ISSN 1453-1437

International Journal of

Chaos Theory *and* Applications

Volume 4, no. 2-3 (1999)

T&T Press
Societa Italiana *Caos & Complessita*

AATI
Fuzzy Logic Systems Institute

Regular paper

Neural Modeling of Non-Linear Process: Relevance of the Takens-Mane Theorem

***Francesco Masulli*[*], *Riccardo Parenti*[**] *and Leonard Studer*[***]**

[*] *Instituto Nazionale per la Fisica della Materia & DISI-Dipartimento di Informatica e Scienze dell'Informazione, Universita di Genova, DISI, via Dodecaneso 35, I-16146 Genova, Italy, E-mail: masulli@ge.INFM. it*

[**] *Ansaldo Ricerche srl, Genova, Italy, E-mail: parenty@ari.ansaldo.it*

[***] *Institut de Physique Nucleaire, Universite de Lausanne, Switzerland, E-mail: leonard.studer@ipn.unil.ch*

Abstract

In this paper, we test a constructive methodology for shaping neural networks models of non-linear dynamic systems. The method is supported by results and prescriptions related to the Takens-Mane theorem and it is based on the measurement of the first minimum of the mutual information of the output signal, and in the application of the method of global false nearest neighbors to determine the embedding dimension. We present a numerical experiment to assess this constructive approach to the identification of a non-linear dynamic system and the application to the design of a neural network to forecasting a time series generated by an accelerometer coupled to a 150 MW steam turbine.

Keywords: ***neural network, prediction, predictor design, vibrations, steam engine***

1. Introduction

The problem of controlling systems characterized by non-linear dynamic behavior is widely encountered in the industrial world, in particular when the process requires a very demanding and ambitious performance, both in terms of precision and dynamic behavior.

An interesting example of application of the forecasting of a dynamical behavior can be found in the automatic condition monitoring of rotating machines. That is a well-known problem for many industrial plants. The most advanced systems record and analyze the vibration data measured by sensors mounted on the bearings of the machine and apply FFT analysis in order to calculate all the significant harmonics of the so-called "Rotor revolution equivalent frequency". The analysis of these data can be used to highlight various malfunctions and/or incoming mechanical fault. Unfortunately, using that approach, a very large amount of data has to be managed (on a typical steam turbine they are about 576,000/hour). Moreover, so far, there are no available "intelligent methods" to automatically check the present behavior of the machine comparing it with reference statuses, even an innovative method, based on Neural Networks, has been developed in Ansaldo and is now under industrial test [15, 16]. The above-mentioned system automatically and continuously compare the vibrational values read from the field to the Neural Network estimated ones. The above-mentioned system automatically and continuously compare the vibrational values read from the field to the Neural Network estimated ones. The neural estimator output is here considered a good "status specimen" for every working condition, being trained on the data collected when the behavior of the machine and the vibration phenomena are recognized as normal.

Even if the Ansaldo system should answer to the plant manager request, it leaves open the way for further improvement. In particular the neural system, being trained on a collection of different steady-state working points, of course tends to be less precise in forecasting during the transient.

In the last years, neural networks have been extensively tested on non-linear dynamic systems modeling and forecasting [8]. Those applications are supported by the *universal approximation theorem* [1, 2], that, unfortunately, is not a constructive one: no information can be extracted from the theory in order to define the structure of the neural network-based approximator.

On the other hand, results in the theory of chaotic systems point out relevant elements that can be extracted from the measurement of time series of one variable of the non-linear dynamic system. One of these results is provided by the Takens-Mane theorem [20-12] about the sufficient dimension of an Euclidean space to secure a fair representation of the true strange attractor of the underlying system. In [14, 19], this result has been applied to the shaping of a neuro-fuzzy system for forecasting chaotic time series.

In this paper, we test a constructive methodology for shaping a neural model of a synthetic non-linear process, that is supported by results and prescriptions related to the Takens-Mane theorem. Next, we apply the methodology to a test case consisting in forecasting the time series output of a vibration sensor located on a steam turbine at a given running.

In the next section we present the multilayer perceptron that is the neural network that we used in the experiments presented in this paper. In Section 3, the problem of dynamical system forecasting is discussed and we introduce the Takens-Mane theorem and the method of embedding. In Section 4 and 5, we show how to construct the neural network on the basis of the previous theoretical suggestions. In Section 6, we show an application to a simple non-linear dynamic system simulated in Matlab/Simulink. In Section 7, we apply the presented methodology to the design of a multilayer perceptron for the prediction of the vibrations of a steam turbine. In Section 8, we discuss the results. Conclusions are drawn in Section 9.

2. Multilayer Perceptron and Function Approximation

Theoretical experimental results support the use of neural networks in many applicative tasks. In particular, it has been shown that such systems can perform function approximation [1], [2], Bayesian classification [18], unsupervised vector quantization clustering of inputs [10], content-addressable memories [9], linear and non-linear principal component analysis and independent component analysis [7].

Artificial neural networks are made up of simple interconnected *nodes* or *neurons*.

Generally speaking, a node of a neural network can be regarded as a block that measures the similarity between the input vector and the parameter vector, or *weight,* associated to the node, followed by another block that computes an activation, normally not linear [8, 13]. The transfer function of an artificial neuron is given by the equation:

$$y = H\left(\sum_i w_i x_i - \theta\right) \qquad (1)$$

where H is the activation function, w_i are weights, θ is the threshold and x_i are the inputs to the neuron i.

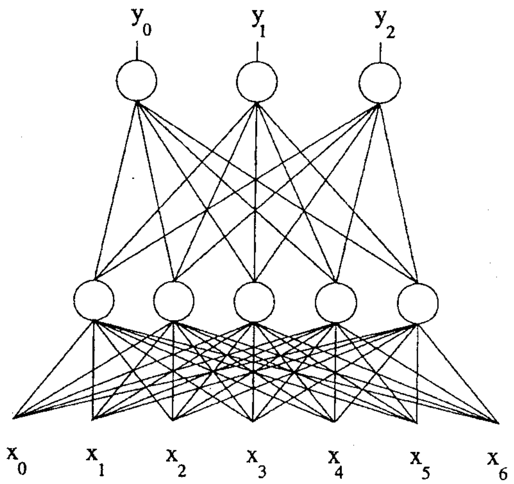

Figure 1

The most used neural network is the multilayer perceptron (MLP) that is a feed-forward model based on layers of neurons (Fig. 1). Nodes of each layer are interconnected with all nodes of the following layer. In this way multilayer perceptrons perform non-linear maps from an input space to an output space.

Moreover, as demonstrated by the *Universal Approximation Theorem* [1, 2], an MLP with a single hidden layer[1], and using sigmoid activation functions:

$$H(x) = 1/[1+\exp(-ax)],$$

where a is the slope parameter of the sigmoid function, is sufficient to uniformly approximate any continuous function with support in a unit hypercube.

The non-linear map can be automatically learned from data by a MLP through supervised learning techniques based on the minimization of a cost function, such as the Mean Square Error. A common learning technique is the *Error Back-Propagation, which* is an efficient application of the Gradient Descent method [8], [17].

The universal approximation property implies that, if the non-linear dynamical process can be represented by a continuous function, an efficient non-linear model can be built from data, using a multilayer perceptron. Using MLPs, the costly detailed design step of the first principles model usually implemented in the non-linear system identification is transformed in a simpler structuring step of the MLP, plus an optional pre-processing of the raw data coming from the field. (The preprocessing is eventually driven understanding of the physical model of the process.)

Even if, in principle, the function approximation property of MLP guarantees the feasibility of data-based models of non-linear dynamical systems, the neural network theory don't give any suggestion about many details. For example, no general prescriptions are available concerning the dimension of the data window (i.e. input layer of the MLP), the sampling rate of the input data, the dimension of the hidden layer, or the dimension of the training set. Consequently, usually those fundamental design parameters have to be obtained by experiments and heuristics [8].

3. Dynamical Systems and Chaos Theory

3.1 State Space

A deterministic dynamical system is described by a set of differential equations. Its evolution is represented by the trajectory in state space (of dimension *N*) of the vector:

$$\mathbf{Q} = (x, \dot{x}, y, \dot{y}, z, \dot{z}, \ldots)^T$$

where $x, \dot{x}, y, \dot{y}, z, \dot{z}, \ldots$ are variables of the system. The figure made in state space by **Q** is the attractor of the system.

For non-linear systems, the dynamical variables (*x, y, z,* ...) are coupled. The evolution of one variable (let say *x)* is not independent of all the other ones (*y, z,* ...). Except for few simple phenomena, the set of differential equations is unknown. Even the whole set of relevant effective dynamical variables is not always well defined in very complex systems such as in a steam turbine. But, as the variables are interdependent, the observation of only one of those brings information – maybe in an implicit way – on the other ones and consequently on the complete dynamical system. This is the reason why time series of non-linear dynamic systems are so useful.

3.2. Embedding Theorem

The question is now: "How to reconstruct the complete dynamical system with only the one-variable time series $(s_1, s_2, s_3, \ldots)$? " Here the theory of chaotic systems gives an answer. Based on topology, Takens and Mane [12, 20] have been able in 1981, independently, to prove the so-called-em Embedding Theorem related to the above question. A vector **Y** is built with *d* elements of the time series. From the Takens-Mane theorem the dimension of the vector has to be greater than two times the box-counting dimension D_0 of the attractor of the system.

A vector **Y** satisfying the Takens-Mane bound cited in the previous paragraph will evolve in a reconstructed state space. The evolution of **Y** will be in a diffeomorphic relation the signal **Q** state space point (a diffeomorphic is a smooth one-to-one relation). In other words, the evolution of **Y** is a fair copy of the evolution of **Q** for every practical purpose.

It is worth noting that there is a distinction between the order of the differential equation *n* which is the dimension of the state space where live the state point **Q** and *d* the sufficient dimension of a reconstructed state space where lives the reconstructed vector **Y**.

3.3. An Example

In order to illustrate the embedding theorem, let consider a sine wave $s(t) = A\sin(t)$. In $d =$ 1 (i.e. the $s(t)$ space) this wave oscillates in the interval [-*A*, +*A*]. Two points which are close in the sense of Euclidean (or other distance) may have quite different values of the derivative $\dot{s}(t)$. In this way two "close" points may move in opposite directions along the single spatial axis. In a two-dimensional space $[s(t), s(t+T)]$ (*T* negative) the ambiguity of the dynamics of points is resolved. The system evolves on a figure (in general, an ellipse) that is topologically equivalent to a circle. If we draw the since wave inn the three dimensions $[s(t), s(t+T), s(t+2T)]$, no further unfolding occurs and the sine is represented as a new ellipse.

3.4. The Method of Embedding

In this paper, we use the time delay embedding method [3]. It consists in building *d*-dimensional state vectors:

$$\mathbf{Y}(k) = [s(k), s(k+T), \ldots, s(k+(d-1)T)].$$

In principle, it suffices that $d \geq n$. However, the *effective* dimension d is not directly related to the dynamical dimension n – as in the case of weak coupled variables.

4. Choosing the time delay

A key feature of chaotic systems is that two initially close chaotic trajectories will diverge exponentially in time. This makes the chaotic system very unstable. It implies that the time delay T used in the embedding is a parameter, which has to be chosen carefully. If it is too long, the scalar variables:

$$[s(k),\, s(k+T), \ldots, s(k+(d-1)T)]$$

are not correlated and so the chaotic dynamical system can not be reconstructed. If it is too short, every scalar is essentially a copy of the previous one, bringing very little information on the dynamical system.

The careful choice of T is not needed only for chaotic systems, but for all non-linear ones. The *average mutual information* between the set of measurements of the observable $s(n)$ and the set of measurements of $s(nt)$ is defined as:

$$I(t) = \sum_{s(n),s(n+t)} P(s(n),s(n+t) \times \times \log_2 \left[\frac{P(s(n),s(n+t))}{P(s(n))P(S(N+T))} \right] \qquad (2)$$

It has been suggested [6, 5, 21, 3] to take the time T, where the first minimum of $I(t)$ occurs, as the value to use at the delay in the phase space reconstruction. In this way, the values of $s(n)$ and $s(n+T)$ are the most independent of each other in an information-theoretic sense. Moreover, the first minimum of average mutual information is a good candidate for the interval between the components of the state vectors that will be input to the neural network model of the non-linear dynamical process.

5. Evaluating the Global Embedding Dimension

From the Embedding Theorem, the box counting dimension D_0 should be evaluated. In principle, it can be estimated directly from the time series itself, but this task is very sensitive to the noise and needs large set of data points (order of 10^{D_0} data points) [3]. In order to avoid those problems, we can estimate the *embedding dimension* d_E, defined as the lowest (integer) dimension, which unfolds the attractor, i.e. the minimal dimension for which foldings due to the projection of the attractor in a lower dimensional space are avoided. The embedding dimension is a *global* dimension and in general is different from the local dimension of the underlying dynamics.

The Embedding Theorem guarantees that if the dimension of the attractor is D_0, then we can unfold the attractor in a space of dimension $d_E (d_E > 2D_0)$. It is worth noting that d_E is not a necessary condition for unfolding, but it is a sufficient condition. The dimension of the input layer in the multilayer perceptron will be then of high enough dimension for the deterministic part of the dynamics of the system is unfold.

5.1. Global False Nearest Neighbors

In practice, the method of *Global False Nearest Neighbors* [3] can be used to evaluate the embedding dimension d_E. Given a data space reconstruction in dimension d, with data vector $\mathbf{Y}(k) = [s(k), s(k+T), \ldots, s(k+(d-1)T]$, where the time delay T is the first minimum of the average mutual information. Let be:

$$\mathbf{Y}^{NN}(k) = \\ = [s^{NN}(k),\, s^{NN}(k+T), \ldots, s^{NN}(k+(d-1)T)],$$

the nearest neighbor vector in phase space. If the vector $\mathbf{Y}^{NN}(k)$ is a *false neighbor* (FNN) of $\mathbf{Y}(k)$, having arrived in its neighborhood by projection from a higher dimension because the present dimension d does not unfold the attractor, then by going to the next dimension $d + 1$ we may move this false neighbor out of the neighborhood of $\mathbf{Y}(k)$.

We define the distance ξ between points in the dimension $d + 1$ space relative to the distance in dimension d as:

$$\xi(k) \equiv \sqrt{\frac{R_{d-1}^2(k) - R_d^2(k)}{R_d^2(k)}}, \qquad (3)$$

Then:

$$\xi(k) = \frac{|s(k+dT) - s^{NN}(k+dT)|}{R_d(k)}. \qquad (4)$$

As suggested by Abarbanel [3], $\mathbf{Y}^{NN}(k)$ and $\mathbf{Y}(k)$ can classified as a false neighbor if $\xi(k)$ is a number greater than a threshold $\theta(\xi(k) \geq \theta)$. In many applications, a good value for θ is 15.

In case of clean data from a dynamical system, we expect that the percentage of FNNs will drop from nearly 100% in dimension one close to zero when d_E is reached.

It is worth noting that, as we go to higher dimensional spaces, the volume available for data grows as the distance to the power dimension, and no near neighbor will be classified close neighbor. In this case, we can modify the eq. 4 as:

$$\xi(k) = \frac{|s(k+dT) - s^{NN}(k+dT)|}{R_A}, \qquad (5)$$

where A is the nominal "radius" of the attractor, defined as the RMS value of data about its mean, e.g.:

$$R_A = \frac{1}{N}\sum_{k=1}^{N} |s(k) - s_{av}|, \qquad (6)$$

$$s_{av} = \frac{1}{N}\sum_{k=1}^{N} s(k). \qquad (7)$$

5.2. Bells Whistles and Pitfalls of FNN

- The global FNN calculation is simple and fast;
- the FNN calculation applied to signals coming from two different outputs of the same dynamical system gives, in general, two different values of d_E. Then from each signal we will obtain different reconstructed coordinate systems, but booth consistent with the original dynamical system;
- FNN method is valid even if the signal of interest results from a filtered output of a dynamical system [3, 4];
- If the signal is contaminated by noise (assumed to be generated by a high dimensional system), it may be that the contamination will dominate the signal of interest and FNN will show the dimension required to unfold the contamination. Here, a simple byproduct of FNN calculation is an indication of noise level in a signal.

6. Numerical Experimental

6.1. MATLAB / Simulink model

In order to test the constructive methodology, a dedicated computer experiment has been developed. Using a Matlab/Simulink environment, a non-linear dynamic system able to show different dynamic behavior for different amplitude of its input has been implemented (see Fig. 2). The target of this experiment was to build-up a non-parametric model using only knowledge extracted from output signals of the simulated system.

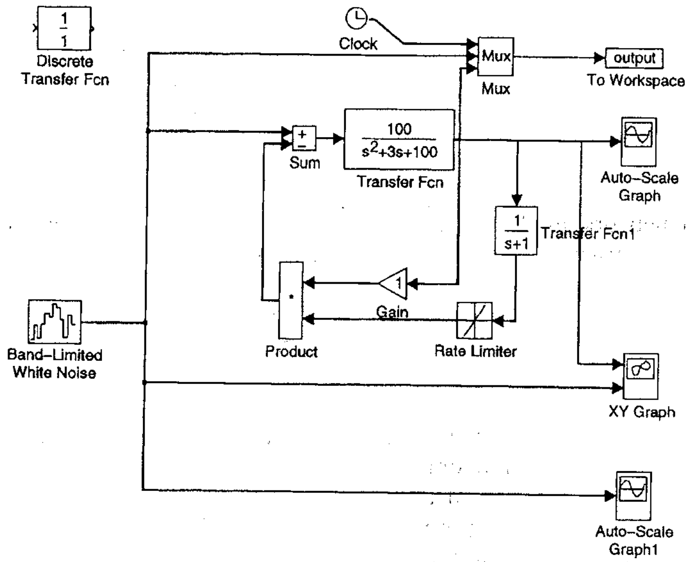

Figure 2

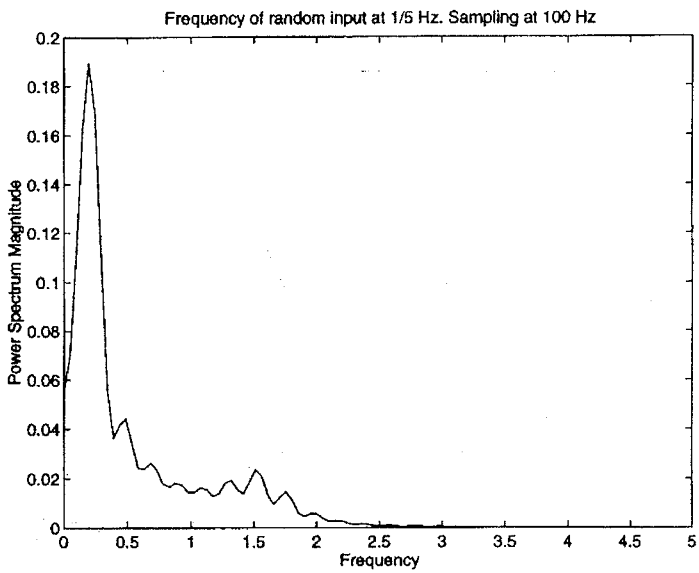

Figure 3

In preliminary experiments (not reported in this paper), the application of classical linear identification methods lead to poor results, such as a strong dependency of the model from the working point, and unpredictable results connection with changes on time scale.

In the experiment presented below, we applied a constructive methodology for shaping a neural network model of the non-linear dynamic system, results and prescriptions related to the Takens-Mane theorem.

6.2. Random Stimulation Input

A possible approach to study the input/output relations of a system is to input random series (random simulation input – RSI), and then to study its inputs and outputs. This approach is mimicking a blind acquisition on a real plant where it is very easy to collect data but it is very difficult to have any control of the system inputs shape and amplitude.

In this numerical experiment, the input of the system is a train of steps with random amplitudes. The length of the plateau is 3 time the period of the fundamental frequency of the dynamical system.

6.3. Power Spectrum

If we stimulate the circuit with RSI changing every 1/5 Hz, the power spectrum (see Fig. 3) shows a peak at the left end, in correspondence of the frequency of changing of the RSI.

6.4. Average Mutual Information

In Fig. 4, we present a plot of the average mutual information $I(T)$ of the output signal. The minimum of $I(T)$ is for $T = .16$ sec.

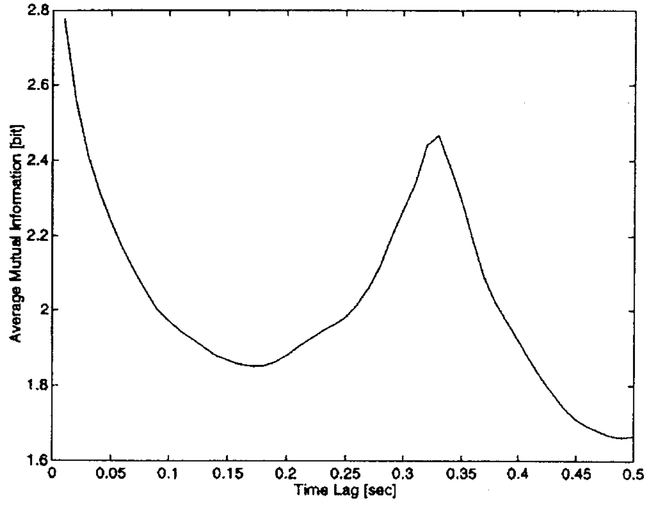

Figure 4

6.5. Embedding

As shown in Fig 5, the percentage of FNN goes to a minimum for $d \geq 4$. Also, as the input signal is a train of steps, two past inputs are sufficient to describe the external excitation signal. The condition $d \geq 4$ and two past inputs imply we should use two past outputs to reconstruct the dynamical system.

6.6. Constructive Approach to Neural Network Modeling

In order to test the effectiveness in practical non-linear dynamic process control problems of the theory related to the Takens-Mane theorem, we modeled the dynamical system by a multilayer perceptron with l inputs $2l$ sigmoidal nodes in a first hidden layer, and 1 output linear node. We assumed that the dimension l of the input layer of the MPL must be equal to the dimension of the reconstructed space of the dynamical system, so $l = 4$.

6.7. Results of the Numerical Experiment

A database of labeled patterns:

$$P_k = [(u(k+T), u(k), s(k+T), s(k), s(k-2T)]$$

(T is negative in our conventions) was obtained by stimulating the non-linear circuit with a RSI of 2 sec. $T = 0.16$ sec, $u(.)$ are the inputs and $s(.)$ the outputs of the circuit.

The data base was subdivided in a learning, a test and a validation sets of 5000, 1000, and 1000 patterns. Then, the learning set was shuffled and a 4-8-4-1 MLP was trained on it.

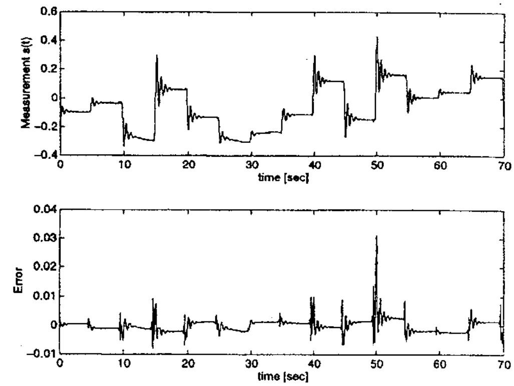

Figure 6

In Fig. 6, the very good quality of approximation of the behavior of the dynamical system obtained by the MLP can be noticed.

7. Application to a steam turbine time series

In this section, we tested the methodology described above, to the design of a Neural Network for forecasting the vibration corresponding to the correct working state of 150 MW Siemens steam turbine at a given running. Information of the working state of the turbine are obtained through the measurements of piezoelectric accelerometers coupled to part of the turbine and collecting the various vibrations.

We use time series of 400 ms (1024 points each) recorded with an interval of 8 hours.

7.1. Power Spectrum

The power spectrum of the first time series is displayed in Fig. 7.

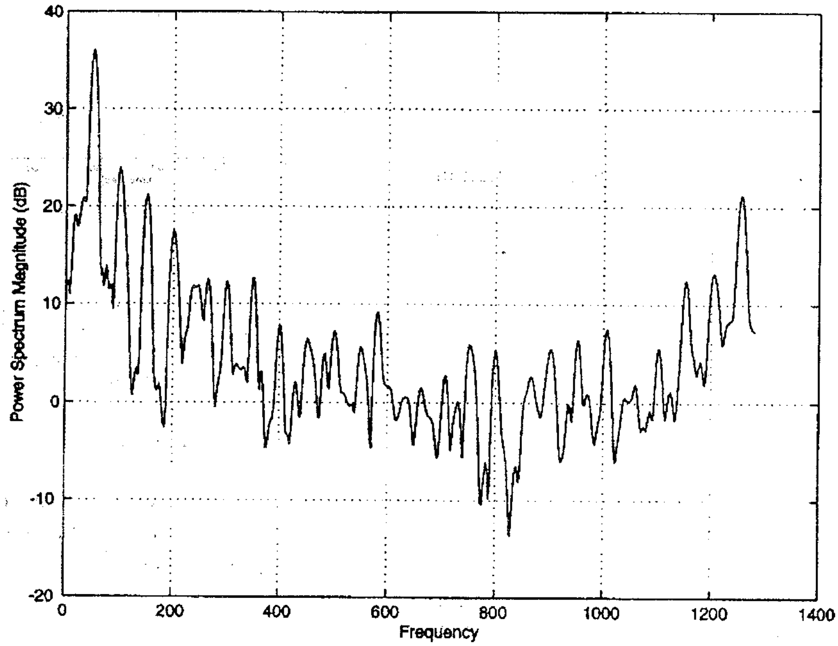

Figure 7

The main period of the time series is clearly visible, but a lot of secondary frequency peaks should be noticed, apart of the harmonics of the main period. The multiplicity of the secondary peaks together with the decay of the power spectrum signal the nonlinearity of the system [3].

7.2. First Minimum of Average Mutual Information

As shown (in Fig. 8), the first minimum of the average mutual information is located at 12 time lags of 0.39 ms (i.e. 4.68 ms), i.e. data points separated by 4.68 ms are minimally correlated. We will take this interval as a good candidate for the interval between the components of the state vectors that will be input to the neural network forecaster.

7.3. Global False Nearest Neighbors

Applying the false nearest neighbors algorithms lead us to consider that the dimension of the state vector should be at least 5 (see Fig. 10).

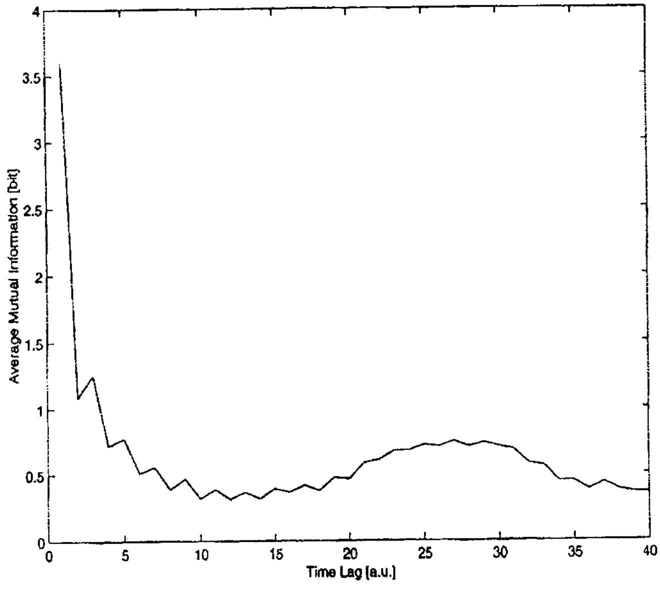

Figure 8

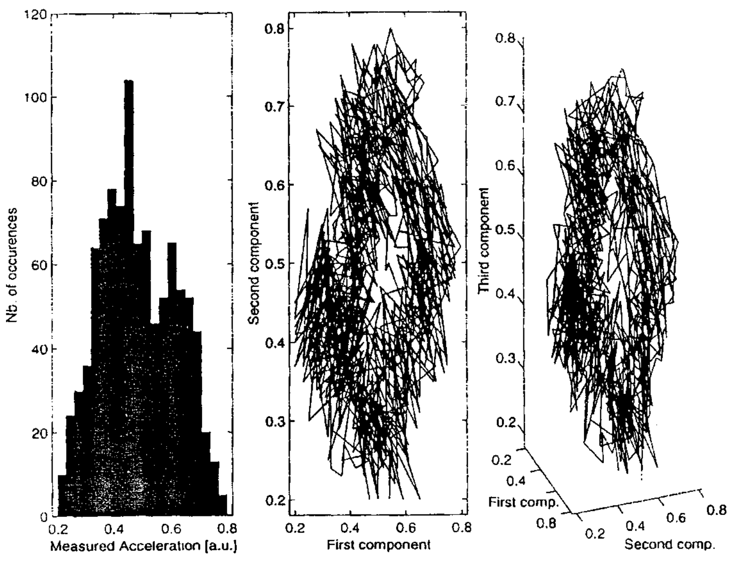

Figure 9

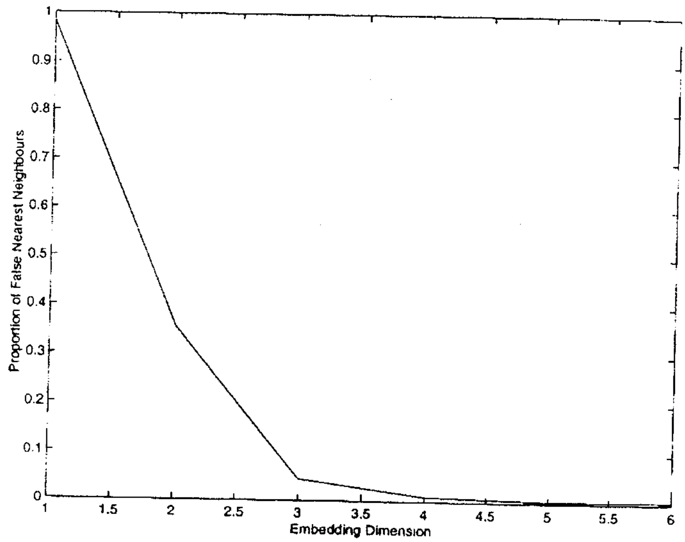

Figure 10

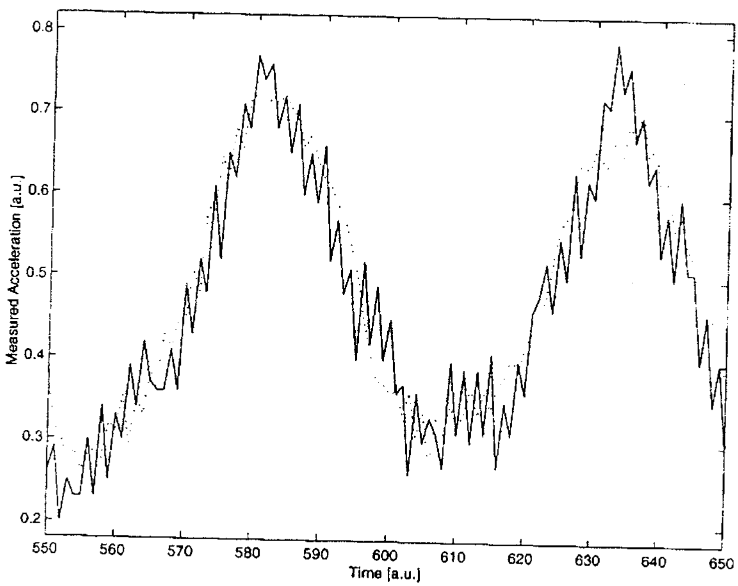

Figure 11

In Figure 9 we show how the trajectory of state vectors looks like for increasing dimensions.

7.4. Neural Networks Forecaster Design and Results

Having computed the optimal time lag (4.68 ms) and the minimal embedding dimension (5), we can design part of the architecture of a forecaster based on a multilayer perceptron with sigmoidal activation functions. The minimal embedding dimension enforced the number of inputs of the network to be 5. The training patterns are the 5 dimensional state vectors with components separated by 4.68 ms. The requested output is the value of the time series 1.56 ms latter that the most recent component of the state vector. We obtained the best results using a 5-7-1 MLP.

The training set of the network is made of 600 vectors. A test set of 379 vectors is used in order to stop before over-training. These two sets are built with the first time series. The obtained forecaster is validated with 977 vectors built on the second time series (which was recorded eight hours after the first one). It can be seen that the forecaster is able to follow the second time series (see Fig. 11). Notice that the power spectrum of the second time series is slightly different from the first one.

8. Discussion

By using the Takens-Mane theorem, we have been able to straightly build a neural network forecaster of rather good quality. In order to demonstrate the usefulness of the described approach, we have not taken care to finely tune the internal structure of the multilayer perceptron.

In [14] two of us have already examined the sensitivity of our method to the exact value of the embedding dimension for the chaotic system obtained by the Mackey-Glass equation [11]. The lower bound in the number of components of the state vector was given by the Takens-Mane theorem. In neural models of non-linear dynamic systems presented in this paper, we notice a similar sharp transition in the quality of the forecaster with respect to the number of components of the temporal window.

On the contrary, the quality of the forecaster is less sensitive to the time lag between the components of the state vector. This is expected because the Takens-Mane theorem is valid for any time lag. In practice, it can be expected that the time lag depend on the noise affecting the dynamical system.

9. Conclusions

In the years, networks have been extensively tested on non-linear dynamic systems modeling and forecasting [8]. Those applications are supported by the *universal approximation theorem* [1, 2], that, unfortunately, is not a constructive one: no information can be extracted from the theory in order to define the structure of the neural network based approximator. The neural network theory does not provide any general suggestion about the dimension of the data window (i.e., input layer of the MLP), the sampling rate of the input data, the dimension of the hidden layer, or the dimension of the training set.

On the other hand results achieved in the theory of chaotic systems point out very relevant elements, which can be extracted from the measurement of time series of one variable of the non-linear dynamic system. One of these results is provided by the Takens-Mane theorem [12, 20], namely on the sufficient dimension of an Euclidean space needed to secure a fair representation of the true strange attractor of the underlying system.

In this paper, we have tested a constructive methodology for shaping a neural model of a synthetic non-linear process, that is supported by results and prescriptions related to the Takens-Mane theorem [12, 20]. Then, we have applied the methodology to the design neural model of the vibration dynamic of a Siemens steam turbine. The proposed constructive methodology has been shown to be very easy to use, leading to useful results.

The integration between the Neural Network estimation and the Chaos Theory proposed in the present work, may be very useful in developing a new generation of the predictive state estimator for non linear dynamic systems. In particular it is foreseen to test in practice integrating it in the existing Ansaldo Condition Monitoring System for Steam turbine, in order improving its behavior to the transients.

Acknowledgements. Dr. Andre Perrenoud at VIBRO-Meter S.A., Fribourg in Switzerland, has kindly supplied us with data recorded form the steam turbine.

References

[1] G. Cybenko: Approximation by Superposition of a Sigmoidal Function. *Math. Of Control Signals and Systems,*2: 303-314, 1989.

[2] K. Hornik, M. Stinchcombe and H. White: Multilayer Feedforward Network are Universal Approximators. *Neural Networks*, 2: 356-366, 1989.

[3] H.D.I. Abarbanel: *Analysis of Observed Chaotic Data*. Springer, New York, USA, 1996.

[4] M.E. Dave: Reconstruction attractors from filtered time series. *Physica D*, 101: 195-206, 1997.

[5] A. Fraser: Information theory and strange attractors. Technical Report PhD THESIS, University of Texas, Austin, 1989.

[6] A. Fraser and L. Swinney: Independent coordinates for strange attractors from mutual information. *Physical Review,*33: 1134-1140, 1986.

[7] M. Girolami and C. Fyve: A temporal model of linear Anti-Hebbian Learning. *Neural Processing Letter*, 4, 1997.

[8] S. Haykin: *Neural Networks: A Comprehensive Foundation*. Macmilan, New York, Toronto, 1944.

[9] J.J. Hopfield: Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, 79, 1982.

[10]T. Kohonen: *Self-Organization and Associative Memory*. Springer-Verlag, Berlin, 3 edition, 1989.

[11]M.C. Mackey and L. Glass: Oscillation and chaos in physiological control systems. *Science,*197: 287, 1977.

[12]R. Mane: On dimension of the compact invariant sets of certain non-linear maps. In D.A. Rand and B.S. Young, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pp. 230-242, Warwick 1980, 1981. Springer-Verlag, Berlin.

[13]F. Masulli: Bayesian Classification by Feedforward Connectionist System. In Masulli and P.G. Morasso and A. Schenone, editors, *Neural Networks in Biomedicine*. pp. 145-162, 1994, World Scientific, Singapore.

[14] F. Masulli and L. Studer: Neuro-fuzzy system for chaotic time series forecasting. In Application of Soft Computing, *Proc. of the SPIE,*3165: 204-2158, 1997, ed. Bellingham WA.

[15] R. Parenti, C. Penno, F. Perotti, A. Rebizzo and D. Ronconi: Steam Turbine Vibration Diagnostic System based Neural Hologram. In POWERGEN'98, 1998, Milan.

[16] R. Parenti and C. Penno: A neural network based approach to the on-line vibration diagnostic of rotating makines. In EANN'98, Engineering Application of Neural Network, 1998, Gibraltar.

[17] D.E. Rumelhart, G.E. Hinton and R.J. Williams: Learning Internal Representations by Error Propagation. In D.E. Rumelhart, D.E. and J.L. McClelland,

Eds): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, pp. 318-362, MIT Press, Cambridge.

[18] D.W. Ruck, S.K. Rogers, M. Kabrisky, M.E. Oxley and B.W. Suther: The multilayer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Transactions on Neural Networks*, 1: 296-298, 1990.

[19] L. Studer and F. Masulli: Building a neuro-fuzzy system to efficiently forecast chaotic time series. *Nuclear Instruments and Methods A*, a389: 264-267, 1997.

[20] F. Takens: Detecting strange attractors in turbulence. In D.A. Rand and B.-S. Young, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pp. 366-381, Warwick 1980, 1981, Springer-Verlag, Berlin.

[21] J. Vastano and L. Rahman: Information transport in spatio-temporal chaos. *Physical Review Letters*, 72: 241-275, 1989.