use this information to model and predict its evolution.

The idea for making short-term predictions in chaotic time series was first introduced by Farmer & Sidorowich (1987). Since we have information on the temporal evolution of orbits $\boldsymbol{y}(k)$, and these orbits lie on a compact attractor in phase space, each orbit has near it a whole neighbourhood of points in phase space which also evolve under the dynamics to new points, see Figure 20. We can combine this knowledge of the evolution of whole neighbourhoods of phase space to enhance our ability to predict in time by building local or global maps with parameters $\boldsymbol{a}$: $\boldsymbol{y}\rightarrow\boldsymbol{F}(\boldsymbol{y},\boldsymbol{a})$, which evolve each $\boldsymbol{y}(k)\rightarrow\boldsymbol{y}(k+1)$. Using the information about how neighbours evolve, we use phase-space information to construct the map, and then we can use the map to extend the evolution of the last points in our observations forward in time.

Table 2 | Root mean square (in cm) prediction error obtained using local linear and quadratic map models for the Venice Lagoon level data.

Number of steps ahead (h)	P=1	P=2
1	8.86	9.59
4	12.86	13.54
12	13.70	14.13
24	15.20	15.57
28	17.72	17.90

There are different approaches, see Casdagli (1989), to find a predictor $\boldsymbol{F}$. These approaches can be divided, mainly, into local and global models. In local models, one considers maps from local neighbourhood to local neighbourhood, whereas in global models one tries to fit all the data points in the reconstructed phase space at once.

Local models

Local models map local-neighbourhood to local-neighbourhood in the reconstructed phase space. We start with a specified local functional form for the dynamics $\boldsymbol{x}\rightarrow\boldsymbol{F}(\boldsymbol{x},k)$ in the neighbourhood of the observed point $\boldsymbol{y}(k)$:

$$\boldsymbol{F}(\boldsymbol{x},k) = \sum_{m=1}^{M} \boldsymbol{c}(m,k)\phi_m(\boldsymbol{x})\,. \tag{16}$$

This $\boldsymbol{y}(k)$ evolves to $\boldsymbol{y}(k+1)$ through:

$$\boldsymbol{y}(k+1) = \boldsymbol{F}(\boldsymbol{y}(k),k) = \sum_{m=1}^{M} \boldsymbol{c}(m.k)\phi_m(\boldsymbol{y}(k))\,. \tag{17}$$

These $\phi_m(\boldsymbol{x})$ functions can be polynomials, radial basis functions or other types of functions. The discussion of which type of functions to use is related to the problem being treated and the quantity of data points available. To determine the coefficients in the model, the N_B nearest neighbours are located and the error minimised. This is a

linear least-squares problem which can be solved efficiently using standard techniques. In this way an optimal polynomial (or other function) predictor of a given degree d is obtained. When the $\boldsymbol{c}(m,k)$ coefficients have been determined, it is possible to construct a lookup table for interpolation and hence, there will be a local model associated with each observed $\boldsymbol{y}(k)$ point on the attractor. To predict ahead from a new point $\boldsymbol{z}(0)$ we search through the $\boldsymbol{y}(k)$ to find one nearest to $\boldsymbol{z}(0)$, $\boldsymbol{y}(J)$. We now look up the model local to $\boldsymbol{y}(J)$. This is $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{J})$ and it should be valid as an interpolating function in the neighbourhood of $\boldsymbol{y}(J)$ and $\boldsymbol{z}(0)$. Next we evaluate $\boldsymbol{F}(\boldsymbol{z}(0),J)$, and this gives us the next point on the orbit which starts with $\boldsymbol{z}(0)$ as initial condition: $\boldsymbol{z}(1) = \boldsymbol{F}(\boldsymbol{z}(0),J)$. Next find the nearest neighbour of $\boldsymbol{z}(1)$, call it $\boldsymbol{y}(K)$, and look up the required local map $\boldsymbol{F}(\boldsymbol{x},K)$ to proceed to $\boldsymbol{z}(2) = \boldsymbol{F}(\boldsymbol{z}(1),K)$. This procedure is called iterative forecasting (Abarbanel 1996).

The root mean square error in this forecast should scale, in going from $\boldsymbol{z}(1)$ to $\boldsymbol{z}(L)$ in L steps, approximately (Farmer & Sidorowich 1987; Casdagli 1989) as:

$$N^{-(P+1)/d_L}\, e^{L\cdot h}, \tag{18}$$

where N is the number of points learned, P is the maximum order of the polynomials used and h is the metric entropy which is equal to the sum of the positive Lyapunov exponents. In the most favourable situation, the scaling law will be proportional to λ_1 which is the largest Lyapunov exponent:

$$N^{-(P+1)/d_L}\, e^{L\cdot\lambda_1}, \tag{19}$$

Global models

The collection of local maps (polynomial functions or other) form a model which is useful over the whole attractor. The shortcomings of such a local model are its discontinuities from neighbourhood to neighbourhood and the large number of adjustable parameters. For polynomial models of order M in d_L local dimensions we have approximately d_L^M parameters at each time step. At the same time it would be nice to have a relatively simple continuous model describing the whole collection of data. In a sense, this is similar to what has been described for local models but the data fitting is applied over the whole attractor. However, in this case, owing to the extremely large number of data points we need to use rather high-order polynomials, with the associated stability problems. Brown (1993) has suggested an alternative approach based on orthogonal polynomials whose weights are determined by the invariant density on the attractor, whereas Casdagli (1989) used radial basis function and neural networks.

An overview and a comparison between different global nonlinear modelling techniques for modelling and forecasting hydrological time series can be found in Babovic (1998).

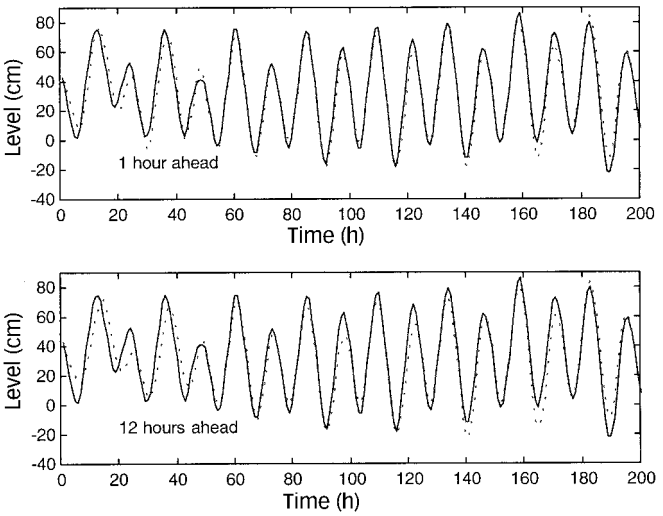

Figure 21 | Observed (continuous) and predicted (dashed) water level values from Venice Lagoon, 1 and 12 hours forecasting. The predictions were made using local linear polynomial predictors whose coefficients were learned from 68,000 data points. These were embedded in $d_E=8$ using a $d_L=8$ dimensional model.

Forecasting water level at Venice Lagoon

Table 2 shows the root mean square error for local linear, quadratic and cubic polynomial prediction functions. The computation was done using 68,000 data points in a reconstructed phase space $d_E = 8$ and local maps with $d_L = 8$. Two thousand different initial conditions were

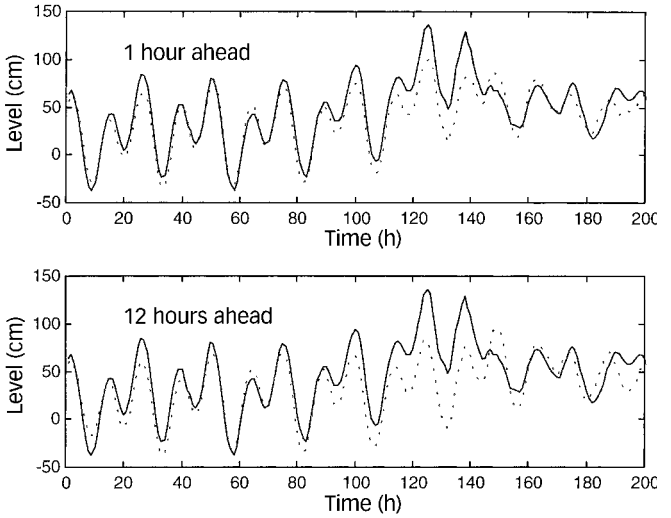

Figure 22 | Observed (continuous) and predicted (dashed) water level values from Venice Lagoon, 1 and 12 hours forecasting. The predictions were made using local linear polynomial predictors whose coefficients were learned from 68,000 data points. These were embedded in $d_E=8$ using a $d_L=8$ dimensional model.

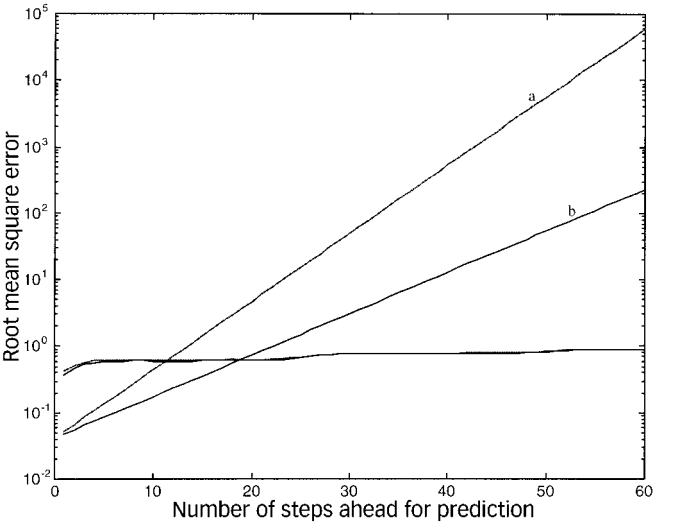

Figure 23 | Standard errors for local linear and local quadratic map models for the Venice Lagoon water level data in units of the size of the attractor. The coefficients for the polynomials were learned from 68,000 data point. The slopes of a and b represent the scaling laws given by Equations (18) and (19), respectively.

examined, and we have displayed the average over these starting locations. As can be seen, surprisingly, the predictions using quadratic predictors are worse than, or equal to, those using linear predictors. Casdagli (1989) pointed out that for complicated systems with a large number of data points there are no advantages in using quadratic predictors over linear predictors.

Different specific predictions for the Venice Lagoon level are shown in Figures 21 and 22. In both figures forecasting results 1 and 12 hours ahead are presented. As can be seen from Figure 21 the evolution of the water level data is well predicted for the two cases. On the other hand, in Figure 22 the prediction fails for extraordinarily high tides. The fact that predictions are similar at 1 and 12 hours is in contradiction with the findings in the calculation of the Lyapunov exponents, since an exponential decrease in our forecasting ability is expected.

Nonlinear prediction methods have been used to distinguish between deterministic chaos and uncorrelated (white) noise added to periodic signals (Sugihara & May 1990). The idea behind using nonlinear prediction as a signature of chaos is simple. Chaotic systems obey certain rules. The limited predictive power of chaotic dynamic systems is because they are sensitive to initial conditions and because we cannot have infinite precision measurements. This property can be used to differentiate between chaos and additive uncorrelated noise. Additive noise produces a fixed amount of error regardless of the prediction time, as has been demonstrated. From the analysis of the water level data of the Venice Lagoon, this decrease of predictive power with time steps into the future we are forecasting has been observed, see Figure 23. However, this decrease does not scale as an exponential function of the largest Lyapunov exponent, as expected for a chaotic signal. In our case it seems that as we have a superimposed predictable periodic signal then the limiting case for our predictions should be the mean square error of the difference between this periodic signal and the water level data. In fact, calculating the root mean square (in cm) prediction error obtained using only the periodic signal, the error would be 25.9 cm, which is close to the values we obtain for predictions a long time ahead.

We have also fitted a global polynomial function (second-order polynomials) to the reconstructed phase space. As the embedding dimension is 8 then the number of adjustable parameters is 45 (Casdagli 1989). The model

was also trained using 68,000 data points by minimising the error: $\sum[\boldsymbol{y}(k+1) - \boldsymbol{f}_p(\boldsymbol{y}(k))]^2$. The root mean square (in cm) prediction error obtained with a prediction of 1 hour ahead is 10.84 cm, i.e. 1 or 2 cm higher than those obtained with local predictors.

NONLINEAR NEURAL MODELS

In order to compare the prediction results obtained from nonlinear time series analysis with other nonlinear techniques, two different nonlinear models based on multilayer neural networks have been developed. The models are built up with the purpose of multistep prediction, i.e. starting from the information at instant k, $x(k-m)$, ..., $x(k)$, the goal is to predict the behaviour of the time series in the future, $x(k+1)$, ..., $x(k+h+1)$, where h is a natural number named prediction horizon.

First neural approach: predicting the time interval $[k+1, k+h+1]$

The first nonlinear neural model consists of approximating the function $\boldsymbol{F}$ appearing in Equation (1) by a multilayer feedforward neural network (Rumelhart *et al*. 1986) as follows:

$$\tilde{x}(k+1) = \tilde{\boldsymbol{F}}(x(k), \ldots, x(k-d), W_1), \tag{20}$$

where W_1 is the parameter set of the model, which is obtained using the backpropagation algorithm (Rumelhart *et al*. 1986). The update of the parameter set is based on the local difference between the measured and predicted values at the current instant, i.e.

$$e(k+1) = \tfrac{1}{2} \cdot (x(k+1) - \tilde{x}(k+1))^2 . \tag{21}$$

When the model given by Equation (20) has to predict the behaviour of the time series in the future, i.e. along the interval $[k+1, k+h+1]$, its structure has to be modified. The model has to be used in a recurrent form because the predictive network output must be fed back as an input for the next prediction. If the aim is to predict h sampling times in the future, the input layer of the network is formed by a group of h neurones that memorise previous network outputs, and the remaining neurones in the input layer receive the original or measured time series data. Thus, the predicted model outputs along the interval $[k+1, k+h+1]$ are given by the following equations:

$$\tilde{x}(k+1) = \tilde{\boldsymbol{F}}(x(k), \ldots, x(k-m), W_1) \tag{22}$$

$$\tilde{x}(k+2) = \tilde{\boldsymbol{F}}(\tilde{x}(k+1), x(k), \ldots, x(k-m+1), W_1) \tag{23}$$

...

$$\tilde{x}(k+h+1) = \tilde{\boldsymbol{F}}(\tilde{x}(k+h), \ldots, \tilde{x}(k+1), x(k), \ldots, x(k-m+h), W_1) . \tag{24}$$

Second neural approach: predicting the prediction horizon $x(k+h+1)$

The structure of the second neural model consists of using a multilayer feedforward network to predict, directly, the time series value at instant $k+h+1$ from the information available at the current instant k, $x(k)$, ..., $x(k-m)$, instead of using the immediate d previous values as in the first model, see Equation (20). In this case, the nonlinear model is written as follows:

$$\tilde{x}(k+h+1) = \tilde{\boldsymbol{F}}(x(k), \ldots, x(k-m), W_2), \tag{25}$$

where h is the prediction horizon. The set of parameter W_2 is updated using the backpropagation algorithm and following the negative gradient direction of the error measured at instant $k+h+1$, i.e.

$$e(k+h+1) = \tfrac{1}{2} \cdot (x(k+h+1) - \tilde{x}(k+h+1))^2 . \tag{26}$$

Comparative study between both neural models

The neural approaches presented above are two different alternatives when a multistep prediction problem is formulated.

The main disadvantage of the first model when it is used for multistep prediction is that the parameter set has been obtained with the purpose of one-step prediction, i.e. to minimise the local errors given by Equation (21). During the training phase, the model captures the relation between the actual observations of the original time series, $x(k), \ldots, x(k-m)$ and the next sampling time, $x(k+1)$. However, when the model is acting as a multistep prediction scheme a group of the input neurones receives the earlier approximated values, see Equations (22)–(24). Hence, errors which occur for the predicted output network at some instant may be propagated to future sampling times and the quality of the approximations at next instants may be affected by those errors.

The number of the predictive network outputs feed back as the input of the network is given by the prediction horizon value. Therefore, the capability of the first neural model to predict the future may decrease when the prediction horizon is increased.

The second neural approach directly provides the prediction of the time series at instant $k+h+1$ from the information at instant $k+1$, see Equation (26). Hence, the inputs to the network when the model is used to predict the future are measured time series values and no outputs of the network must be fed back into the input network. Thus, the problem concerning the propagation of errors disappears when the second model is used as a nonlinear multistep prediction scheme.

A disadvantage of the second model is relative to the structure of the model. As previously mentioned, in this case the model predicts directly the time series value at instant $k+h+1$. The inputs to the model may not contain sufficient information about the time series in order to predict that instant. That is, the input vector, $x(k), \ldots, x(k-m)$, may be very distant in the time from the prediction horizon, $k+h+1$, and it may not have any relation with that instant. In this case, the second neural model cannot be used to predict the future. This structure only has sense when a relation exists between the information available at the current instant and the prediction horizon.

On the other hand, it is necessary to point out that the second model, Equation (26), has only been prepared to predict the time series value at instant $k+h+1$, while the first neural model can be used to predict each sampling time until the prediction horizon is reached. Thus, if the purpose is to predict the overall prediction interval $[k+1, k+h+1]$, h different neural models of the second approach must be trained.

Forecasting water level at Venice Lagoon using NAR models

In this section, the nonlinear neural approaches have been used to predict the dynamic behaviour of the water level in the Venice Lagoon in the future.

As the analysis of the power spectrum of the water level time series indicates the existence of periodicities related to the diurnal and semidiurnal tides with periods of 12 and 24 hours, respectively, it has been decided to consider NAR models owning information about those periodicities in the time series. Furthermore, as from the nonlinear time series analysis we know the predictability then, we considered as a first approach NAR models in which the current value $x(k+1)$ is expressed as a function of the 24 previous values of the time series, which corresponds to the values measured 24 hours earlier, i.e.

$$x(k+1) = \boldsymbol{F}(x(k), \ldots, x(k-24)). \tag{27}$$

However, the model given by Equation (27) has a large number of input variables. The immediate question that arises is whether the model may be simplified. After training NAR models with a different number of input variables, it has been observed that some input variables can be eliminated because they did not provide the model with useful information and because they did not provide better predictions of the water level in the Venice Lagoon. As a consequence, the NAR models used in this work to predict the dynamic behaviour of the water level have the following structure:

$$\begin{aligned} x(k+1) = \boldsymbol{F}(&x(k), x(k-4), x(k-8), x(k-12), \\ &x(k-16), x(k-20), x(k-24)). \end{aligned} \tag{28}$$

The functional $\boldsymbol{F}$ has been approximated by multilayer feedforward networks. The networks had one hidden layer and the number of the neurones in this layer was fixed at seven.

Table 3 | Root mean square prediction error (in cm) using NAR models for the Venice Lagoon level data.

Number of steps ahead (h)	First neural model	Second neural model
1	3.30	3.30
4	9.75	9.55
12	12.38	11.38
24	13.15	11.64
28	16.91	15.74

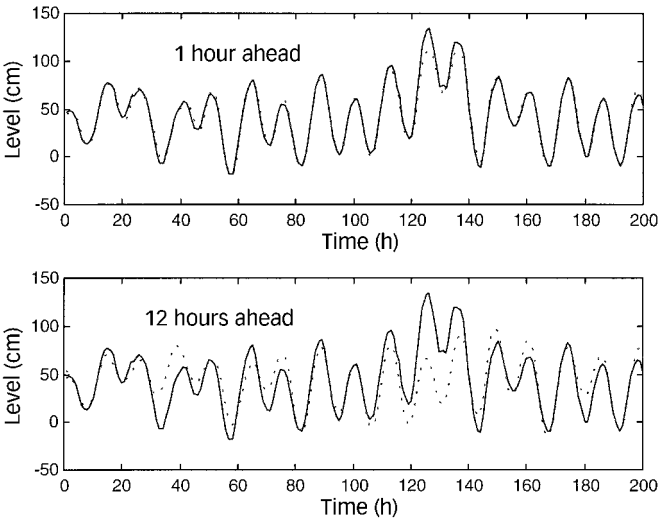

Figure 24 | Observed (continuous) and predicted (dashed) water level values from Venice Lagoon. Model 1 neural network predictions, 1 and 12 hours forecasting.

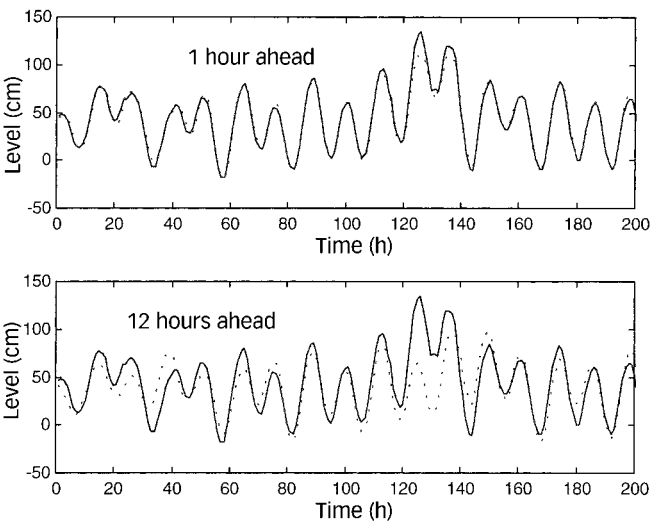

Figure 25 | Observed (continuous) and predicted (dashed) water level values from Venice Lagoon. Model 2 neural network predictions, 1 and 12 hours forecasting.

For the first model structure, the parameters have been determined to approximate the immediate sampling time, $x(k+1)$; after that, the model has been used to predict the water level in the Venice Lagoon for several prediction horizons ($h=1$, $h=4$, $h=12$, $h=24$ and $h=28$) using the recurrent structure presented above. In the second approach, five neural networks have been trained to approximate the water level at instants $k+1$, $k+4$, $k+12$, $k+24$, $k+28$, respectively, using the structure given by Equation (26).

To train the neural models, data of the water level in the Venice Lagoon corresponding to normal situations as well as abnormal situations ('high waters' phenomena) have been used; these data correspond to a period of two months (1,440 data points). Other data sets corresponding to two months were also used as test patterns. The prediction errors obtained by the two structures of neural models for the different prediction horizons are shown in Table 3, whereas in Figures 24 and 25 different specific predictions for the Venice Lagoon level are shown. In both figures forecasting results of 1 and 12 hours ahead are presented for the case of 'high waters'. As can be seen from Figures 24 and 25 the evolution of the water level data is well predicted for the first case, i.e. 1 hour ahead but fails in predicting the high tide 12 hours in advance.

CONCLUSIONS

Where once time series analysis was shaped by linear systems theory, it is starting now to be possible to recognise when an apparently complicated time series has been produced by a low-dimensional nonlinear system, to characterise its essential properties, and to build a model that can be used for forecasting. Although the actually developed nonlinear time series analysis works nicely in applications to well-controlled laboratory experiments, or

simulations from systems having a limited degree of complexity, the situation is not so clear for complicated natural (uncontrolled) systems. In this case, the literature is full of claims and counterclaims for low-dimensional attractors on one side or 'coloured' noises with power law power spectra or nonlinear stochastic processes on the other side.

From the analysis of these results, it seems that apart from the tidal oscillations due to the relative motion of the earth, moon and sun, there is another dynamic factor which is not white noise. The analysis using the standard time-delay embedding techniques seems to indicate low dimensional chaotic dynamics. Furthermore, the space-time separation plots and the E1 & E2 method seem to indicate that it is not random coloured noise, even though the anomalous scaling of the nonlinear prediction does not follow the typical exponential dependence of chaotic systems. But this is probably due to the coexistence of chaos, noise and ordered motion. In this case, as the dynamic behaviour of the system is much more complex than usual situations, a scaling with a power law function might occur (Mannella *et al*. 1994).

It is always possible to define a time-delayed vector from a time series, but this certainly does not mean that it is always possible to identify meaningful structure in the embedded data. Because the mapping between a delay vector and the system's underlying state is unknown, the precise value of an embedded data point is insignificant. In this sense, it is important to point out the differences found when analysing the level and the residual (level minus the astronomic tide) data, since usual data treatment techniques, such as filtering, should be carefully applied when dealing with nonlinear time series.

Nonlinear predictions seem to follow accurately the 'normal' behaviour of the water level in the Venice Lagoon. However, they fail to recognise the 'high waters' phenomenon more than a few hours ahead. This is probably due to the predictability limit of the system. Neural networks perform better than local (linear and quadratic) and global polynomial predictors (see Tables 2 and 3). However, if one considers that the time delay of the Venice Lagoon water level was $T = 4$, and the embedding dimension was 8 and compares these results with Equation (28), it is easy to see that the results of applying nonlinear time series analysis are identical to the neural network pruning approach. Since the definition of the neural network architecture, i.e. number of neurons and connections, is a tedious iterative procedure, it seems that for the case of nonlinear time series prediction, phase space reconstruction techniques could be employed to speed up the whole procedure. In this case the neural network acts as a global predictor, similar to the one described in the nonlinear prediction section using chaos theory techniques. The main difference, with the global polynomial function used there, lies in the number of adjustable parameters – 56 against 45 – and the characteristics of the basis functions used—polynomials against sigmoidal functionals.

Similar results, time delay and embedding dimension have been obtained by Keijzer & Bavobic (1999) using the residual errors made by a deterministic model (MIKE 21) when forecasting the water level at Venice Lagoon.

Nonlinear forecasting results probably can be improved, as in the case of ARMA methods, using more data, i.e. atmospheric pressure, water levels at different locations along the Adriatic Sea, etc. Our present research is continuing along these lines.

ACKNOWLEDGEMENTS

We gratefully acknowledge I. Shepherd and Dr D. Al-Khudhairy for their invaluable advice, stimulating discussions and critical reading of this manuscript.

REFERENCES

Abarbanel, H. D. I. 1996 *Analysis of Observed Chaotic Data.* Springer, New York.

Abarbanel, H. D. I. & Kennel, M. B. 1993 Local false neighbours and dynamical dimension from observed chaotic data. *Phys. Rev.* E **47**, 3057–3068.

Accerboni, E. & Manca, B. 1973 Storm surge forecasting in the Adriatic Sea by means of a two-dimensional hydrodynamical numerical model. *Boll. Geof. Teor. Appl.* **54**, 25.

Accerboni, E., Castelli, F. & Mosetti, F. 1971 Sull'uso di modelli matematici idrodinamici per lo studio dell'acqua alta a Venezia. *Boll. Geof. Teor. Appl.* **49**, 18.

Babovic, V. 1998 A data mining approach to time series modelling and forecasting. In: *Hydroinformatics 98–Proceedings of the Third International Conference on Hydroinformatics* **2**, pp 847–856 (ed. V. Babovic & L. C. Larsen). Copenhagen, Denmark.

Badii, R., Broggi, G., Derighetti, B., Ravani, M., Ciliberto, S., Politi, A. & Rubio, M. A. 1988 Dimension increase in filtered signals. *Phys. Rev. Lett*. **60**, 979–982.

Bassingthwaighte, J. B. & Raymond, D. C. 1994 Evaluating rescaled range analysis for time series. *Ann. Biomed. Eng*. **22**, 432–444.

Bergamasco, L., Serio, M., Osborne, A. R. & Cavaleri, L. 1995 Finite correlation dimension and positive Lyapunov exponents for surface wave data in the Adriatic sea near Venice. *Fractals* **3**, 55–78.

Breeden, J. L. & Packard, N. H. 1994 A learning algorithm for optimal representation of experimental data. *Int. J. Bif. Chaos* **4**, 311–326.

Brown, R. 1993 Orthogonal polynomials as prediction functions in arbitrary phase space dimensions. *Phys. Rev.* E **47**, 3962–3969.

Cannon, M. J., Percival, D. B., Caccia, D. C., Raymond, G. M. & Bassingthwaighte, J. B. 1997 Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series. *Physica* A **241**, 606–626.

Cao, L. 1997 Practical method for determining the minimum embedding dimension of a scalar time series. *Physica* D **110**, 43–50.

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

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

Cybenko, G. 1989 Approximation by superposition of a sigmoidal function. *Mathematics of Control, Signals and Systems* **2,** 303–314.

Davis, M. E. 1997 Reconstructing attractors from filtered time series. *Physica* D **101**, 195–206.

Eckmann, J. P. & Ruelle, D. 1985 Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys*. **57**, 617–656.

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

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

Gilmore, R. 1998 Topological analysis of chaotic dynamical systems. *Rev. Mod. Physics* **70**, 1455–1526.

Grassberger, P. & Procaccia, I. 1983 Characterization of strange attractors. *Phys. Rev. Lett*. **50**, 346–349.

Hornik, K., Stinchcombe, M. & White, H. 1989 Multilayer feedforward networks are universal approximators. *Neural Networks* **2**, 359–366.

Hurst, H. E. 1951 Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Eng*. **116**, 770–779.

Isliker, H. & Kurths, J. 1993 A test for stationarity: Finding parts in time series apt for correlation dimension estimates. *Int. J. Bif. Chaos* **3**, 1573–1579.

Kantz, H. & Schreiber, T. 1999 *Nonlinear Time Series Analysis*. Cambridge University Press.

Kaplan, J. L. & Yorke, J. A. 1979 Chaotic behavior in multidimensional difference equations. In: *Functional Differential Equations and Approximation of Fixed Points* (Ed. H. O. Peitgen & H. O. Walther). Springer, Berlin, pp 204–227.

Keijzer, M. & Babovic, V. 1999 Error correction of a deterministic model in Venice lagoon by local linear models. To be presented at*: Modelli complessi e metodi computazionali intensivi per la stima e la previsione*, Venice, September.

Kennel, M. B., Brown, R. & Abarbanel, H. D. I. 1992 Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403–3411.

Mandelbrot, B. B. & Van Ness, J. W. 1968 Fractional Brownian motions, fractional noises and applications. *SIAM Rev*. **10**, 422–437.

Mandelbrot, B. B. 1983 *The Fractal Geometry of Nature*. W. H. Freeman, New York.

Mannella, R., Grigolini, P. & West, B. J. 1994 A dynamical approach to fractional Brownian motion. *Fractals* **2**, 81–94.

Manuca, R. & Savit, R. 1996 Stationarity and nonstationarity in time series analysis. *Physica D* **99**, 134–161.

Mees, A. I., Rapp, P. E. & Jennings, L. S. 1987 Singular value decomposition and embedding dimension. *Phys. Rev.* A **36**, 340–346.

Michelato, A., Mosetti, R. & Viezzoli, D. 1983 Statistical forecasting of storm surges: an application to the Lagoon of Venice. *Boll. Ocean. Teor. Appl.* **1**, 67–76.

Moretti, E. & Tomasin, A. 1984 Un contributo matematico all'elaborazione previsionale dei dati di marea a Venezia. *Boll. Ocean. Teor. Appl.* **2**, 45–61.

Ott, E., Sauer, T. & Yorke, J. A. (eds) 1994 *Coping With Chaos*. Wiley Interscience, New York.

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.

Rumelhart, D., Hinton, G. & Williams, R. J. 1986 Learning internal representations by error propagation. In: *Parallel Distributed Processing*. MIT Press, Cambridge, MA.

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

Schreiber, T. 1999 Interdisciplinary approach of nonlinear time series methods. *Phys. Rep.* (in press).

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

Takens, F. 1981 In: *Dynamical Systems and Turbulence*, Vol. 898 of Lecture Notes in Mathematics (Warwick), (ed. A. Rand & L. S. Young). Springer, Berlin, p. 366.

Theiler, J., Galdrikian, B., Longtin, A., Eubank, S. & Farmer, J. D. 1992 Using surrogate data to detect nonlinearity in time series. In: *Nonlinear Modelling and Forecasting*, SFI Studies in the Sciences of Complexity, Proc. Vol. XII, (M. Casdagli & S. Eubank, Eds). Adison-Wesley, Redwood City, California.

Tomasin, A. 1972 Autoregressive prediction of sea level in the Northern Adriatic. *Riv. Ital. Geofis*. **21**, 211.

Tomasin, A. 1973 A computer simulation of the Adriatic Sea for the study of its dynamics and for the forecasting of floods in the town of Venice. *Comp. Phys. Comm*. **5**, 51.

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

Vieira, J., Føns, J. & Cecconi, G. 1993 Statistical and hydrodynamic models for the operational forecasting of floods in the Venice Lagoon. *Coastal Eng.* **21**, 301–331.

Vittori, G. 1992 On the chaotic features of tide elevation in the lagoon of Venice. In *Proceedings of the ICCE'92, 23rd International Conference on Coastal Engineering, 4–9 October* 1992, Venice, Italy, pp 361–362.

Vittori, G. 1993 Caos deterministico nelle oscillazioni di marea a Venezia. *Atti dell'Istituto Veneto di Scienze, Lettere ed Arti* **151**, 23–61.

Wolf, A., Swift, J. B., Swinne, H. R. & Vastan, J. A. 1985 Determining Lyapunov exponents from a time series. *Physica* D **16**, 285–317.