Predicting time series using a neural network as a method of distinguishing chaos from noise

1992 J. Phys. A: Math. Gen. 25 843

(http://iopscience.iop.org/0305-4470/25/4/021)

J. Phys. A: Math. Gen. 25 (1992) 843-850. Printed in the UK

Predicting time series using a neural network as a method of distinguishing chaos from noise

J B Elsner

Department of Meteorology, Florida State University, Tallahassee, FL 32306, USA

Received 25 March 1991, in final form 31 October 1991

Abstract. A neural-network approach is presented for making short-term predictions on time series. The neural network does better at short-term predictions of a chaotic signal than does an optimum autoregressive model. Also, the neural network is clearly capable of distinguishing between chaos and additive noise.

1. Introduction

One of the basic tenets of science is making predictions. If we know previous behaviour, how can we predict future behaviour? The approach in many sciences requires two steps; construct a model based on theoretical considerations and use measured data as initial input. Since, in many cases, the underlying theoretical principles are known, model construction continues to be a primary area of interesting research.

One class of alternative approaches is to build models directly from the available data. For these methods, the data, given as a time series, is usually considered a single realization of a continuous random process. This is appropriate when the randomness is a result of complex interactions involving many independent and ultimately irreducible degrees of freedom. Along these lines, linear models have had some success especially in regards to relating cause and effect to physical phenomena; however, their predictive power is limited. The limitation is perhaps related to their inability to model the evolutionary dynamics of the system [1].

In the past few decades advances in the theory of dynamical systems have demonstrated the existence of dissipative systems whose trajectories that depict their asymptotic final states are not confined on limit cycles (periodic evolutions) or tori (quasi-periodic evolutions) but on attractors which are not submanifolds of the total available phase space. These attractors are fractal sets and are often called strange attractors. The corresponding dynamical systems are called chaotic systems and their trajectories never repeat. Thus their evolution is aperiodic but completely deterministic. Because the evolution is aperiodic any 'signal' measured from a chaotic dynamical system 'looks' quite irregular and exhibits frequency spectra with energy at all wavelengths (broadband spectra) similar to those of random 'signals'. Another important property of chaotic dynamical systems and their strange attractors is the divergence of initially nearby trajectories. Due to the action of the attractor the evolution of the system from two (or more) nearby initial conditions will soon become quite different. Since the measurement of any initial condition is subjected to some error, such a

0305-4470/92/040843+08$04.50

property imposes limits on long-term prediction. Nevertheless, for a short time nearby trajectories may not diverge significantly and thus, even though each individual evolution might be quite complex, the knowledge of the dynamics and especially of the structure of the attractor (dimensions, Lyapunov exponents, etc) may prove beneficial to the art of short-term prediction.

Motivated by the above ideas, very recently a number of techniques for making predictions have been developed to exploit the underlying determinism in complex systems [2-6]. The purpose of this paper is to show that neural networks are capable of making short-term predictions on time-series data that are better than an optimum autoregressive model and to show that such a methodology is capable of distinguishing chaos from additive noise.

2. Examples

In this section we present two examples showing the effectiveness of using a neural network for making predictions on time-series data. In general, neural networks work by iteratively solving for a weight matrix ($\mathbf{W}$) which is used for the inner product that takes inputs ($\mathbf{X}$) to outputs. The equation for the weights is given as

$$\mathbf{W}^{n+1} = \mathbf{W}^{n} - \eta \sum \delta E / \delta \mathbf{W}$$

where the error (E) at each iteration n is given by

$$E = (\mathbf{X} - \sum \mathbf{W}\mathbf{X})^2$$

and the sum is over all inputs (components of $\mathbf{X}$). The parameter η is the learning rate and is set to a value that ensures convergence. For time-series prediction, the inputs are taken to be lagged values of the discrete time sequence. More details concerning neural networks are given in Rumelhart *et al* [7] and Owens and Filkin [8].

The neural network architecture we employ consists of three layers: one input, one hidden and one output layer (figure 1). Learning is achieved using back-propagation of the errors resulting from the difference between predicted and actual values during training [8]. Both time series used in this study consist of 1000 data points. Training is performed on the first 500 values with subsequent predictions made on the remaining 500 values. The number of input nodes is set at eight, the number of hidden nodes is set at three and the number of output nodes is set at one. Numerous trial runs indicated that the accuracy of prediction was not sensitive to small changes in the number of input or hidden nodes. The single output represents some future value of the time series we wish to predict.

The inputs are the components of a reconstructed n-dimensional state space consisting of successive time-delayed values of the series. The method is similar to the one used by Perrett and van Stekelenborg [9] to predict annual sunspot numbers. For example, if we represent the series as $x(t_i)$ where $i = 1, 2, \ldots, L$, then with $\tau = 1$, and using an 8-dimensional phase space beginning with the first value of the time series, the first set of inputs is $\{x(1), x(2), \ldots, x(8)\}$ and the output we are trying to predict is $x(9)$. Similarly, the second set of inputs is $\{x(2), x(3), \ldots, x(9)\}$ and the output we are trying to predict is $x(10)$. Training continues over all training pairs (set of inputs and output) for several thousand iterations.

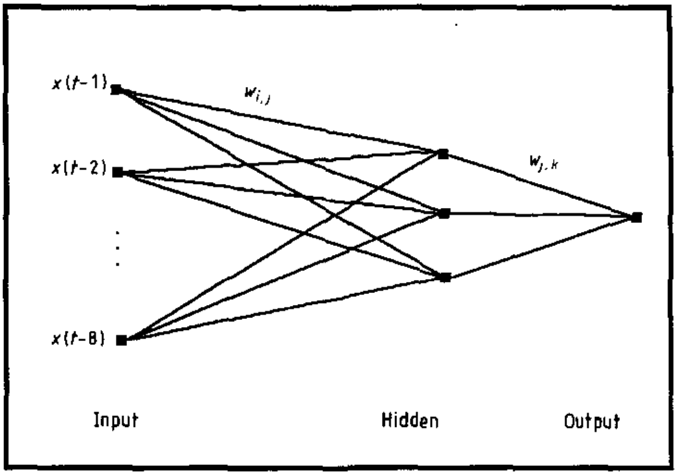

Figure 1. Architecture of the three layer neural network used in this study. The number of output nodes is set at one corresponding to the fact that we are making predictions one step into the future. The number of input nodes is set at eight and the number of hidden nodes is set at three. Results from trial runs indicated that adding more input and/or hidden nodes did not significantly improve the networks prediction capabilities, rather only slowed the convergence. The values at the input nodes are lagged values of the time series.

For the first example, we generate a time series by numerically integrating the Lorenz system [10] consisting of three ordinary differential equations describing convection of a fluid, warmed from below in time. The system is given as

$$dx/dt = -ax + ay$$

$$dy/dt = -xz + bx - y$$

$$dz/dt = xy - cz$$

where x is proportional to the intensity of convective motion, y is proportional to the horizontal temperature variation, z is proportional to the vertical temperature variation, and a, b and c are constants. For a choice of constants, corresponding to sufficient heating, the convection will exhibit chaos. We use a fourth-order Runge–Kutta integration scheme and constants $a = 16.0$, $b = 120.1$ and $c = 4.0$. The time series of convective motion (x-component of the system) after all transients (10^4 iterations) have diminished is shown in figure 2(a). Positive values indicate upward motion in the fluid. We take 1000 values from the time series, train the network on the first 500 values and make predictions on the last 500 values. Results of the neural network at predicting one step into the future (points) compared with the actual values (solid line) are given in figure 2(b). The normalized root-mean-square error (RMSE) between the actual and predicted values is 0.072 where zero implies a perfect forecast. Clearly the network is capable of capturing the underlying chaotic dynamics of the system.

We note here the success of other recent studies concerning accurate predictions of chaos using neural networks [11–14]. Our results indicate that accurate short-term predictions can be achieved for some chaotic systems with only a relatively small number of hidden nodes. We have made no attempt in this study, however, to compare

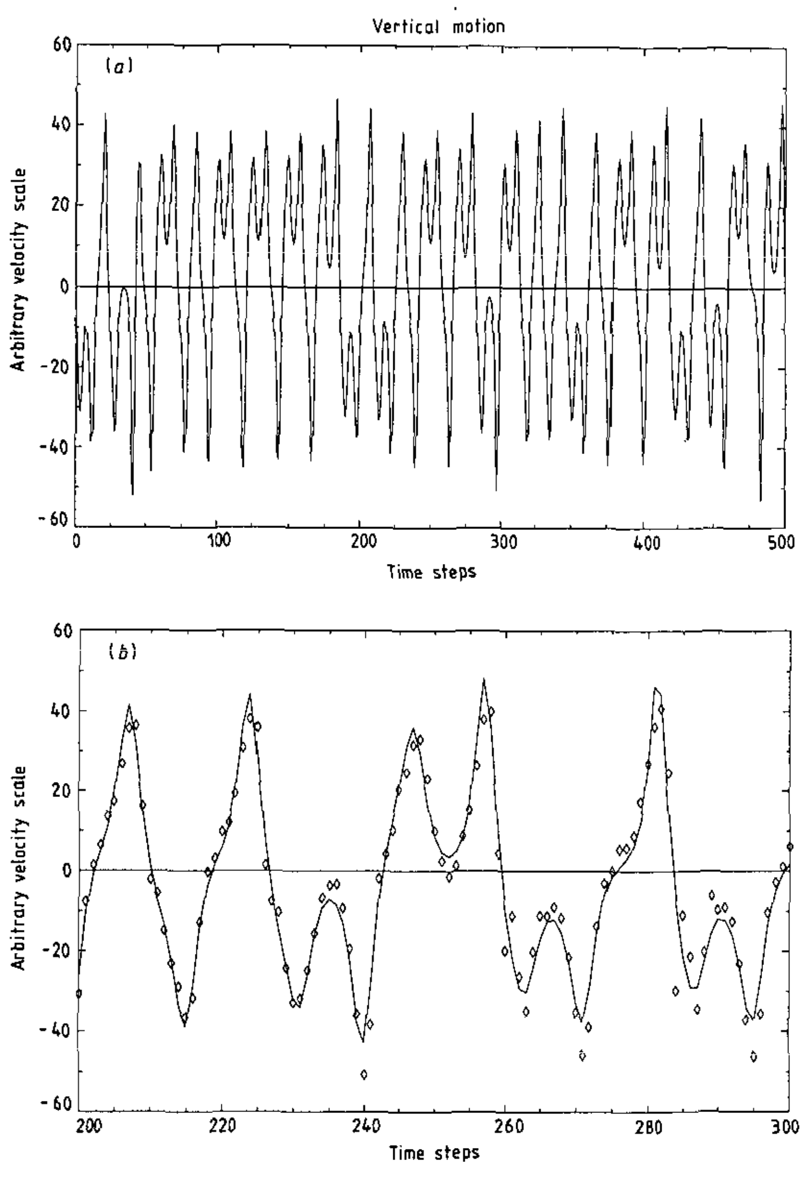

Figure 2. (*a*) Time series of convective motions generated by numerically integrating the Lorenz system using a fourth-order Runge–Kutta scheme with time step of 0.03. The time axis is in integration steps and the magnitude of convection is on an arbitrary velocity scale. The series displays chaotic oscillations. (*b*) Comparisons of the actual time series (line) with a neural network prediction (points). The actual time series represents a novel portion (second half) of the convective signal. Predicted values correspond quite well with actual values.

prediction skill with the number of hidden units and only speculate that there may be some relationship between the optimum number of hidden nodes and attractor dimension or the embedding dimension.

To assess the predictive ability of the neural network against that of a standard statistical model we fit the first half of the time series using an optimum autoregressive process and then compare predictions on the second half of the series from both models. For the autoregressive (AR) model the time series is viewed as a single realization of a stochastic process which is taken to be stationary and having a Gaussian distribution. For model selection we employed the Bayesian Information Criteria [15] and determined that the optimum order of the AR model for the convective time series was twelve.

Comparisons between the neural network and AR models are made by quantifying how the prediction accuracy (skill) decreases as predictions are made further into the future. To do this, we make a prediction one step into the future and then use this predicted value as one of the lagged inputs for the next prediction two time steps into the future. Similarly, the prediction at this second time step, as well as the previous time step, are used as lagged inputs for the next prediction three time steps into the future. Doing this successively allows us to compare the correlation coefficient between actual and predicted values as a function of prediction time where prediction time is given as discrete time steps into the future. The correlation coefficient between actual and predicted values is defined in the standard statistical way and is widely used as a measure of predictive skill. This procedure is followed for both the neural network model and for the optimum AR model. Results are shown in figure 3. For the first few steps into the future predictions from both models are good and the difference between the two models in terms of predictive skill is small. In contrast, the neural network makes significantly better forecasts than does the AR model as prediction time increases. Predictive skill on a non-uniform chaotic attractor will vary in time [16]. However, by using the same segment of the attractor to compare the models, as was done, we ensure a fair comparison. We note that the AR model is essentially a linear model and therefore incapable of capturing the inherent nonlinear nature of such a record. Since the signal is, in fact, chaotic we cannot hope to make accurate predictions with any model too

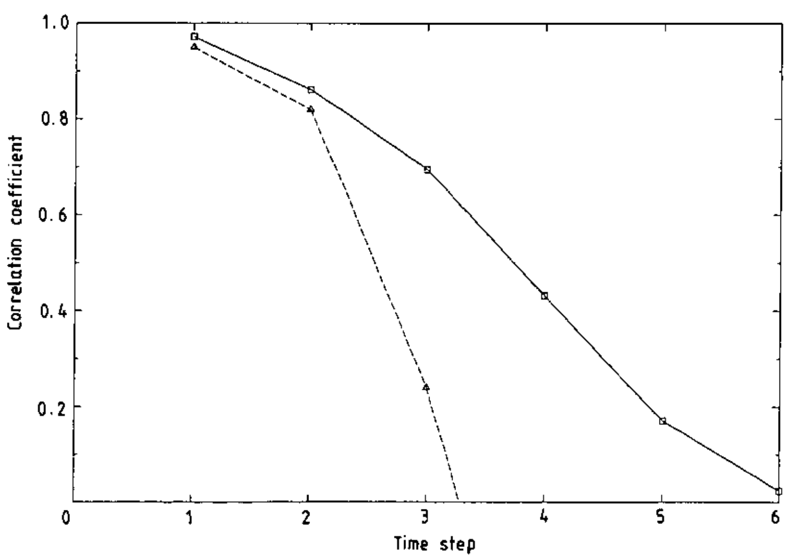

Figure 3. Correlation coefficients computed between actual and predicted values as a function of prediction time for the convective motions using a neural network (solid line) and using an optimum autoregressive model (dashed line). Prediction time is given as discrete time steps into the future. A correlation coefficient of one corresponds to perfect prediction. The neural network model clearly out-performs the autoregressive model.

far into the future and we see the predictive skill of the neural network also drops to near zero after a relatively short time.

Recently it has been suggested that certain nonlinear prediction techniques are capable of distinguishing between chaos and noise in time-series records [5, 17]. We demonstrate that neural networks share this capability by comparing results of the Lorenz system with results from a model trained on a time series generated from discrete points on a sine wave, having a unit amplitude, and adding to it at each step a uniformly distributed random variable in the interval $[-0.5, 0.5]$. Such a time series may display character similar to chaotic systems. Fourier analysis will result in spectra exhibiting peaks superimposed on a continuous background and dimensional analysis may indicate anything from a low-dimensional system (if noise is weak) to a random signal (if noise is strong).

After training the neural network on the first half of the signal composed of a sine wave plus noise we make predictions on the second half and, as was done with the Lorenz system, we compute the correlation coefficient between actual and predicted values as a function of prediction time. The dashed horizontal line in figure 4 is the result of this procedure. The independence of predictive skill with prediction length is in sharp contrast to the rapid decrease of predictive skill for the chaotic signal from the Lorenz system (solid line). From the differences we suggest that predicting time series using neural networks is another method for differentiating additive noise from deterministic chaos. With a simple AR model, such as persistence, predictive skill on a time series containing periodicities and/or noise will show a marked dependency on prediction length making it inappropriate for distinguishing chaos.

Neural network predictions on non-chaotic time series with additive noise will appear to have a fixed amount of error, regardless of how far into the future one tries

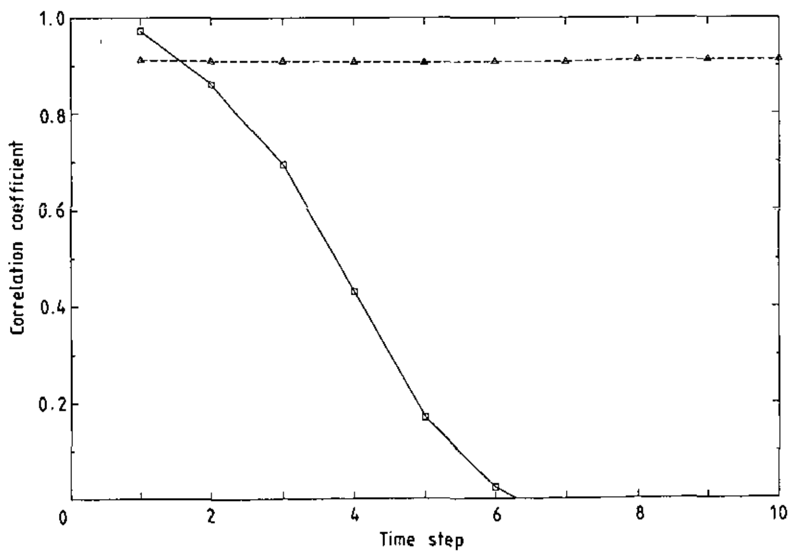

Figure 4. Correlation coefficients between actual and neural-network predicted values as a function of prediction time for the convective motions (solid line) and for a sine wave plus noise (dashed line). Prediction time is given as discrete time steps into the future. In many ways a signal composed of a sine wave plus noise is indistinguishable from chaos (e.g. Fourier spectrum), however, as is clear from the graphs of correlation coefficients, such a distinction can be made using a neural network.

to predict. On the other hand, prediction accuracy on chaotic time series will degrade as one tries to predict too far into the future. It is suggested that it might be possible to quantitatively compare the rates of degradation in prediction skill as an indication of the amount of chaos in a system. For example, one measure of the rate of degradation might simply be how many prediction steps are necessary before the correlation coefficient between actual and predicted values reaches some nearly asymptotic value. Lower numbers would correspond to fully developed chaos. Another measure, suggested recently by Wales [17], is the initial rate in the loss of prediction skill which can be related to the largest Lyapunov exponent in the system. We note that preliminary model and theoretical results indicate that this measure might be successful at distinguishing chaos from multiplicative noise, although more work in this area is needed.

3. *Conclusions*

In applying chaos theory in the analysis of time-series data one usually begins with estimating the dimension of the underlying attractor [18–21] by reconstructing a state space from the time series and then applying some variant of the correlation algorithm [22] on the set of points. The dimension, which is given by the power-law (scaling) behaviour of the correlation integral, gives a measure of the effective number of degrees of freedom of the system. Application of the algorithm, however, is subject to many problems like proper length of time series, proper time delay, etc. Also, because the scaling regions used to estimate the dimension involve only a small number of distances between points in the state space much of the information in the time series is lost, which for relatively short records can cause serious problems [5]. In addition, such methods may not, in certain cases, be able to distinguish self-affine random signals from chaos [23]. In contrast, prediction methods like the one discussed here, have the advantage that standard statistical procedures (such as correlation coefficients between actual and predicted values) can be used to evaluate their performance. Although we have studied only the case of additive noise, their performance should provide a more stringent test of underlying determinism in complex systems [3, 4].

Acknowledgments

I would like to thank J Perrett and J van Stekelenborg of the Bartol Research Foundation and A Owens at DuPont for introducing me to the back-propagation method for neural networks as a model for time-series predictions. Thanks are also extended to A Tsonis and the University of Wisconsin-Milwaukee for stimulating discussions during the course of this research. Part of the research was supported by NOAA's Global Change Program T-POP grant NA16RC-0454-01.

References

[1] Farmer J D and Sidorowich J J 1987 *Phys. Rev. Lett.* **59** 845
[2] Crutchfield J P and McNamara B S 1987 *Complex Syst.* **1** 417
[3] Farmer J D and Sidorowich J J 1988 *Technical Report LA-UR-88-901* Los Alamos National Laboratory
[4] Casdagli M 1989 *Physica* **35D** 335
[5] Sugihara G and May R M 1990 *Nature* **344** 734

[6] Linsay P S 1991 *Phys. Lett.* **153A** 353
[7] Rumelhart D E, Hinton G E and Williams R J 1986 *Nature* **323** 533
[8] Owens A J and Filkin D L 1989 *Int. Conf. on Neural Networks, Washington DC* **2** 381
[9] Perrett J C and van Stekelenborg J T P 1990 *Bartol Res. Found. Newark, DE*
[10] Lorenz E N 1963 *J. Atmos. Sci.* **20** 130
[11] Lapedes A and Farber R 1987 *Technical Report LA-UR-87* Los Alamos National Laboratory
[12] Stokbro K, Umberger D K and Hertz J A 1990 *Complex Syst.* **4** 603
[13] Frison T 1990 *J. Neur. Net. Comput.* **3** 31
[14] Deppisch J, Baur H-U and Geisel T 1991 *Phys. Lett.* **158A** 57
[15] Katz R W 1982 *J. Atmos. Sci.* **39** 1445
[16] Nese J M 1989 *Physica* **35D** 237
[17] Wales D J 1991 *Nature* **350** 485
[18] Nicolis C and Nicolis G 1984 *Nature* **311** 529
[19] Fraedrich K 1986 *J. Atmos. Sci.* **43** 419
[20] Essex C, Lookman T and Neremberg N A H 1987 *Nature* **326** 64
[21] Tsonis A A and Elsner J B 1988 *Nature* **333** 545
[22] Grassberger P and Procaccia I 1983 *Phys. Rev. Lett.* **50** 3446
[23] Osborne A R and Provenzale A 1989 *Physica* **35D** 357