MODELING NONLINEAR DYNAMICS WITH NEURAL NETWORKS: EXAMPLES IN TIME SERIES PREDICTION

Eric A. Wan

Stanford University, Department of Electrical Engineering, Stanford, CA 94305-4055 *

***Abstract*— A neural network architecture is discussed which uses Finite Impulse Response (FIR) linear filters to provide dynamic interconnectivity between processing units. The network is applied to a variety of chaotic time series prediction tasks. Phase-space plots of the network dynamics are given to illustrate the reconstruction of underlying chaotic attractors. An example taken from the *Santa Fe Institute Time Series Prediction Competition* is also presented.**

I. Introduction

The goal of time series prediction can be stated succinctly as follows: given a finite sequence $y(1)$, $y(2)$, $\ldots y(N)$, find the continuation $y(N+1)$, $y(N+2)$... The series may arise from the sampling of a continuous time system, and be either stochastic or deterministic in origin. Applications of prediction range from modeling turbulence to *differential pulse code modulation* schemes for telecommunication to stockmarket portfolio management. The standard prediction approach involves constructing an underlying model which gives rise to the observed sequence. In the oldest and most studied method, a linear autoregression (AR) is fit to the dqata:

$$y(k) = \sum_{n=1}^{T} a(n)y(k-n) + e(k) = \hat{y}(k) + e(k). \qquad (1)$$

This AR model forms $y(k)$ as a weighted sum of past values of the sequence. The single step prediction for $y(k)$ is given by $\hat{y}(k)$.

Neural networks may be used to extend the linear model to form a *nonlinear* prediction scheme. The basic form $y(k) = \hat{y}(k) + e(k)$ is retained; however, the estimate $\hat{y}(k)$ is taken as the output $\mathcal{N}$ of a neural network driven by past values of the sequence:

$$y(k) = \mathcal{N}[y(k-1), y(k-2), \ldots y(k-T)] + e(k). \qquad (2)$$

*We wish to acknowledge the support of NSF under grant NSF IRI 91-12531, of ONR under contract no N00014-92-J-1787, of EPRI under contract RP:8010-13, and of the Department of the Army Belvoir R D & E Center under contract #DAAK70-92-K-0003.

Note that this model is equally applicable for both scalar and vector sequences. The use of this nonlinear autoregression can be motivated as follows. First, *Takens Theorem* (Takens 1981) implies that for a wide class of deterministic systems there exists a *diffeomorphism* (one-to-one differential mapping) between a finite window of the time series $[y(k-1), y(k-2), \ldots y(k-T)]$ and the underlying *state* of the dynamics system which gives rise to the time series. This implies that there exists, in theory, a nonlinear autoregression of the form $y(k) = g[y(k-1), y(k-2), \ldots y(k-T)]$, which models the series exactly (assuming no noise). The neural network thus forms an approximation to the ideal function $g(\cdot)$. Furthermore, it has been shown (Hornik *et al.* 1989; Cybenko 1989; Irie and Miyake 1988) that a feedforward neural network $\mathcal{N}$ with an arbitrary number of neurons and 2 or more layers is capable of approximating any *uniformly* continuous function. These arguments provide the basic motivation for the use of neural networks in time series prediction.

II. Network Architecture and Prediction Configuration

The use of neural networks for time series prediction is not new. Previous work includes (Werbos 1974, 1980; Lapedes and Farber 1987; Weigend *et al.* 1990) to cite just a few. In this paper, we focus on a method for achieving the nonlinear autoregression by use of a Finite Impulse Response (FIR) network (Wan 1990, 1993. The network resembles a standard feedforward network where each synapse is replaced with an adaptive FIR linear filter as illustrated in Fig. 1. The FIR filter forms a weighted sum of past values of its input. The neuron receives the filtered inputs and then passes the sum through a nonlinear squashing functions. Neurons are arranged in layers to form a network in which all connections are made with the synaptic filters. Training the network is accomplished through a modification of the backpropagation algorithm (Rumelhart *et al.* 1986) called *temporal backpropagation* in which error terms are symmetrically filtered backward through the network. A complete description of the architecture along with the

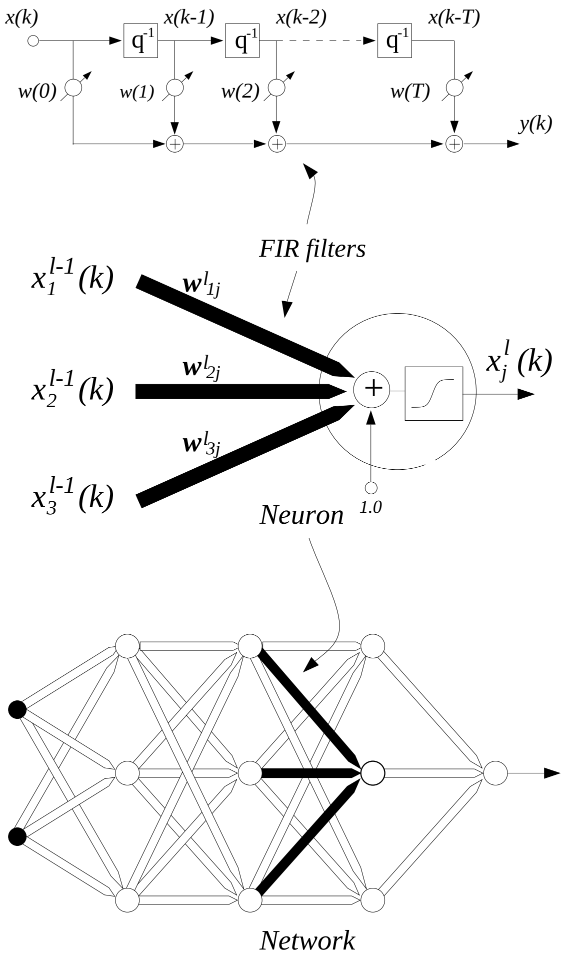


Fig. 1: FIR network architecture (q^{-1} represents a time-domain unit delay operator)

training algorithm can be found in (Wan 1993).

Fig. 2a illustrates the basic predictor training configuration. At each time step, the input to the FIR network is the known value $y(k-1)$, and the output $\hat{y}(k) = \mathcal{N}_q[y(k-1)]$ is the single step estimate of the true series value $y(k)$. Our model construct is thus: $y(k) = \mathcal{N}_q[y(k-1)] + e(k)$. Since the FIR network has only a finite memory of past samples, $\mathcal{N}_q[y(k-1)]$ is equivalent to a finite nonlinear regression on $y(k)$. During training, the squared error $e(k)^2 = (y(k) - \hat{y}(k))^2$ is minimized by using the *temporal backpropagation* algorithm to adapt the network. Note we are performing *open-loop* adaptation; both the input and desired response are provided from the known training series.

Once the network is trained, long-term *iterated* prediction is achieved by taking the estimate $\hat{y}(k)$ and feeding it back as input to the network: $\hat{y}(k) = \mathcal{N}_q[\hat{y}(k-1)]$.

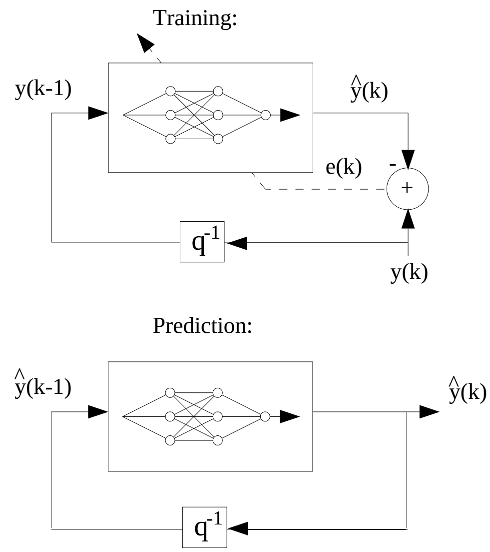


Fig. 2: Network prediction configuration: (a) Training. (b) Iterated prediction.

This closed-loop system is illustrated in Fig. 2b. The system can be iterated forward in time to achieve predictions as far into the future as desired.

III. Examples of chaos prediction and attractor reconstruction

In the examples that follow, networks of various dimensions are trained on time series generated by chaotic processes. The long term iterated prediction of the networks are then analyzed.

A. Results of the SFI Competition

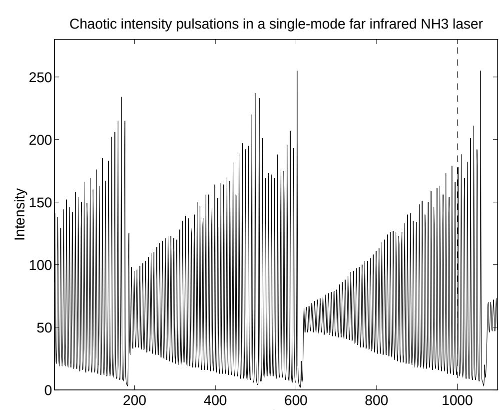


Fig. 3: 1000 points of laser data.

During the fall of 1991, *The Santa Fe Institute Time*

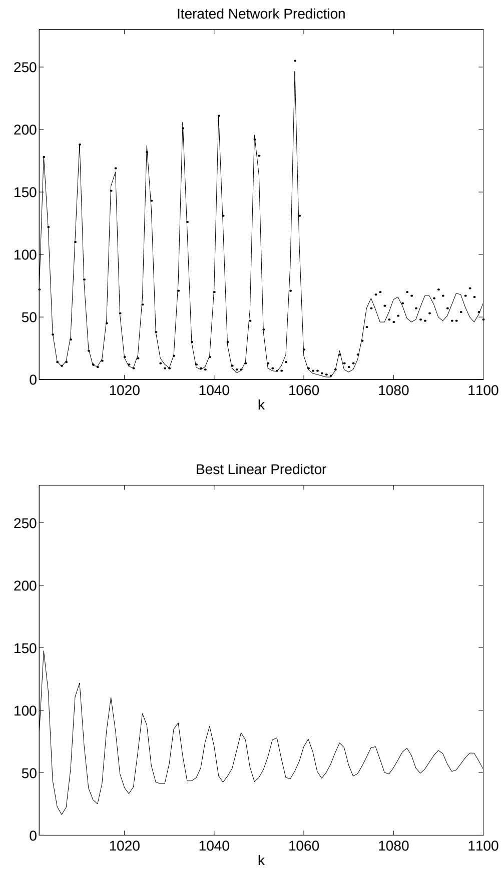


Fig. 4: (a) Network prediction (solid line) and series continuation (dashed line). (b) linear AR prediction.

Series Prediction and Analysis Competition was established as a means for evaluating and benchmarking new and existing techniques in time series prediction (Weigend and Gershenfeld 1993). The plot in Fig. 3 shows the chaotic intensity pulsations of an NH_3 laser * distributed as part of the competition. Contestants were given only 1000 points of data and then invited to send in solutions predicting the next 100 points. During the course of the competition, the physical background of the data set, as well as the 100 point continuation, was withheld to avoid biasing the final prediction results.

The 100 step prediction achieved by using an FIR network (dimensions: 1x12x12x1 neurons with 25:5:5 order filters) is shown in Fig. 4 along with the actual series continuation for comparison. It is important to emphasize that this prediction was made based on only the past 1000 samples. True values of the series for time past 1000 were not provided and were not available when the predictions were submitted. As can be seen, the prediction is remarkably accurate with only a slight eventual phase degradation. A prediction based on a 25th order linear autoregression is also shown to emphasize the differences from traditional linear methods. Other submissions to the competition included methods of k-d trees, piecewise linear interpolation, low-pass embedding, SVD, nearest neighbors, Wiener filters, as well as standard recurrent and feedforward neural networks. As reported by the Santa Fe Institute, the FIR network outperformed all other methods on this data set.

* "Measurements were made on an 81.5-micron 14NH3 cw (FIR) laser, pumped optically by the P(13) line of an N20 laser via the vibrational aQ(8,7) NH3 transition" - (Huebner 1989).

B. Mackey-Glass

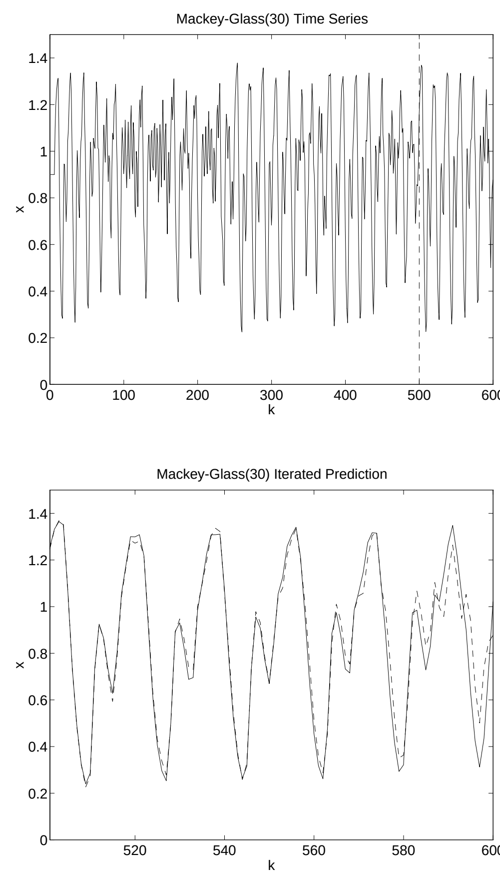


Fig. 5: (a) Mackey-Glass(30) delay-differential equation. (b) Iterated prediction (solid line) and series continuation (dashed-line).

For the next example we consider the Mackey-Glass delay-differential equation (Glass 1977):

$$\frac{dx(t)}{dt} = -0.1x(t) + \frac{0.2x(t-\Delta)}{1+x(t-\Delta)^{10}}. \tag{3}$$

with parameters $\Delta = 17$ and 30, initial conditions $x(t) = 0.9$ for $0 \leq t \leq \Delta$, and sampling rate $\tau = 6$. These

Table 1: Comparison of log normalized single step prediction errors for Mackey-Glass (small numbers mean a better prediction).

	Linear	Poly.	Rational	loc(1)
MG(17)	-0.57	-1.95	-1.14	-1.48
MG(30)	-0.49	-1.40	-1.33	-1.24
	loc(2)	RBF	N.Net	FIR Net
MG(17)	-1.89	-1.97	-2.00	**-2.31**
MG(30)	-1.42	-1.60	-1.50	**-1.79**

parameters were chosen to facilitate comparisons with prior work (Farmer and Sidorowich 1987; Lapedes and Farber 1987; Casdagli 1989). The time series for $\Delta = 30$ is shown in Fig 5a. An FIR network with $1 \times 15 \times 1$ nodes and $8:2$ in taps in each layer was trained on only the first 500 points of the series. The resulting log normalized single step prediction errors for the subsequent 1500 points are given in Table 1 along with results of other methods as summarized in (Casdagli 1989). While the FIR network shows a slight improvement over existing methods, the single step prediction task is really not that difficult. A more challenging problem is the iterated prediction as shown in Fig. 5b. The network receives no new inputs past the point 500. The iterated 100 point prediction is remarkably accurate.

C. Hénon Map

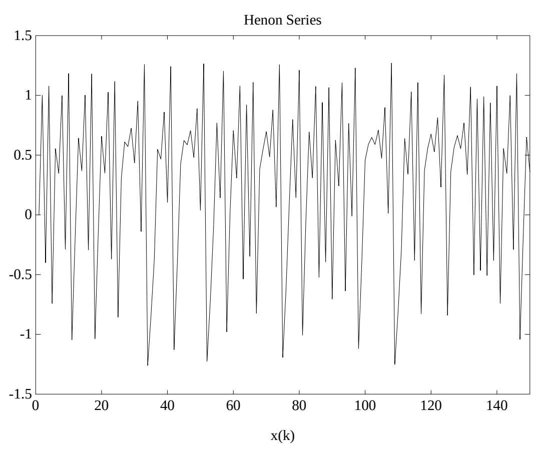


Fig. 6: Hénon series.

Consider the rather benign looking Hénon equations: $x_{n+1} = 1.0 - ax_n^2 + y_n$, $y_{n+1} = bx_n$, with $a = 1.4$ and $b = 1.3$(Hénon 1976). The iterated time series x_n is shown if Fig. 6. The phase-plot x versus y reveals a remarkable structure called a *strange attractor* (see

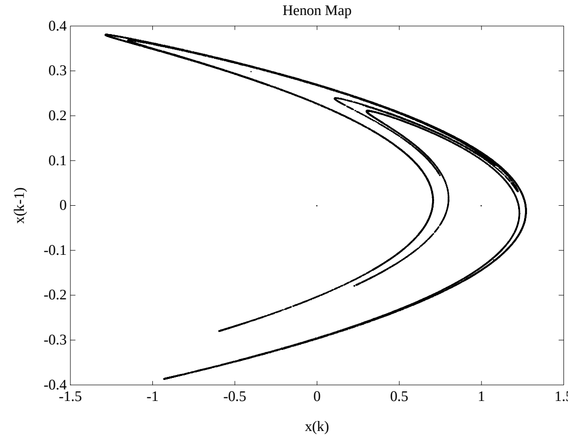


Fig. 7: Hénon attractor

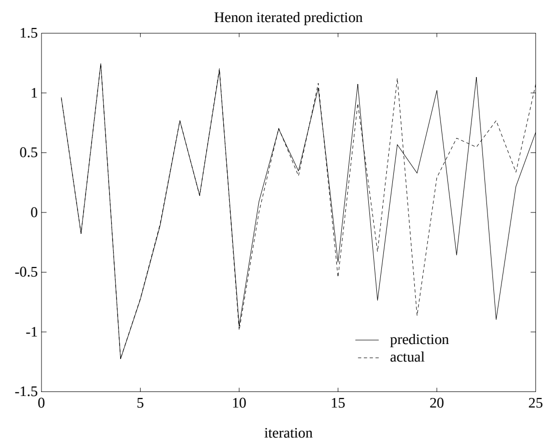


Fig. 8: Iterated prediction

Fig. 7. Increased magnification of the attractor would reveal ever finer detail in a fractal like geometry.

An FIR network was trained on the series using single step predictions. (Dimensions: 1x12x12x1 nodes with 2:1:1 order FIR filters in each layer)[†]. The iterated prediction versus the true series is shown in Fig. 8. As can be seen, the prediction is exceptionally accurate starting out, but then diverges after around 15 time steps. This divergence is unavoidable and reveals one of the fundamental tenants of chaos theory. The network system, however, can still be iterated thousands of time steps into the future and then used to construct its corresponding

[†]While a smaller network with lower embedding dimension would have worked for this problem, in general the actual dimensions of the system is not known in advance. Determining appropriate embedding dimensions for the network is a rich topic of research and beyond the scope of this paper.

attractor. As seen in Fig. 9, the original Hénon attractor emerges indicating that the network has indeed captured the underlying dynamics.

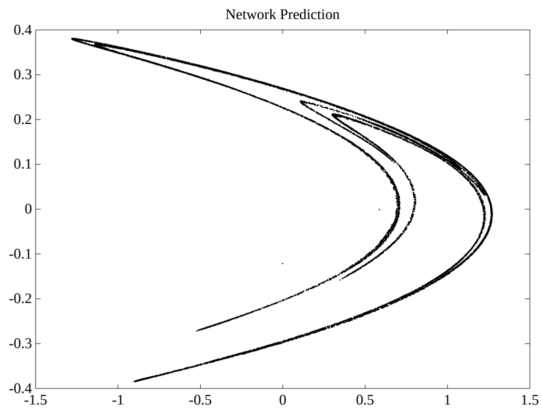


Fig. 9: Predicted attractor

D. *Ikeda Map*

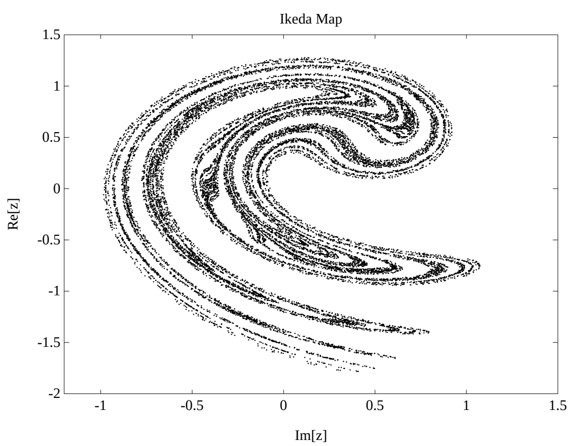


Fig. 10: Ikeda attractor

A more complicated equation corresponding the plane-wave interactivity in an optical ring laser is the Ikeda map: $z_{n+1} = a + Rz_n exp\{i[\phi - p/(1 + |z_n|^2)]\}$, with $a = 1.0, R = 0.9, \phi = 0.4, p = 6$ (Hammel *et al.* 1985). The phase-plot of real vs. imag. is shown in Fig. 10. To make the problem even more difficult for the network, only the imaginary sequence is used to train the network. The network must make a prediction for both the next imaginary value and the next real value. The network is thus performing a state estimation as it must learn the interrelation between the real and imaginary sequences. After training, the network (dimensions: 1x25x25x2 nodes with 5:2:2 taps) was iterated forward in time to form the image shown in Fig. 11. Again it is clear that the network has captured much of the underlying dynamics.

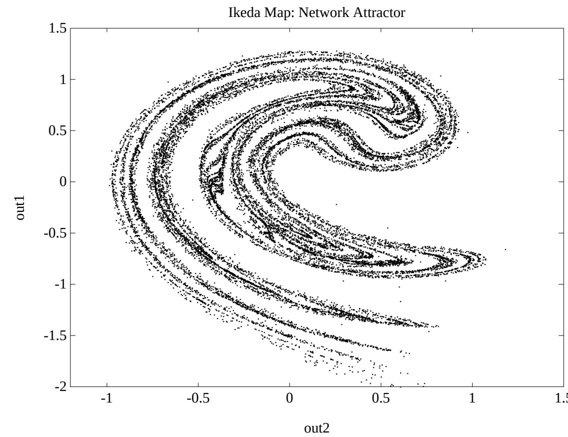


Fig. 11: Network predicted attractor

E. *Lorenz Dynamics*

A Lorenz system (Lorenz 1963) is described by the solution of the three simultaneous differential equations:

$$dx/dt = -\sigma x + \sigma y \quad (4)$$

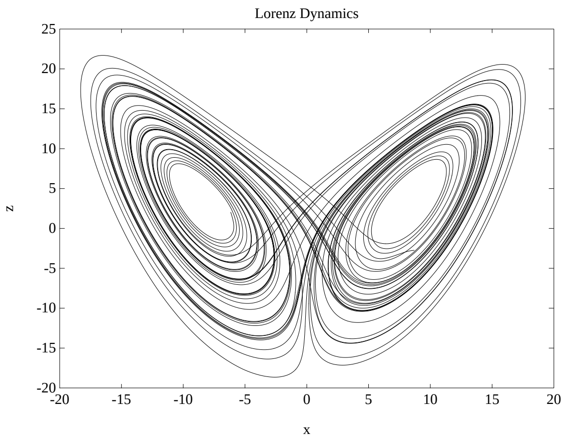
$$dy/dt = -xz + rx - y \quad (5)$$

$$dz/dt = xy - bz \quad (6)$$

A projection of the trajectory in the $x - z$ plane for parameters values $\sigma = 10$, $r = 28$, and $b = 8/3$ is shown in Fig. 12a. A network (dimensions: 1x12x12x1 nodes with 2:5:5 taps) was trained on observations of only the x state sampled at a period of 0.05 seconds. The output of the network was a prediction of both x and z at the next time step. This is again a state estimation problem. In Fig. 12b the iterated network prediction $\hat{x}(t)$ vs. $\hat{z}(t)$ is shown. The ability of the network to capture the underlying dynamics is evident.

IV. Conclusion

In this paper we have provided several examples illustrating the potential of using neural networks for time series prediction and modeling. While we have focused on autonomous deterministic scalar series, it should be clear that the basic principles directly apply to many problem of interest in dynamic modeling, system identification, and control.



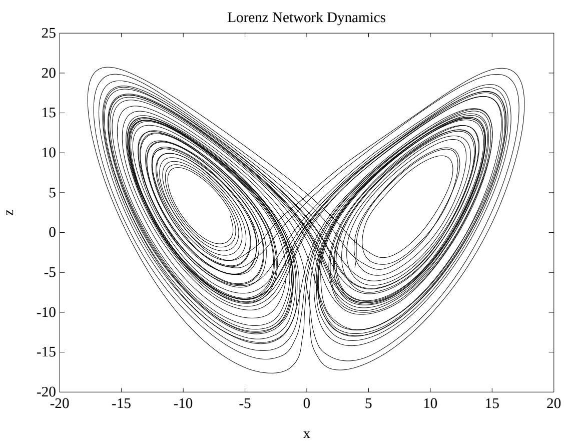


Fig. 12: (a) Lorenz attractor. (b) Network attractor

V. References

Casdagli, M. 1989. “Nonlinear Prediction of Chaotic Time Series”, *Physica D*, 35, 335-356.

Cybenko, G. 1989 “Approximation by superpositions of a sigmoidal function,” *Mathematics of Control, Signals, and Systems*, Vol. 2 No. 4, 303-314

Farmer, J.D. and Sidorowich, J.J. 1987. “Predicting chaotic time series”, *Phys. Rev. Lett.* 59, 854.

Hénon, M. 1976. ”A two-dimensional mapping with a strange attractor”, *Commun. Math. Phys. 50*, 69

Hammel, S. Jones, C.K.R.T. and Moloney, J.V. 1985. “Global dynamical Systems, and Bifurcations of Vector Fields”, *J. Opt. Soc. Am., B 2*, 552.

Hornik, K. Stinchombe, M. White, H. 1989. “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol 2, 359-366.

Huebner, U. Abrahm, N.B., and Weiss, C.O. 1989. “ Dimension and entropies of a chaotic intensity pulsations in a single-mode far-infrared NH3 laser,” *Physics Review*, A 40, 6354.

Irie, B. and Miyake, S. 1988. “Capabilities of three-layered perceptrons”, In *Proceedings of the IEEE Second International Conference on Neural Networks*, (San Diego, CA, July), Vol. I, 641-647.

Lapedes, A. and Farber, R. 1987. “Nonlinear Signal Processing Using Neural Networks: Prediction and system modeling”, Technical Report LA-UR-87-2662, Los Alamos National Laboratory.

Lorenz, E.N. 1963. “Deterministic Nonperiodic Flow”, *J. Atmos. Sci.*, 20, 130-141.

Mackey, M. Glass, L. 1977. “Oscillations and chaos in physiological control systems”, *Science* 197, 287.

Rumelhart, D.E. McClelland, J.L. 1986. and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, Cambridge, MA: MIT Press.

Takens, F. 1981 “Detecting Strange Attractors in Fluid Turbulence”. In *Dynamical Systems and Turbulence*, D. Rand and L.S. Young, eds. Springer, Berlin.

Wan, E. 1990 “Temporal backpropagation for FIR neural networks,” In *International Joint Conference on Neural Networks*, (San Diego, 1990) Vol. 1, 575-580.

Wan, E. 1993, “Time series prediction using a neural network with distributed time delays”, In *Proceedings of the NATO Advanced Research Workshop on Time Series Prediction and Analysis*, (Santa Fe, New Mexico, May 14-17 1992), A. Weigend and N. Gershenfeld, eds. Addison-Wesley.

Weigend, A. Huberman, B. and Rumelhart, D. 1990. “Predicting the future: a connectionist approach”, in *International Journal of Neural Systems*, vol. 7, no.3-4, 403-430.

Weigend, A. and Gershenfeld, N. eds. 1993. *Proceedings of the NATO Advanced Research Workshop on Time Series Prediction and Analysis*, (Santa Fe, New Mexico, May 14-17 1992), Addison-Wesley.

Werbos P. 1974. “Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences”, Ph.D. thesis, Harvard University, Cambridge, MA.

Werbos P. 1988. “Generalization of Backpropagation with Application to a Recurrent Gas Market Model”, *Neural Networks*, vol. 1, 339-356.