A STATISTICALLY DEPENDENT APPROACH FOR THE MONTHLY RAINFALL FORECAST FROM ONE POINT OBSERVATIONS

J. Pucheta [1,*], D. Patiño [2], B. Kuchen [2]

[1] LIMAC, Departments of Electronic Engineering and Electrical Engineering, National University of Córdoba, Vélez Sarsfield ave. 1611, ARGENTINA X5016GCA.

[2] Institute of Automatics, Faculty of Engineering, National University of San Juan, Lib. San Martín ave., 1109, ARGENTINA J5400ARL.

[*] Corresponding author, Address: LIMAC, Departments of Electronic Engineering and Electrical Engineering, National University of Córdoba, Vélez Sarsfield ave. 1611, ARGENTINA. X5016GCA. Email: julian.pucheta@gmail.com.

Abstract: In this work an adaptive linear filter model in a autoregressive moving average (ARMA) topology for forecasting time series is presented. The time series are composed by observations of the accumulative rainfall every month during several years. The learning rule used to adjust the filter coefficients is mainly based on the gradient-descendent method. In function of the long and short term stochastic dependence of the time series, we propose an on-line heuristic law to set the training process and to modify the filter topology. The input patterns for the predictor filter are the values of the time series after applying a time-delay operator. Hence, the filter's output will tend to approximate the current value available from the data series. The approach is tested over a time series obtained from measures of the monthly accumulative rainfall from La Perla, Córdoba, Argentina. The performance of the presented approach is shown by forecasting the following 18 months from a hypothetical actual time for four time series of 102 data length.

Keywords: Adaptive filter, rainfall forecasting, Autoregressive moving average

1.   INTRODUCTION

This work presents an approach to the future rainfall water availability problem for agricultural purposes. There are several approaches based on non linear autoregressive moving average filters that face the rainfall forecast problem for water availability by taking an ensemble of measurement points (Liu and Lee, 1999; Masulli *et al*., 2001). Here, the proposed approach is based on the classical linear autoregressive filter moving average using time lagged feedforward approach, by considering the historical data from one geographical point. One of the motivations for this study follows the closed-loop control scheme (Pucheta et al., 2007a) where the controller considers future conditions for the control law's design as shown Fig. 1. In that scheme the controller takes into account the actual state of the crop by a state observer and the monthly accumulative rainfall. However, this paper presents only the controller portion concerning with the rainfall forecast. The controller design is inspired on the one presented in (Pucheta et al., 2007a).

The main contribution of this work lies on the tuning process and filter structure, which employs the gradient descendent rule and considers the long and 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 see the performance of the proposed prediction model, sinusoidal and square signals are used. Then, the predictor filter is applied to the monthly accumulative rainfall from La Perla -Córdoba, Argentina- as the time series to forecast the next 18 values given a historical data set.
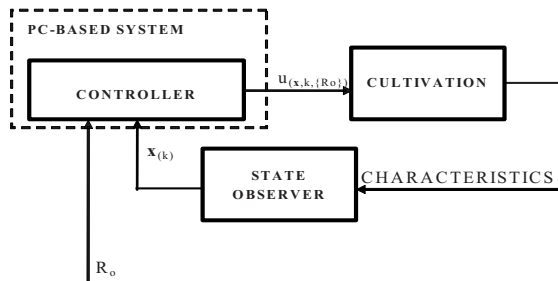

Fig. 1. PC-based control approach, which considers an accumulative rainfall R_o.

1.1   **Overview on Fractional Brownian motion**

In this work the Hurst's value is used in the learning process to modify on-line the number of patterns and number of iterations presented. The H parameter is useful for the definition of the Fractional Brownian Motion (fBm). The fBm is defined in the pioneering work by Mandelbrot (1983), through its stochastic representation

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

(1)

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

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

(2)

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   (1). Here, it is assumed that B is defined on some probability space (Ω, F, P), where Ω, F and P are the sample space, the sigma algebra (event space) and the probability measure, respectively. So, an fBm is a continuous-time Gaussian process depending on the so-called Hurst parameter 0<H<1. It generalizes the ordinary Brownian motion corresponding to H=0.5, and whose derivative is the white noise. The fBm is self-similar in distribution and the variance of the increments is given by

$$Var\left(B_H(t) - B_H(s) \right) = v\left| t - s \right|^{2H}$$

(3)

where, v 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 (Dieker, 2004). In this work, the algorithm uses a wavelet-based method for estimating H from a trace path of the fBm with parameter H (Abry *et al.*, 2003; Dieker, 2004). Three trace path from fBm with different values of H are shown in Fig. 2, where can be noted the difference in the velocity and the amount of its increments.
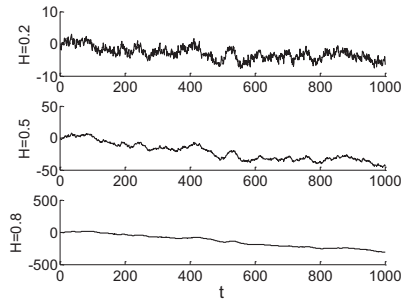
Fig. 2. Three sample path from fractional Brownian motion for three values of H.

2.   PROBLEM STATEMENT

The classical prediction problem may be formulated as follow. Given past values of a process that are uniformly spaced in time, as shown by *x(n-T)*, *x(n-2T), . . . , x(n-mT)*, where T is the sampling period and m is the prediction order, it is desired to predict the present value x(n) of such process. Therefore, obtain the best prediction (in some sense) of the present values from a random (or pseudo-random) time series is desired.

The predictor system may be implemented using an ARMA linear filter. Here, the model follows the classic linear schemme (Ljung, 1999). The linear model structure is self tuned in such a way that smaller the prediction error is (in a statistical sense), the better the filter serves as model of the underlying physical process responsible for generating the data. In this work, time lagged feedforward scheme are used. Thus, the present value of the time series is used as the desired response for the adaptive filter, and the past values of the signal supply as input of the adaptive filter. Then, the adaptive filter output will be the one-step prediction signal. In Fig. 3 the block diagram of the linear prediction scheme based on a ARMA 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 18 values sequence. 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

$$e(k) = x_n(k) - x_e(k),$$

which is used for the learning rule to adjust the filter's coefficients. The coefficients of the filter are adjusted on-line in the learning process, by considering a criterion that modifies 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) 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. A similar algorithm was presented in (Pucheta *et al.*, 2007b).
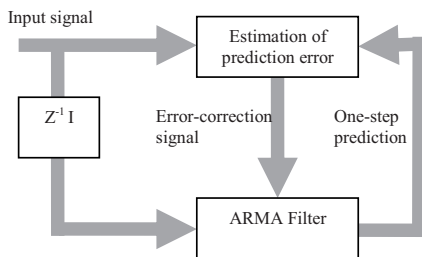

Fig. 3. Block diagram of the linear prediction.

3.   PROPOSED APPROACH FOR PREDICTION

3.1   Autoregressive Linear Model

Now, a linear autoregressive filter model (Haykin, 1999; Ljung, 1991) is proposed. The filter used is a time lagged feedforward type. The filter topology consists of one input with l_x taps, and one output. The rule used in the tuning process is based on the standard descendent gradient (Ljung, 1991). The tuning rule modifies the number of patterns and the number of iterations at each time-stage according to the Hurst's parameter H, which gives short and long term dependence of the sequence $\{x_n\}$ or —from a practical point of view, it gives the ruggedness of the time series. In order to predict the sequence $\{x_e\}$ one-step ahead, the first delay taken off from the tapped-line x_n is used as input. Therefore, the output prediction can be denoted by

$$x_e(n+1) = F_p\left(Z^{-1}I(\{x_n\})\right) \qquad (4)$$

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

3.2   The Proposed Learning Process

The filter's coefficients are tuned by means of the gradient-descendent rule in a batch scheme, which in turn considers the long and short term stochastic dependence of the time series measured by the Hurst's parameter H. The proposed learning process consists on changing both the number of

patterns and the number of iterations in function of the parameter H for each corresponding time series. Here, the tuning process is performed using a batch model. In this case the update of the coefficients is being performed after the presentation of all tuning examples, which forms an epoch. The pairs of the used input-output patterns are

$$(x_i, y_i) \quad i = 1,2,...., \ N_p \tag{5}$$

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 define as

$$X_i = Z^{-1} I(\{x_i\}), \tag{6}$$

and its corresponding output vector as

$$Y_i = x_i, \tag{7}$$

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

$$l_x \leq N_p \leq 4 \cdot l_x,$$

where, l_x is the dimension of the input vector.

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

$$l_x \leq i_t \leq 4l_x.$$

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 is an fBm. The dependence is evaluated by the Hurst's parameter H, which is computed using a wavelet-based method (Abry *et al.*, 2003).

Then, a heuristic adjustment for the pair (i_t, N_p) in function of H according to the membership functions shown in Fig. 4 is proposed. Finally, the number of inputs of the 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 —$\{x_n\}$ and $\{\{x_n\},\{x_e\}\}$, should have the same H parameter. The error between $H(\{x_n\})$ and $H(\{\{x_n\},\{x_e\}\})$ is used for tuning the value of l_x, by mean of a well-known PID scheme. Thus, l_x is updated by,

$$l_x = l_x + round \ (K_P \cdot e_H + K_D \cdot \dot{e}_H). \tag{8}$$

where K_P, K_D are constants, and e_H is defined as

$$e_H = H(\{x_n\}) - H(\{\{x_n\},\{x_e\}\}). \tag{9}$$

Thus, the objective is to forecast the time series in such a way that the predicted time series and the data time series present the same H.
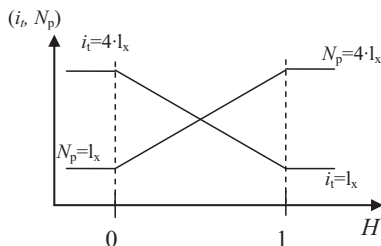
Fig. 4. Heuristic adjustment of (i_t, N_p) in terms of H.

4.   MAIN RESULTS

4.1   Set-up of Filter and Tuning Process

Table 1. Initial condition of the tuning algorithm.

Variable	Initial Condition
l_x	12
i_t	$5 \cdot l_x$
H	0.5
η	0.1

The initial conditions for the filter and tuning algorithm are shown in Table 1. Note that the initial number of iteration is set in function of the input number. The variable η is the step-size of the gradient descendent method, used for tuning the filter parameter. These initial conditions of the tuning algorithm were used for forecasting the monthly accumulative rainfall time series, whose sizes have a length of 102 values each.

4.2   Preliminary Results Using Test Time Series

In order to test the proposed design procedure of the linear predictor, an experiment with sinusoidal and square signals was carried out. The performance of the filter is evaluated using the mean Symmetric Mean Absolute Percent Error (SMAPE):

$$SMAPE_s = \frac{1}{n} \sum_{t=1}^{n} \frac{\left| X_t - F_t \right|}{\left(X_t + F_t + 0.1 \right) / 2} \cdot 100 \qquad (10)$$
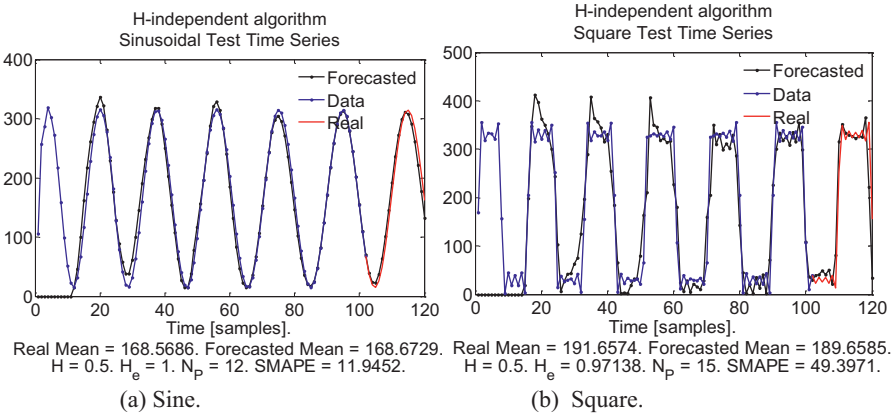
where, t is the time observation, n is the (Data series) test set size, s each time series, X_t and F_t are the actual and the forecast time series values at time t respectively. The shortcoming arise when the denominator is 0 in

$$SMAPE_{s} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|X_{t} - F_{t}\right|}{\left(X_{t} + F_{t} + 0.1\right)/2}\cdot 100 \qquad (10), \text{ so we opt to}$$

include the value 0.1 mm of accumulative rainfall as a dummy minimum

only for evaluating $SMAPE_{s} = \dfrac{1}{n}\sum_{t=1}^{n}\dfrac{\left|X_{t} - F_{t}\right|}{\left(X_{t} + F_{t} + 0.1\right)/2}\cdot 100 \qquad (10)$

to avoid the division by zero. The SMAPE of each series s calculates the symmetric absolute error in percent between the actual X_t and its corresponding forecast F_t value, across all observations t of the test set of size n for each time series s. Fig. 5 (a) shows the filter response, when it forecasts the 18 future values for a sinusoidal time series. The used sine time series has a period T=0.48 s, and it is sampled at T_0=0.05 s. The initial length of the tapped-delay line was set-up at 12 taps and at the end of the tuning process was equal to 12, given that Eq $l_{x} = l_{x} + round\left(K_{P}\cdot e_{H} + K_{D}\cdot \dot{e}_{H}\right)$. (8) does not apply because H equals 1 for all the trials. For a square time series, Fig. 5 (b) presents the forecasted 18 values. Here the value of H, across for the complete time series $\{x_n\}$ and $\{x_e\}$, differs at a 5%. To improve the forecasting performance of the H-dependent filter, it is used as initial condition l_x= 15, in order to increase H of the $\{x_e\}$. The new results are shown in Fig. 5 (c) and Fig. 5 (d), where the percentage is declined in the order of 1%.
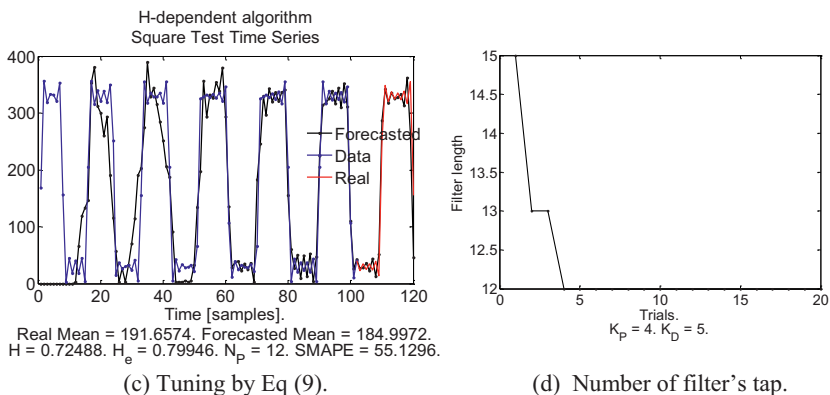

Real Mean = 168.5686. Forecasted Mean = 168.6729. H = 0.5. H_e = 1. N_P = 12. SMAPE = 11.9452.

(a) Sine.

Real Mean = 191.6574. Forecasted Mean = 189.6585. H = 0.5. H_e = 0.97138. N_P = 15. SMAPE = 49.3971.

(b) Square.

(c) Tuning by Eq (9).

(d) Number of filter's tap.

Fig. 5. Algorithm performance on test series.

4.3   **Prediction Results for the Rainfall Time Series**

Each one of the time series are composed by observations from the monthly rainfall, which over ten years are yields 120 values. However, the 18 last values where used to validate the performance of the prediction system. So, 102 values forms the Data set, and the Forecasted set are 120, and the Real data are 18 values. The Data was obtained along 4 decades, which is the laps from January, 1962 to December, 1971, from Jan 1972 to Dec 1981, and so on up to Dec 2001. Obtained results are show in Fig. 6.
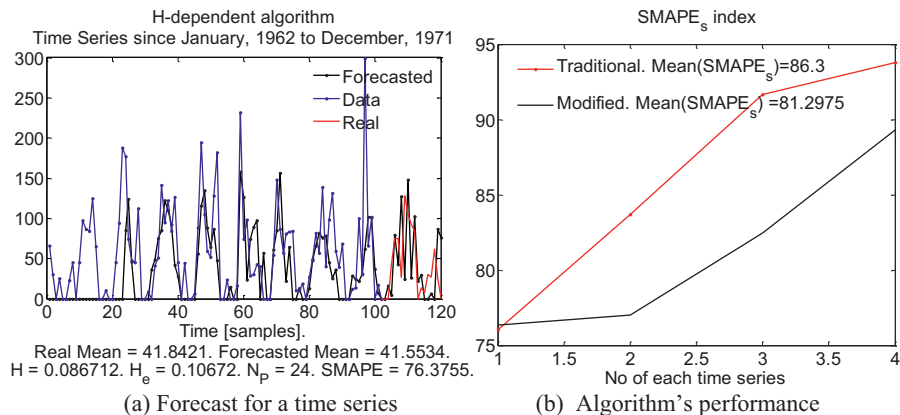

(a) Forecast for a time series

(b) Algorithm's performance

Fig. 6. Performance on the prediction of rainfall time series.

4.4   **Comparative Results**

The performance of the ARMA predictor filter is evaluated by

$$SMAPE_{\,s} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|X_{t}-F_{t}\right|}{\left(X_{t}+F_{t}+0.1\right)/2}\cdot 100$$

Eq.   (10) across the accumulative rainfall time series measured at La Perla (Córdoba, Argentina). Fig. 6 (b) shows by red trace the evolution of the SMAPE index for an H-independent filter, which uses a tuning algorithm with fixed parameters. The other filter, labeled H-dependent filter is proposed in this work and use the H parameter to adjust heuristically either the filter's structure or parameters of the learning rule -black trace in Fig. 6 (b). It can be noted the improvement since the SMAPE index diminish from 86.3 to 81.29, averaging over the four time series.

5.   DISCUSSION

The evaluation of the obtained results has been realized by comparing the performance of the proposed filter against that of the classic filter, both based on linear scheme. Although the difference between both filters only resides in the adjustment algorithm, the coefficients that each filter has perform 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, although such parameters and filter's structure are changed by the proposed algorithm but they are not modified by the classic algorithm. In the algorithm of adjustment for the proposed filter, the coefficients and the structure of the filter are tuned by considering their stochastic dependency. It can be noted that in figures Fig. 5 and Fig. 6 the computed value of the Hurst's parameter is denoted either H_e or H when it is obtained from the Forecasted time series or from the Data series, respectively, since the Real (future time series) are unknown. Index SMAPE is computed between Data series and the Forecasted   one,   as   indicates   the   Ec.

$$SMAPE_{\,s} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|X_{t}-F_{t}\right|}{\left(X_{t}+F_{t}+0.1\right)/2}\cdot 100$$

(10).

6.   CONCLUSION

In this work a statistically dependent linear filter for forecasting time series has been presented. The tuning rule proposed to adjust the filter

coefficient is based on the standard gradient-descendent method. Furthermore, in function of the long and 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 filer topology at each time-stage, which is the number of input taps, the number of patterns and the number of iterations. The main result shows a good performance of the predictor system applied to the accumulative rainfall time series when the observations are taken from a single point, due to similar roughness for both the original and the forecasted time series, evaluated by H and He respectively was obtained. These results encourage one to go on working with this new tuning algorithm, applying to other filter models (such as non linear autoregressive moving average), due to the time series generated by humans interaction presents short and long term stochastic dependence.

ACKNOWLEDGMENTS

REFERENCES

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, 2003, pp. 527-556.

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

Haykin, S, Neural Networks: A comprehensive Foundation, 2nd Edition, Prentice Hall, 1999.

Liu, J.N.K.; Lee, R.S.T., Rainfall forecasting from multiple point sources using neural networks, Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on, Vol.3, Iss., 1999, Pages:429-434 vol.3.

Ljung, Lennart. System Identification Theory for the user. 2nd Ed, Prentice Hall. 1999.

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

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.

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) 2007a, pp. 618—623.

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, 2007b.