Nonlinear Dynamic Systems Modeling Using Neural Networks with Reduced Interference on the Plant Operation

Joarez B. Monteiro
e-mail: joarez@lps.ufrj.br

Luiz P. Calôba
e-mail: caloba@lps.ufrj.br

COPPE/EE/UFRJ C.P. 68504
CEP 21945-970, Rio de Janeiro, RJ, Brasil - Fax: (021) 260-1092

Abstract - This paper presents a nonlinear dynamic plant identification technique using neural networks that causes small interference in the normal operation point of the system. So, it allows to identify plants that don't accept the frequently proposed procedure of applying as input a large amplitude white noise signal capable of exploiting the whole amplitude and frequency ranges of interest of the system.

This method identifies the order of the system and may also furnish topological information indicating the presence of linearities, without any previous knowledge of the plant characteristics.

The used neural network is a slightly modified feedforward network with delayed output feedback and the error minimization is performed by a modified version of a simulated annealing algorithm. All tested examples pointed out that stable plants bring out precise and stable models.

Keywords - Nonlinear dynamic systems, Modeling and Identification, Neural networks.

I. INTRODUCTION

Systems modeling is a fundamental engineering area whose objective is to build a model capable of producing an output y(k) that approaches the output of the plant yd(k), with a small error e(k), when both are submitted to the same input signal u(k) (Figure 1).

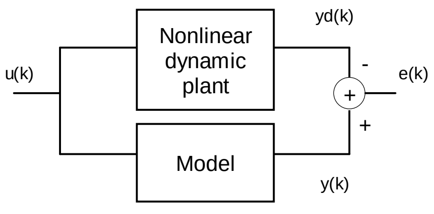


Fig. 1 The Plant and its Model

The neural networks modeling technique is a tool capable of generating compact and accurate solutions, associated with a great computational simplicity. It has been largely used for modeling static nonlinear systems [1] [2]. However, to model real dynamic plants using recurrent neural networks may be complex. To guarantee the stability of the model it is necessary to precisely identify the behavior of the plant, mainly in the region of its singularities[1].

Some studies presenting dynamic systems identification processes using recurrent neural networks [3][4] or memory neural networks [5] proposes to get the data for the training set (input-output pairs) by the application of a large amplitude and wide frequency spectrum signal to the input of the system. The large amplitude is necessary in order to obtain at the system output the same dynamic range of the operation signal, and conveniently characterize the nonlinearities. The large frequency spectrum is necessary to exploit high frequency singularities not excited by the normal operation signals. Failing to fulfill these two conditions will lead to a model that poorly represents the nonlinearities or usually presents high frequency instability. Although theoretically effective, in practice this procedure imposes a severe limitation because most real plants cannot or shall not excessively deviate from its normal operation point.

The identification method proposed in this paper causes just a small interference in the operating point of the plant. It's able to model the system identifying its order and, eventually, to give information on its topology.

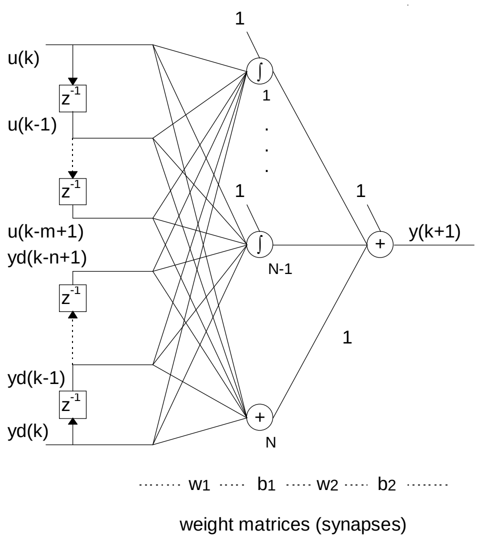
II. THE GENERAL NEURAL NETWORK STRUCTURE

Without loss of generality, let us restrict to SISO plant predictors. A general neural network predictor for such plants may be implemented by the two layers neural network shown in Figure 2. The neural network input **x**(k) at time k is composed by the present and delayed plant inputs u(.) and outputs yd(.), and the neural network output y(k+1) is the prediction of the plant output at time k+1, yd(k+1).

1 In this work, "singularities" means the singularities of the system linearized at all its possible operating points.

The structure of the proposed neural network uses a linear neuron at the output, and an intermediate layer with (N-1) sigmoidal neurons (tanh activation function) plus a linear neuron. The objective of the linear neurons is to accelerate the identification and to allow the model simplification for plants with quasi-linear partial transfer functions. This general model is able to identify all types of plants, as it corresponds to model IV accordingly to the nonlinear dynamic systems classification proposed by Narendra and Parthasarathy [3] (Figure 4). But now the linear transmissions identification made possible by introduction of the linear neurons will allow this model to degenerate in the simplest models I, II and III when possible.

An open loop training method, the series-parallel identification model, is used to avoid instability during the training process [3].



$$\underline{\mathbf{x}}(k) = [u(k),\ u(k-1),\ \ldots,\ u(k-m+1),\ yd(k),\ yd(k-1),\ \ldots,\ yd(k-n+1)]^T$$

$$y(k+1) = \sum_{i=1}^{N-1} w2_i \cdot \mathrm{tgh}\left(\sum_{j=1}^{m+n} w1_{ij}\, x_j(k) + b1_i\right) + b2 + \sum_{j=1}^{m+n} w1_{Nj}\, x_j(k)$$

Fig. 2 Neural network structure during the training period and its mathematical behavior. The symbols + and ∫ hold for linear and tanh neurons, respectively. Note that biasing the linear neuron in the intermediate layer is redundant. At operation, yd(k) is replaced by y(k), the delayed NN output, closing the feedback loop around the neural network, as shown in Plant IV of Figure 4.

III. PILOT SIGNAL

For a correct identification, the training set must contain enough information about the nonlinearities and the singularities of the plant. Because the normal operation signal usually swings the whole output dynamic range, it allows to characterize the nonlinearities. But as it is usually strongly frequency limited, it has the inconvenience of not to produce the appropriate information about the system high frequency singularities. The spectral analysis of the plant output signal suggests the maximum frequency f_o (say, the –40 dB frequency) where information is still supplied by the normal operation signal. To overcome that inconvenience we add a low amplitude wide frequency spectrum pilot signal q(.) to the normal operation input signal of the plant, Figure 3, during the training procedure. The pilot signal may be white noise or, more practically, a square wave with small and constant amplitude, zero mean and random pulse width (uniform distribution). The limits of the pulse width variation must be chosen such that for frequencies from f_o to the highest frequency of interest to the plant identification the pilot signal spectrum is approximately flat. In this way only a small disturbance is introduced in the plant operation point by the pilot signal, but now the output signal will contain complete information on the system: the amplitude and frequency characteristics necessary to the plant identification.

IV. IDENTIFICATION METHOD

The Figure 3 presents the identification process block diagram using the series-parallel training method.

A. *Amplitude adjustment of the pilot signal*

Basically, the pilot signal q(.) amplitude adjustment will depend on two characteristics: the plant frequency response and its internal noise level r(.). The more accentuated the high frequency attenuation and the plant noise, higher it must be the input pilot signal amplitude in order to assure that the output signal contains enough information about the system in the high frequency range of interest.

To adjust the pilot signal amplitude we construct a low pass (LP) and a high pass (HP) filter with cutoff frequency f_o (say –3 dB), and accept as an estimative of the plant output signal and noise at normal operation (q(.)=0) the outputs of the LP and HP filters, respectively. Then we apply the pilot signal q(.) and use the output signal/noise ratio in dB before (SNR1) and after (SNR2) the application of the pilot signal estimated using the above approximation to adjust the pilot signal amplitude.

The linear rate between the pilot signal power Pq and the noise power Pr is given by:

$$\frac{Pq}{Pr} \approx 10^{\left(\frac{SNR1-SNR2}{10}\right)} - 1 \qquad (1)$$

The approximation holds because the pilot signal is much smaller than the operation signal.

B. *Identification method*

Although the output error yd(.) - y(.) contains all the informations on the system, those informations are strongly unbalanced: the low amplitude of the pilot signal causes very little influence on the output error. The high and low pass filtering allows to compute separately the high and low frequencies errors, and to compose a balanced cost function C (2).

$$C = k1 \cdot lme + k2 \cdot hme = \qquad (2)$$

$$= k1\left(\frac{1}{N}\sum_{k=1}^{N} el^2(k)\right) + k2\left(\frac{1}{N}\sum_{k=1}^{N} eh^2(k)\right)$$

Where,

$$k2 = \frac{\sum_{k=1}^{N} ydl^2(k)}{\sum_{k=1}^{N} ydh^2(k)} = \left.\frac{lme}{hme}\right|_{y(k)=0,\forall k} \qquad (3)$$

Initially k1=1 and k2 balances the importance of the errors, as k2 equals the power ratio between the low and high frequency components of the plant output signal.

Along the training procedure, every time each parcel of the cost function becomes greater than the other, it will have more chance of being reduced in the next optimization step. So, both low (lme) and high (hme) frequency mean square error are reduced during the process. Because k2 >> k1 = 1, a small decreasing in hme will cause great reduction on the total cost. Therefore, high frequency optimization will take place first but the lme continues reducing until it reaches about 10% of the high frequency parcel. At this time it turns difficult to reach lower costs and the process stops progressing fast.

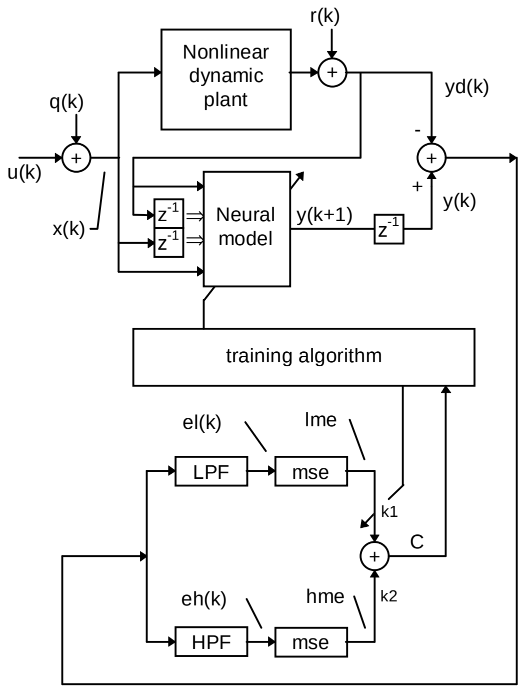


Fig. 3 Identification process: block diagram. q(.) is the pilot signal and r(.) is the plant internal noise.

To increase the efficacy of the training process, k1 shall be gradually increased to stimulate the lme reduction, but always respecting k1 ≤ k2. At the end of the optimization process k1 = k2 and both lme and hme will have the same magnitude order.

C. *Order identification and topological information*

When the order of the plant is unknown, the problem of identification becomes more complex. Beginning with an arbitrary higher order neural model it will be necessary to eliminate the unimportant inputs of the neural network based on the quantification of the importance of each one in the composition of the output signal.

Although it's a nonlinear system , the individual response of each input may be used as an efficient way to quantify its relative importance as can be seen through the results presented in section IV. Using the training data as shown in sections II and III-A, the order identification can be obtained by the following heuristic procedure:

1. Start the identification method (section IV-B) with a neural model composed by an arbitrary large number of inputs for both groups, u(.) and yd(.);

2. From the trained neural network, get the individual response $y_i(k+1)$ for each network input i, vanishing all the others network inputs;

3. Calculate the variance σ_i^2 for all outputs $y_i(k+1)$ and normalize them by the greatest of its group (u(.) or yd(.)) in order to obtain the relative importance of each input in the composition of the output signal;

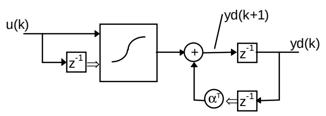
$$\sigma_i^2 = \frac{1}{N}\sum_{k=1}^{N}[y_i(k+1) - \bar{y}_i(k+1)]^2$$

4. Inputs with small relative contribution (say less than 5%) are discarded. At this time, it is recommended to eliminate no more than one input of each group;

5. Repeat steps 1 to 4 until isn't possible to discard any input.

Besides the order identification, visual inspection of the individual transfer curves from each network input to the network output may suggests the existence of linearities. So, eventually the neural model achieved may be simplified to plants I, II or III. This procedure and its results are presented in section V-C.

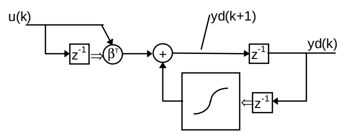
Type I plant



$$yd(k+1) = 0{,}3yd(k) + 0{,}6yd(k-1) + \Phi[x(k)]$$

$$\text{where:} \quad \Phi(x) = x^3 + 0{,}3x^2 - 0{,}4x$$

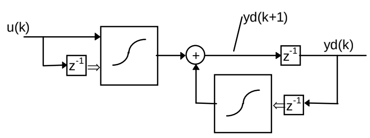
Type II plant



$$yd(k+1) = x(k) + \Phi[yd(k), yd(k-1)] \quad ;$$

$$\text{where:} \quad \Phi[yd(k), yd(k-1)] = \frac{yd(k)\cdot yd(k-1)\cdot[yd(k) - 2.5]}{1 + yd(k)^2 + yd(k-1)^2}$$

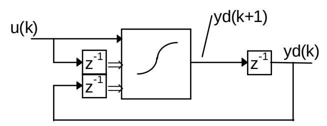
Type III plant



$$yd(k+1) = \Psi[x(k)] + \Phi[yd(k)] \quad ;$$

$$\text{where:} \quad \Psi[x(k)] = x(k)^3 \quad \text{and}, \quad \Phi[yd(k)] = \frac{yd(k)}{1 + yd(k)^2}$$

Type IV plant



$$yd(k+1) = \Phi[x(k), x(k-1), yd(k), yd(k-1), yd(k-2)] \quad ;$$

$$\text{where:} \quad \Phi[x(k), x(k-1), yd(k), yd(k-1), yd(k-2)] =$$

$$= \frac{yd(k)\cdot yd(k-1)\cdot yd(k-2)\cdot x(k-1)\cdot[yd(k-2) - 1] + x(k)}{1 + yd(k-2)^2 + yd(k-1)^2}$$

Fig. 4 Plants presented by Narendra and Parthasarathy [3].

D. *Optimization Algorithm*

The error minimization for training the neural network was performed by a modified version of a simulated annealing algorithm [6] that will be detailed in other paper [7].

V. RESULTS

In order to compare results, we identified the plants presented by Narendra and Parthasarathy [3], whose structures and difference equations are reproduced in Figure 4.

The input training signal was a simple low frequency sinusoid. For the verification of the results we choose, for its complexity, a test signal proposed by Sasty et al. [4]:

$$x(k) = \begin{cases} \sin(\pi k/25) & , k < 250 \\ +1 & , 250 \le k < 375 \\ -1 & , 375 \le k < 500 \\ 0{,}3\sin(\pi k/25) + 0{,}1\sin(\pi k/32) + 0{,}6\sin(\pi k/10) & , 500 \le k < 900 \end{cases}$$

A neural model with 7 neurons in the intermediate layer (6 tanh + 1 linear) was used in all simulations. As pilot signal it was used a square wave with constant amplitude, zero mean and random pulse width. To quantify the accuracy between the plant and the model outputs, the values of signal/error ratio (SER) are indicated for various time gaps. An error smaller than 20% corresponds to a SER higher than 14 dB.

A. *Order identification*

As an example, the step by step results of the order identification process applied to type III plant are listed on Table 1. The line "structure" presents the models in the order they appear as a result of the application of the method of order reduction described in the section IV-C. Values in boxes point the discarded inputs in each model.

Beginning with a arbitrary higher order neural model (structure 1) the final result (structure 3) correctly identified the dynamic structure of the plant.

Satisfactory results were obtained for all four plant types proposed in [3]. In some cases the plant was identified without an exact correspondence of the order. But in these cases the rejected or changed inputs didn't had significative importance, indicating that a different order suitable model of the plant could also be obtained. But stable plants always generated stable models.

The Figure 5 shows the output of the type III plant and that of its model.

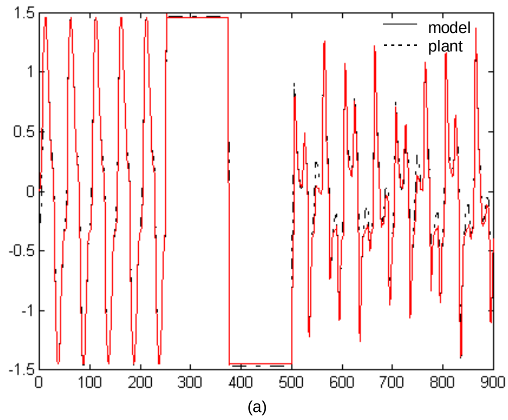
B. *Noisy plant identification*

Adjusting the pilot signal power (Pq) for about twice the noise power (Pr) good results was obtained in the noisy plants identifications (SNR1 around 43 dB). Only type I plant required Pq ≈ 4Pr.

The Table 2 shows the power rate (Pq / Pr) adjusted for all plants and the Figure 6 presents both the neural model and type I plant outputs as an example of noisy plant identification.

Table 1 Type III plant : Order identification

pilot signal			2%		
structure			1	2	3
NN output y(.)			[0 1 3]	[0 3]	[0]
NN input u(.)			[0 1 2]	[0 1]	[0]
relative importance of the NN inputs	u	0	1.0000	1.0000	1.0000
	u	1	0.0603	0.0157	
	u	2	0.0024		
	y	0	1.0000	1.0000	1.0000
	y	1	0.0151		
	y	3	0.1087	0.0226	



Time	0-250	250-500	500-900	0-900
SER (dB)	24,70	29,05	14,68	22,03

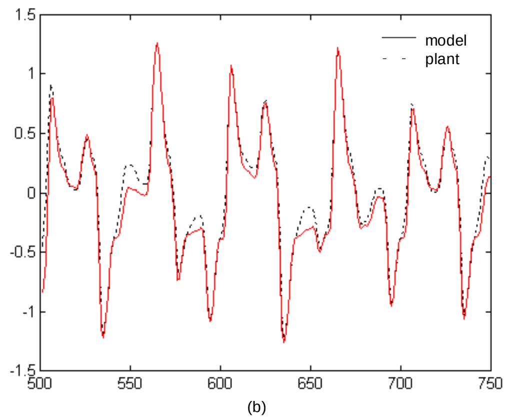
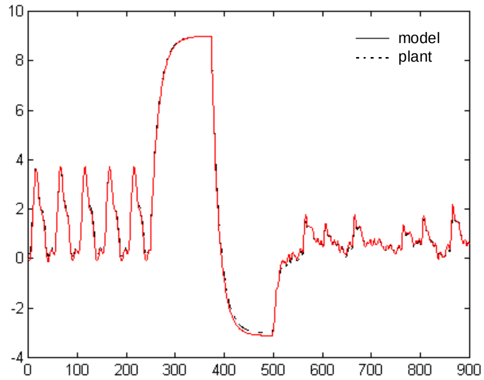


Fig. 5 Type III plant without noise:
(a) the plant and the trained neural model outputs;
(b) detail of time interval 500-750.

Table 2 Type I to IV plants: Adjusted signal pilot to noise power rate (Pq / Pr).

Plant	Pilot Signal versus Input Signal magnitude (%)	SNR1 (dB)	SNR2 (dB)	Pq / Pr
type I	8,4	42,9	35,9	3,95
type II	5,9	42,9	38,1	2,04
type III	1,5	42,9	38,0	2,12
type IV	2,4	43,0	38,2	2,00



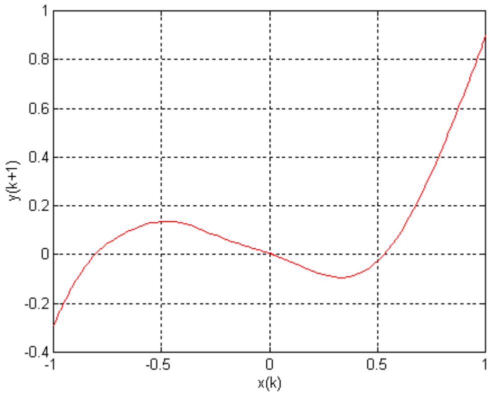
Time	0-250	250-500	500-900	0-900
SER (dB)	24,79	33,39	21,32	30,73

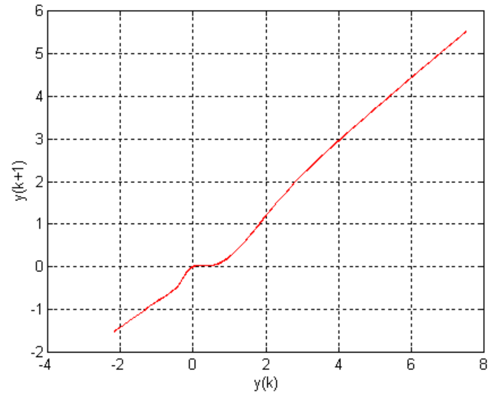
Fig. 6 Type I noisy plant: the plant and the trained neural model outputs (Pq ≈ 4Pr).

C. *Topological informations*

After getting a trained model able to represent the plant, we may plot the transfer curves for each neural network input component (vanishing the others) to the output. Visual inspection then suggests the existence or not of linearities. Inputs with quasi-linear transfers to the output may be tied up only to the linear neuron of the hidden layer, reducing the synapses number. A new training should then be performed to fine tune the simplified structure.

As an example, the transfer curves of the type I plant are shown in Figure 7.





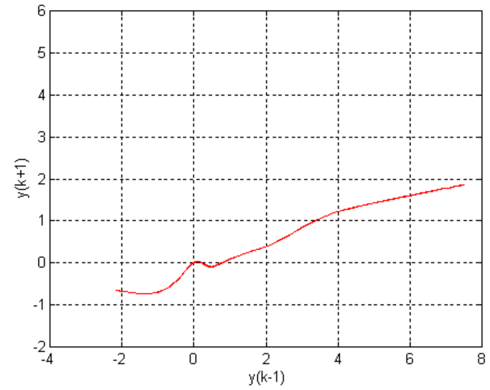


Fig. 7 Type I plant: input-output transfer curves

They suggest the possibility of linearization of both second yd(k) and third yd(k-1) elements of the network input, allowing the construction of a specific neural model with a simplified structure (Figure 8), where the dashed line synapses are eliminated. Evidently, when the topological simplifications are completed, the linear neuron in the intermediate layer is eliminated and the connections are made directly to the output layer neuron.

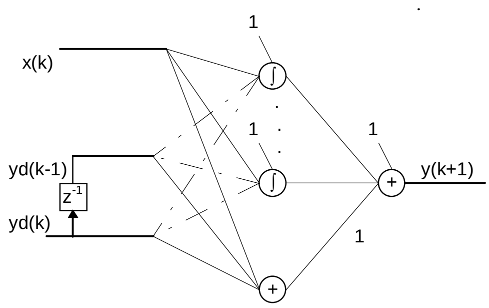


Fig. 8 Simplified structure based on the topological information.

VI. CONCLUSIONS

Because of its small interference in the operation point and for requesting no previous knowledge of the characteristics of the plant, the proposed method seems to be a progress in the modeling and identification techniques of nonlinear dynamic plants that cannot or should not excessively deviate from its normal operation point.

The automatic order identification and the eventual information on the topology are other interesting characteristics of the method.

Two aspects played an important role in coming up with the precise identification of the plant singularities. First, the fact that the modeling is efficiently made not only in the frequency operation range, but in the whole frequency spectrum of the system output. Second, the use of a training algorithm (modified simulated annealing) able of reaching solutions very close to the global minimum. This suggests that, applying the proposed method, if the input-output pairs are generated from a stable plant, the model obtained will be precise and stable. The obtained results evidence this observation: all tested cases always came up with stable models in spite of no specific care taken in the sense of guaranteeing their stability.

References

[1] HAYKIN, S., *Neural Networks: A Comprehensive Foundation*, 1 ed., New York, Macmillan College Publishing Company, 1994.

[2] CHEN, T., CHEN, H., “Approximations of Continuous Functionals by Neural Networks with Application to Dynamical Systems”, *IEEE Transactions on Neural Networks*, v. 4, n. 6, pp. 910-918, Nov. 1993.

[3] NARENDRA, K. S., PARTHASARATHY, K., “Identification and Control of Dynamical Systems Using Neural Networks”, *IEEE Transactions on Neural Networks*, v. 1, n. 1, pp. 4-27, Mar. 1990.

[4] ALIPPI, C., PIURI, V., “Identification of Nonlinear Dynamic Systems in Power Plants”. In: *38th Midwest Symposium on Circuits and Systems*, pp. 493 - 496, Rio de Janeiro, Aug. 1995.

[5] SASTY, P. S., SANTHARAM, G., UNNIKRISHNAN, K. P., “Memory Neuron Networks for Identification and Control of Dynamical Systems”, *IEEE Transactions on Neural Networks*, v. 5, n. 2, pp. 306-319, Mar. 1994.

[6] MONTEIRO, J. B., *Identificação de Sistemas Dinâmicos Não Lineares Usando Redes Neurais*. M.Sc. Thesis, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 1997.

[7] MONTEIRO, J. B., CALÔBA, L. P., “Simulated Annealing: Fast Convergence with Initial Conditions Independence”. This Proceedings.