A NN-based Model for Time series forecasting in function of energy associated of series

C. RODRÍGUEZ RIVERO, J. PUCHETA, J. BAUMGARTNER, M. HERRERA, D. PATIÑO AND B. KUCHEN
Departments of Electrical and Electronic Engineering, Mathematics Research Laboratory applied to Control (LIMAC), Faculty of Exact, Physical and Natural Sciences - National University of Córdoba, Córdoba, ARGENTINA.
cristian.rodriguezrivero@gmail.com, julian.pucheta@gmail.com, josef.s.baumgartner@gmail.com
Departments of Electrical Engineering, at Faculty of Sciences and Applied Technologies
National University of Catamarca, Catamarca, Argentina.
martinictohache@gmail.com
Institute of Automatics INAUT, Faculty of Engineering-National University of San Juan, San Juan, ARGENTINA.
hdpatino@gmail.com, bkuchen@unsj.edu.ar

Abstract:- In this work an algorithm to adjust parameters of time series forecasting in function of energy associated of series using a feed-forward NN-based nonlinear autoregressive model is presented. The criterion for fitting comprises to yield value time series from forecasted time series area. These values are approximated by the NN to generate a primitive calculated as an area by the predictor filter. The NN output will tend to approximate the current value available from the series which has the same Hurst Parameter as the real time series. The approach is tested over a time series obtained from samples of the Mackey-Glass delay differential equations (MG) and serve to be applied for meteorological variables measurements such as soil moisture series, daily rainfall and monthly cumulative rainfall time series forecasting.

Keywords:- Neural networks, time series forecast, primitive, Hurst's parameter, Mackey-Glass.

1 Introduction

Natural phenomena prediction is a challenging topic, useful for control problems from agricultural activities and decision-making that helps the producer to decide. There are several approaches based on NN that face the rainfall forecast problem for energy demand purposes [5], for water availability [18] and seedling growth [21] by taking an ensemble of measurement points [12], [14]. Here, the proposed approach is based on the classical NAR filter using time lagged feed-forward neural networks, where the data from the MG benchmark equation whose forecast is simulated by a Monte Carlo [4] approach. The number of filter parameters is put function of the roughness of the time series, in such a way that the error between the smoothness of the time series data and the forecasted data modifies the number of the filter parameters.

1.1 MG and fBm Overview

The MG equation serves to model natural phenomena and has been used by different authors to perform comparisons between different techniques for foretelling and regression models [7] [19]. Here we propose an algorithm to predict values of time series taken from the solution of the MG equation [9]. The MG equation is explained by the time delay differential equation defined as,

$$\dot{y}(t) = \frac{\alpha y(t-\tau)}{1+y^{c}(t-\tau)} - \beta y(t), \quad (1)$$

where α, β, and c are parameters and τ is the delay time. According as τ increases, the solution turns from periodic to chaotic. Equation is solved by a standard fourth order Runge-Kutta integration step. By setting the parameter β ranging between 0.1 and 0.9 the stochastic dependence of the deterministic time series obtained varies according to its roughness. The performance of the proposed method is tested with the SMAPE index and it is compared with a traditional NN based predictor. Due to the random process, it is proposed to use the Hurst's parameter H in the learning process to modify on-line the number of patterns, of iterations and filter inputs. This *H* gives information about the roughness of a signal, and also to determine its stochastic dependence. The definition of the Hurst's parameter is defined by Mandelbrot through its stochastic representation,

$$B_H(t)=\frac{1}{\Gamma\left(H+\frac{1}{2}\right)}\left(\int_{-\infty}^{0}\left((t-s)^{H-\frac{1}{2}}-(-s)^{H-\frac{1}{2}}\right)dB(s)+\int_{0}^{t}(t-s)^{H-\frac{1}{2}}dB(s)\right) \quad (2)$$

where, $\Gamma(\cdot)$ represents the Gamma function

$$\Gamma(\alpha)=\int_{0}^{\infty}x^{\alpha-1}e^{-x}dx \quad (3)$$

and $0<H<1$ is called the Hurst parameter. The integrator B is a stochastic process, ordinary Brownian motion. Note, that B is recovered by taking $H=1/2$ in (2) and B is defined on some probability space (Ω, F, P). Thus, an fBm is a time continuous Gaussian process depending on the so-called Hurst parameter $0<H<1$. The ordinary Brownian motion is generalized to $H=0.5$, and its derivative is the white noise. The fBm is self-similar in distribution and the variance of the increments is defined by

$$Var(B_H(t)-B_H(s))=\nu|t-s|^{2H} \quad (4)$$

where ν is a positive constant.

This special form of the variance of the increments suggests various ways to estimate the parameter H . In fact, there are different methods for computing the parameter H associated to Brownian motion [6]. In this work, the algorithm uses a wavelet-based method for estimating H from a trace path of the fBm with parameter H [8].

1.2 NN Approach overview

One of the motivations for this study follows the closed-loop control scheme [16] where the controller considers meteorological future conditions for designing the control law as shown in Fig. 1. In that scheme the controller considers the actual state of the crop by a state observer and the meteorological variables, referred by $x(k)$ and R_o, respectively. However, in this paper only the controller's portion concerning with the prediction system is presented by using a benchmark time series.

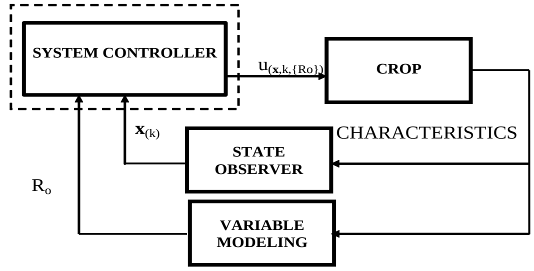

Fig. 1. Closed-loop PC-based control approach.

The main contribution of this work is in the learning process, which employs the Levenberg-Marquardt rule and considers the long or short term stochastic dependence of passed values of the time series to adjust at each time-stage the number of patterns, the number of iterations, and the length of the tapped-delay line, in function of the Hurst's value, H of the time series. According to the stochastic characteristics of each series, H can be greater or smaller than 0.5, which means that each series tends to present long or short term dependence, respectively. In order to adjust the design parameters and show the performance of the proposed prediction model, solutions of the MG equation are used. The NN-based nonlinear filter is applied to the time series obtained from MG to forecast the next 18 values out of a given historical data set of 102 values.

2 Problem statement

The best prediction of the present values from a random (or pseudo-random) time series is desired. The predictor system may be implemented using an autoregressive model-based nonlinear adaptive filter. In this work, time lagged feed-forward neural networks are used. The adaptive filter output will be the one-step prediction signal. In Fig. 2 the block diagram of the nonlinear prediction scheme based on a NN filter is shown. Here, a prediction device is designed such that starting from a given sequence $\{x_n\}$ at time n corresponding to a time series it can be obtained the best prediction $\{x_e\}$ for the following sequence of 18 values. Hence, it is proposed a predictor filter with an input vector l_x, which is obtained by applying the delay operator, Z^{-1}, to the sequence $\{x_n\}$. Then, the filter output will generate x_e as the next value, that will be equal to the present value x_n. So, the prediction error at time k can be evaluated as:

$$e(k)=x_n(k)-x_e(k) \quad (5)$$

which is used for the learning rule to adjust the NN weights.

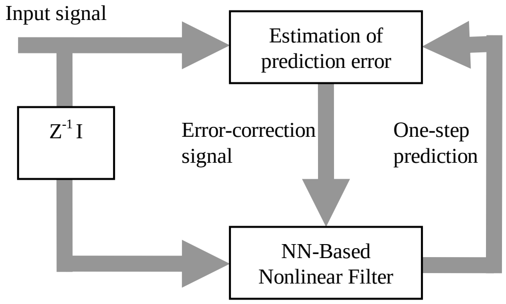

Fig. 2. Block diagram of the nonlinear prediction.

The coefficients of the nonlinear filter are adjusted on-line in the learning process, by considering a criterion that modifies at each pass of the time series the number of patterns, the number of iterations and the length of the tapped-delay line, in function of the Hurst's value H calculated from the time series. According to the stochastic behavior of the series, H can be greater or smaller than 0.5, which means that the series tends to present long or short term dependence, respectively [17].

3 Proposed approach

3.1 The Proposed Learning Process

The NN's weights are tuned by means of the Levenberg-Marquardt rule, which considers the long or short term stochastic dependence of the time series measured by the Hurst's parameter H. The proposed learning approach consists of changing the number of patterns, the filter's length and the number of iterations in function of the parameter H for each corresponding time series. The learning process is performed using a batch model. In this case the weight updating is performed after the presentation of all training examples, which forms an epoch. The pairs of the used input-output patterns are

$$(x_i, y_i) \quad \mathrm{i} = 1,2,\ldots,N_p \tag{6}$$

where, x_i and y_i are the corresponding input and output pattern respectively, and N_p is the number of input-output patterns presented at each epoch. Here, the input vector is defined as:

$$X_i = \mathrm{Z}^{-1}\mathrm{I}(\{x_i\}) \tag{7}$$

and its corresponding output vector as:

$$Y_i = x_i. \tag{8}$$

Furthermore, the index i is within the range of N_p given by

$$l_x \le i_t \le 2 \cdot l_x. \tag{9}$$

where l_x is the dimension of the input vector.

In addition, the number of iterations performed by each epoch it is given by

$$l_x \le N_p \le 2 \cdot l_x \tag{10}$$

The proposed criterion to modify the pair (i_t,N_p) is given by the statistical dependence of the time series $\{x_n\}$, supposing that it is an fBm. The dependence is evaluated by the Hurst's parameter H, which is computed by a wavelet-based method [1] [8]. Then, a heuristic adjustment for the pair (i_t,N_p) in function of H according to the membership functions shown in Fig. 3 is proposed.

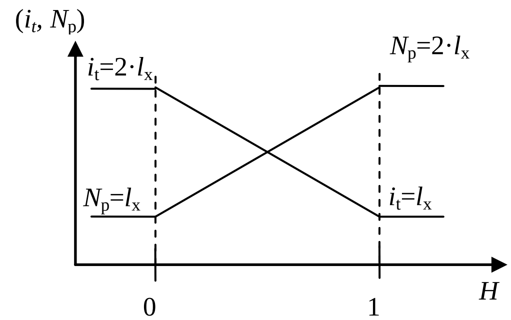

Fig. 3. Heuristic adjustment of (i_t,N_p) in terms of H after each epoch.

In order to predict the sequence $\{x_e\}$ one-step ahead, the first delay is taken from x_n data and used as input. Therefore, the output prediction can be denoted by

$$\mathrm{x_e}(+1) = \mathrm{F_p}(\mathrm{Z}^{-1}\mathrm{I}(\{\mathrm{x_n}\})) \tag{11}$$

where $\mathrm{F_p}$ is the nonlinear predictor filter operator, and $\mathrm{x_e}(n+1)$ is the output prediction at n+1.

$$\mathrm{I} \in \Re^{lx} \times \Re^{lx} \tag{12}$$

3.2 Approximation by Primitive

The area resulting of integrating the data time series of MG equation is the primitive that is obtained by considering each value of time series its derivate;

$$\int_{t_k}^{t_{k+1}} y_t\,dt \cong y_t\left(t_{k+1} - t_k\right) \tag{13}$$

where y_t is the original value time series. The area approximation by its periodical primitive is:

$$I_{t_n} = \int_{t_n}^{t_{n+p}} y_t\,dt = Y_t\Big|_{t_n}^{t_{n+p}}, n = 1,2,\ldots N. \tag{14}$$

During the learning process, those primitives are calculated as a new entrance to the NN, in which the prediction attempts to even the area of the forecasted area to the primitive real area predicted. The real primitive integral is used in two instances, firstly from the real time series an area is obtained and run by the algorithm proposed. The H parameter from this time series is called H_A. On the other hand, the data time series is also forecasted by the algorithm, so the H parameter from this time series is called H_S. Finally, after each pass the number of inputs of the nonlinear filter is tuned —that is the length of tapped-delay line, according to the following heuristic criterion. After the training process is completed, both sequences -$\{\{I_n\}, \{I_e\}\}$ and $\{\int\{y_n, y_e\}\}$, in accordance with the hypothesis that should have the same H parameter. If the error between H_A and H_S is greater than a threshold parameter θ the value of l_x is increased (or decreased), according to $l_x \pm 1$. Explicitly,

$$l_x = l_x + sign(\theta). \tag{15}$$

Here, the threshold θ was set about 1%.

4 Main results

4.1 Generations of areas from MG equations

Primitives of time series are obtained from the MG equations with the parameters shown in Table 1, with τ=100 and α=20. This collection of coefficients was chosen to generate time series whose H parameters vary between 0 and 1. The chosen one was selected in accordance to its roughness.

Table 1. Parameters to generate the times series.

Series No.	β	H
1	0.32	2.71
2	1.6	0.73

4.2 Performance measure for forecasting

In order to test the proposed design procedure of the NN -based nonlinear predictor, an experiment with time series obtained from the MG solution was performed. The performance of the filter is evaluated using the Symmetric Mean Absolute Percent Error (SMAPE) proposed in the most of metric evaluation, defined by

$$SMAPE_S = \frac{1}{n}\sum_{t=1}^{n}\frac{|X_t - F_t|}{(X_t + F_t)/2}\cdot 100 \tag{16}$$

where t is the observation time, n is the size of the test set, s is each time series, X_t and F_t are the actual and the forecasted time series values at time t respectively. The SMAPE of each series ***s*** calculates the symmetric absolute error in percent between the actual X_t and its corresponding forecast value F_t, across all observations t of the test set of size n for each time series s.

4.3 Prediction Results for the MG Time Series

Each time series is composed by sampling the MG solutions. However, there are three classes of data sets: one is the original time series used for the algorithm in order to give the forecast, which comprises 102 values. The other one is the primitive obtained by integrating the values of original time series and the last one is used to compare if the forecast is acceptable or not where the 18 last values can be used to validate the performance of the prediction system, which 102 values form the data set, and 120 values constitute the Forecasted and the Real ones. A comparison is made between a linear and nonlinear NN dependent and independent filter and an ARMA predictor filter.

The Monte Carlo method was used to forecast the next 18 values from MG time series and its primitive acquired by integration. Such outcomes are shown from Fig. 4 to Fig. 8 . The plot shown in Fig. 4 is provided for a linear ARMA filter outcome.

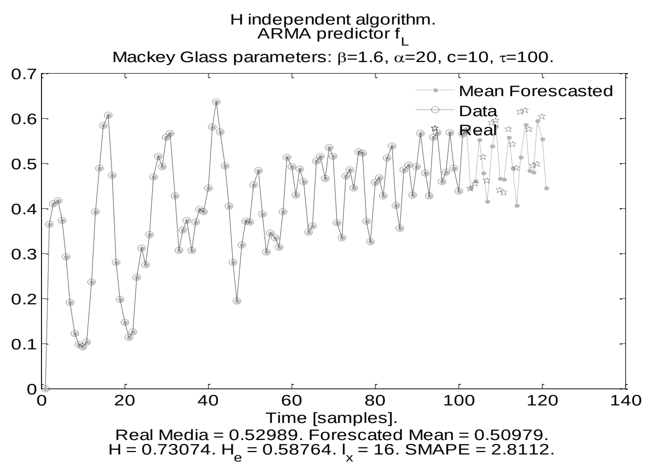

Fig. 4. Primitive of MG time series obtained from ARMA predictor filter.

In the figure, the legend “Data” represents the values obtained by Eq. (11), and the legend “Real” denotes the actual values -not available in practice- used here for verification purposes only. From time k equal 103 to time 120 the inputs of Eq. (11) include the outputs delayed one time interval. The obtained time series has a mean value, which is indicated at the foot of the figure by “Forecasted Mean”. The “Real Mean” it is not available at time 102. In the learning process, the primitive is calculated as area by the predictor filter, in which each value of time series represents its derivate. In Fig. 7, the MG equation in which the H parameter is equal to 2.71 was discarded due to its low roughness.

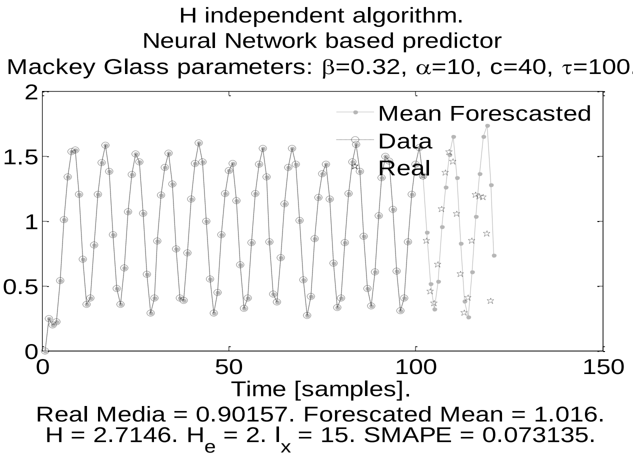

Fig. 5. Primitive of MG time series with β=0.32.

The algorithm proposed to predict the area related with the energy of the time series is shown in Fig. 6 and Fig. 7. These are calculated by a dependent and independent NN filter.

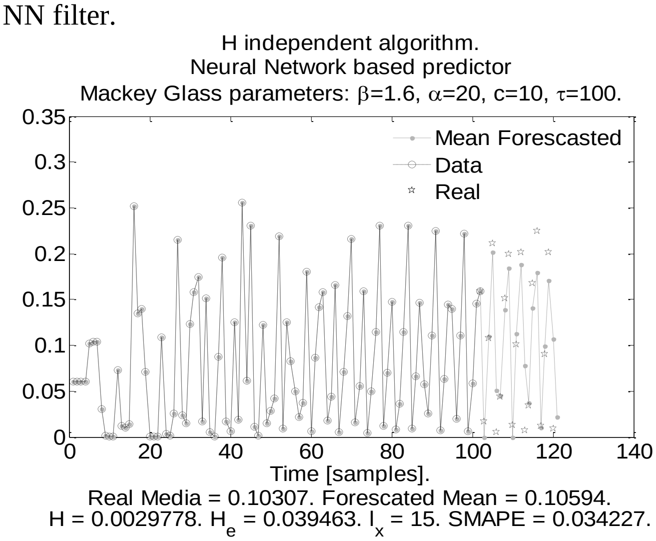

Fig. 6. H independent algorithm.

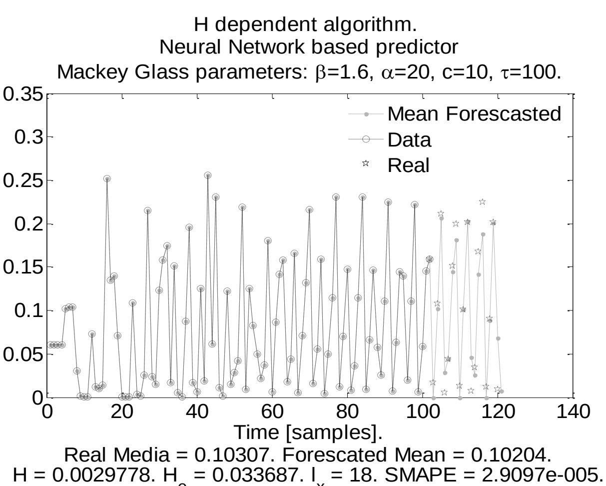

Fig. 7. H dependent algorithm.

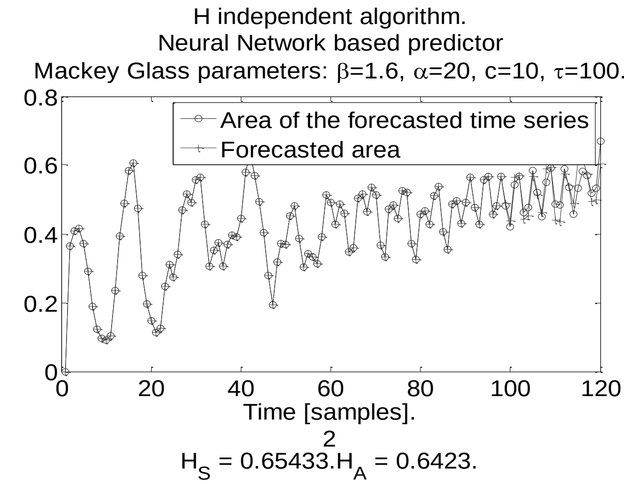

Fig. 8. Comparisons.

4.4 Comparative Results

The performance of the stochastic NN-based predictor filter is evaluated through the SMAPE index —Eq. (16), shown in Table 2 and

Table 3 along the time series from MG solutions with β=1.6.

Table 2. Figures obtained by the proposed approach

Series No.	H	H_e	Real mean	Mean Forecasted	SMAPE
Fig. 6	0.73	0.587	0.529	0.509	2.811
Fig.8	0.002	0.039	0.103	0.105	0.034
Fig.9	0.002	0.033	0.103	0.102	2.9 10-5

Table 3. Comparisons obtained by the proposed approach

Series No.	H_S	H_A
Fig. 10	0.6543	0.6423

The comparison between the deterministic [19], the stochastic approach [22] and the present forecasted time series is shown through Fig. 6 to Fig. 8. In addition, an area of a primitive value acquired of MG time series was incorporated in order to use the proposed approach. The SMAPE index for each time series is obtained by a deterministic NN-based filter, which uses the Levenberg–Marquardt algorithm with fixed parameters

(i_t, N_p). In addition, the results of the SMAPE obtained by the stochastic NN-based filter proposed here are also shown in Fig. 6 and Fig. 7 where the Forecasted mean is closer the Real media in Fig. 9 than in Fig. 8 due to the fact that H is dependent of the algorithm. Furthermore, the SMAPE value is improved of order of 10^{-5} for a class of time series with high roughness of the signal, in this case with β=1.6 which is one of worst condition for signal prediction.

5 Discussion

The assessments of the obtained results by comparing the performance of the proposed filter with the classic filter, both are based on NN. Although the difference between both filters resides only in the adjustment algorithm, the coefficients that each filter has, each ones performs different behaviors. In the four analyzed cases, the generation of 18 future values from 102 present values was made by each algorithm. The same initial parameters were used for each algorithm, but these parameters and the filter's structure are changed by the proposed algorithm that is not modified by the classic algorithm. The adjustment of the proposed filter, the coefficients and the structure of the filter are tuned by considering their stochastic dependency. It can be noted that in each of the figures —Fig. 4 to Fig. 8— the computed value of the Hurst's parameter is denoted either by H_e or H, both taken from the Forecasted time series or from the Data time series, respectively, since the Real time series (future time series) are unknown. Index SMAPE is computed between the complete Real time series (it includes the series Data) and the Forecasted one, as indicates the Eq. (16) for each filter. Note that there is no improvement of the forecast for any given time series, which results from the use of a stochastic characteristic to generate a deterministic result, such as a prediction.

6 Conclusions

In this work a NN-based autoregressive model for time series forecasting that considers the energy associated with the data series was presented. The learning rule proposed to adjust the NN's weights is based on the Levenberg-Marquardt method. Likewise, in function of the long or short term stochastic dependence of the time series evaluated by the Hurst parameter H, an on-line heuristic adaptive law was proposed to update the NN topology at each time-stage. The major result shows that the area predictor system supplied to time series has an optimal performance from several samples of MG equations, in particular, those whose H parameter has a high roughness of signal, which is assessed by H_S and H_A, respectively. This fact encourages us to be applied for meteorological variables measurements such as soil moisture series, daily rainfall and monthly cumulative rainfall time series forecasting when the observations are taken from a single point.

6.1 Acknowledgments

This work was supported by National University of Córdoba (UNC), FONCYT-PDFT PRH N°3 (UNC Program RRHH03), SECYT UNC, National University of Catamarca, National Agency for Scientific and Technological Promotion (ANPCyT) under grant PICT-2007-00526 and Departments of Electrotechnics - National University of Cordoba.

References:

[1] Abry, P.; P. Flandrin, M.S. Taqqu, D. Veitch., Self-similarity and long-range dependence through the wavelet lens. Theory and applications of long-range dependence, Birkhäuser, pp. 527-556. 2003.

[2] Bardet, J.-M.; G. Lang, G. Oppenheim, A. Philippe, S. Stoev, M.S. Taqqu. Semi-parametric estimation of the long-range dependence parameter: a survey. Theory and applications of long-range dependence, Birkhäuser, pp. 557-577. 2003.

[3] Bishop, C. Neural Networks for Pattern Recognition. Pp. 290 292. University Press. Oxford, 1995.

[4] Bishop, C. Pattern Recognition and Machine Learning. Springer. Boston, 2006.

[5] Chow, T.W.S.; Leung, C.T. Neural network based short-term load forecasting using weather compensation. *Power Systems, IEEE Transactions on*, Vol.11, Iss.4, Nov 1996, Pp. 1736-1742.

[6] Dieker, T. Simulation of fractional Brownian motion. MSc theses, University of Twente, Amsterdam, The Netherlands. 2004.

[7] Espinoza Contreras, Adriana Eliza. El Caos y la caracterización de series de tiempo a través de técnicas de la dinámica no lineal. Universidad Autónoma de Mexico. Campus Aragón. 2004.

[8] Flandrin, P. Wavelet analysis and synthesis of fractional Brownian motion". *IEEE Trans. on Information Theory*, 38, pp. 910-917. 1992.

[9] Glass L. and M. C. Mackey. From Clocks to Chaos, The Rhythms of Life. Princeton University Press, Princeton, NJ, 1988.

[10] Haykin, S. Neural Networks: A comprehensive Foudation. 2nd Edition, Prentice Hall. 1999.

[11] Istas, J.; G. Lang. Quadratic variations and estimation of the local Hölder index of a Gaussian

process. Ann. Inst. Poincaré, 33, pp. 407-436. 1994.

[12] Liu, J.N.K.; Lee, R.S.T. Rainfall forecasting from multiple point sources using neural networks. *In Proc. of the International Conference on Systems, Man, and Cybernetics*, Pages:429-434 Vol.3.1999.

[13] Mandelbrot, B. B. The Fractal Geometry of Nature, Freeman, San Francisco, CA. 1983.

[14] Masulli, F., Baratta, D., Cicione, G., Studer, L. Daily Rainfall Forecasting using an Ensemble Technique based on Singular Spectrum Analysis. *In Proceedings of the International Joint Conference on Neural Networks IJCNN 01*, pp. 263-268, vol. 1, IEEE, Piscataway, NJ, USA, 2001.

[15] Mozer, M. C. Neural Net Architectures for Temporal Sequence Processing. A. S. Weigend and N. A. Gershenfeld, eds. Time Series Predictions: Forecasting the Future and Understanding the Past. pp. 243-264. Reading, M.A.: Addison-Wesley. 1994.

[16] Pucheta, J., Patiño, H., Schugurensky, C., Fullana, R., Kuchen, B. Optimal Control Based-Neurocontroller to Guide the Crop Growth under Perturbations. *Dynamics Of Continuous, Discrete And Impulsive Systems Special Volume Advances in Neural Networks-Theory and Applications. DCDIS A Supplement, Advances in Neural Networks*, Watam Press, Vol. 14(S1), pp. 618—623. 2007.

[17] Pucheta, J., Patiño, H.D. and B. Kuchen, Neural Networks-Based Time Series Prediction Using Long and Short Term Dependence in the Learning Process. In proc. of the 2007 *International Symposium on Forecasting*, 24th to 27th of June 2007 Marriott Marquis Times Square, New York. 2007.

[18] Pucheta, J., Patino, D. and Kuchen, B. “A Statistically Dependent Approach For The Monthly Rainfall Forecast from One Point Observations”. *In IFIP International Federation for Information Processing Volume 294, Computer and Computing Technologies in Agriculture II*, Volume 2, eds. D. Li, Z. Chunjiang, (Boston: Springer), pp. 787–798. (2009).

[19] Velásquez Henao, Juan David, Dyna, Red. Pronóstico de la serie de Mackey glass usando modelos de regresión no-lineal. Universidad Autónoma de Mexico. Campus Aragón. 2004.

[20] Zhang, G.; B.E. Patuwo, and M. Y. Hu. Forecasting with artificial neural networks: The state of art. J. Int. Forecasting, vol. 14, pp. 35-62. 1998.

[21] Patino, H.D.; Pucheta, J.A.; Schugurensky, C.; Fullana, R.; Kuchen, B., “Approximate Optimal Control-Based Neurocontroller with a State Observation System for Seedlings Growth in Greenhouse,” Approximate Dynamic Programming and Reinforcement Learning, 2007. *ADPRL 2007. IEEE International Symposium* on , vol., no., pp.318-323, ISBN: 1-4244-0706-0. 1-5 April 2007.

[22] Rodríguez Rivero, C.; Pucheta, J.; Baumgartner, J.; Patiño y H., Kuchen, B. An Approach for Time Series Forecasting by simulating Stochastic Processes through time lagged feed-forward neural network. *Proceedings of the 2010 International Conference on Data Mining*, Las Vegas, USA. Vol.6, pp. 287-293, ISBN 1-60132-138-4. CSREA Press.