Forecasting SO_2 air pollution in Salamanca, Mexico using an ADALINE.

M.G. Cortina[a], U.S. Mendoza[a], J.M. Barrón-Adame[a], D.Andina[b], A. Vega-Corona[a]

[a,] *Universidad de Guanajuato, Facultad de Ingeniería Mecánica, Eléctrica y Electrónica, México.*
[b] *Universidad Politécnica de Madrid, ETSI Telecomunicación, España.*

Abstract

A comparison between a linear regression model and a Non-linear regression model is presented in this work for forecasting of pollution levels due to SO_2 in Salamanca city, Gto. Prediction is performed by means of an Adaptive Linear Neural Network (ADALINE) and a Generalized Regression Neural Network (GRNN). Prediction experiments are realized for 1, 12 and 24 hours in advance, and the results for linear regression have been satisfactory. The performance estimation of both models are determined using the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Obtained results are compared. The final results indicated that ADALINE outperforms the past approach using GRNN.

Keywords: ADALINE, GRNN, SO_2 concentration

1. Introduction

Salamanca city is catalogued as one of the most polluted cities in Mexico. The main causes of pollution in Salamanca are due to fixed emission sources such as Chemical Industry, and Electricity Generation, being the more important pollutants in Air, Sulphur Dioxide (SO_2), measured in part per billion (PPB), and Particulate Matter less than 10 Micrometers in diameter (PM_{10}), measured in micrometers in diameter. This article focuses on forecasting SO_2 concentration.

In an effort to fight pollution of the zone, in July 2005, the Program of Environmental Contingency was launched, the purpose of it being to protect the health of population, especially that of vulnerable groups. This program contemplates the urgent and immediate reduction of SO_2 emissions and PM_{10} when measurements of these pollutants register levels above those established by Health Authorities. To accomplish it, 3 phases were established: Pre-contingency, Contingency Phase I and Contingency Phase II for Sulphur Dioxide, PM_{10} particles and for a combination of both [1].

Prediction of pollutant concentrations in the Atmosphere would allow taking preventive measures, reducing the emission of pollutants before reaching levels of an environmental contingency.

In this work, the use of a Neural Network ADALINE (ADA) is proposed to predict pollution levels 1, 12 and 24 hours in advance for the zone of Salamanca before an environmental contingency

occurs, and results obtained are compared with those obtained with a Generalized Regression Neural Network (GRNN) [2].

2. Methodology

Figure 1 shows the flow diagram of the methodology that was followed for the realization of this work, which consists of 3 main phases: i) Select Training and Test data sets, ii) Neural Network Design and iii)Simulation and Results Evaluation.

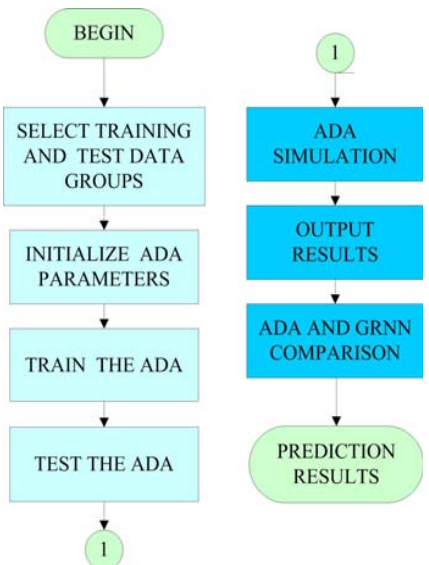

Fig. 1. Flow Diagram for SO_2

2.1. Training and Test Data Sets

Data base used for this experiment have been previously processed as did in [2]. Data base selected to train the net are those corresponding to the months of January and February 2005, and data selected to make the predictions are those of March 2005.

2.2. Neural Network Design

ADA is a generalisation of the perceptron training algorith. The main functional difference with the perceptron training rules is the way the output of the system is used in the learning rule.

ADA transfer function is a linear function instead of a hard limit transfer function of the Perceptron. ADA and the multiple version Madaline (MADA) use a learning mechanism known as Delta Rule of Widrow and Hoff, also known as the Least Square Mean Error (LSM) Rule [3], based on the search of the minimum error between the desired output and the linear output obtained.

2.2.1 Network Structure

In general terms, the output function of the network is given by equation (1)

$$a = W^T p \qquad (1)$$

where **a,** is the output vector of the linear neurons, **W**, is the weight matrix, and **p** is the input vector.

2.2.2 Learning Rule

ADA is a supervised learning network that needs a priori knowledge of the associated values to each input, denominated Widrow-Hoff Rule, also known as LMS .

The Rule of Widrow-Hoff in general terms is expressed as indicated in equation (2).

$$\mathbf{W}(k+1) = \mathbf{W}(k) + 2\alpha \mathbf{e}(k)\mathrm{p}^T(k) \qquad (2)$$

where k represents the current iteration of the weights updating process, **W**(k+1) is the next value that vector **W** is going to take, and **W**(k) is the current weights vector; **e** is the vector of current error, defined as the difference between the desired response and the network's output shown in equation (1); α is the learning rate. The gain updating process is given by equation (3)

$$\mathbf{b}(k+1) = \mathbf{b}(k) + 2\,\alpha\,\mathbf{e}(k) \qquad (3)$$

where **e** is the vector representing the error, **b**(k+1) is the Gain updating vector, and **b**(k) is the current Gain Vector.

2.3. Training and Simulation of the net

Input neurons are equal to the number of observations. In this work, input neurons are 1344 which is the total number of observations corresponding to the training period (January and February 2005). Test group consists of 672 observations corresponding to March 2005.

Training vector have been made only to predict SO_2 concentrations levels, that means, no other variables were used to perform the forecast. Vector ***X,*** is the input vector at times *t=i-n-1*, ..., *t=i-1*, *t=i*, where *i,* is the current hour and *n* is the number of forecast hours is done. Vector ***Y,*** is the output vector, whose elements correspond to the estimation of SO_2 levels at times *t=i+1*, *t=i+2*, ..., *t=n*, where *i,* is the current hour and *n* is the number of forecast hours is done.

Training and simulation for ADA and GRNN were performed using two different Pattern Schemes, since the scheme used in [2] produces an apparent time-shift in the prediction made by ADA, due to how the patterns for the training and test matrices were formed. Due to this situation, two different Training Schemes were used, the second one to correct the apparent time-shift in the forecast for ADALINE network. These schemes were named ADA I and ADA II.

In Training Scheme I (ADA I), input patterns $\mathbf{x}_i$ are formed as indicated in [2], where in time *t=i,* each pattern is $\mathbf{X} = \{x_{i-n-1}, \ldots x_{i-1}, x_i\}$, where x_i is the SO_2 concentration in the current hour, and *n* is the number of forecast hours is done. This means that input patterns are formed with the current and past concentrations. However, the first pattern was formed by all zeros, since for the first data, we had not apriori information. Output Training patterns are formed with the next *n* hours concentrations, $\mathbf{Y} = \{y_{i+1}, y_{i+2}, \ldots, y_{i+n}\}$.

In Training Scheme II (ADA II), patterns were formed as in ADA I, but with the difference that for ADA II $\mathbf{x_1}$ is equal to ADA I $\mathbf{x_2}$, ADA II $\mathbf{x_2}$ is equal to ADA I $\mathbf{x_3}$ and so on. Another difference is that ADA I was formed with N patterns, and ADA II with N-1 patterns. Due to the structure of the patterns it is necessary that we use N-1 patterns.

Evaluation of the forecasting Performance was accomplished using the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE).

Mean-squared error is the most commonly used measure of success of numeric prediction, and root mean-squared error is the square root of mean-squared-error, taken to give it the same dimensions as the predicted values by themselves. This method exaggerates the prediction error - the difference between prediction value and actual value of a test case – of test case in which the prediction error is largest than the others. If this number is significantly greater than the mean absolute error, it means that there are test cases in which the prediction error is significantly greater than the average prediction error. Balaguer et al. [4], have used RMSE as an indicator of the relationship between predicted and observed data.

Root Mean Squared Error is computed according to equation (4)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \qquad (4)$$

where $\hat{y}_i$, is the predicted value for a determined time *t=i*, y_i, is the real value for the same time and *n* is the number of observations.

Mean Absolute Error is the average of the difference in all test cases. Mean Absolute Error (MAE) is computed according to equation (5)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \qquad (5)$$

where $\hat{y}_i$, is the predicted value for a given time *t=i*, y_i, is the real value for the same time, and *n* is the number of observations.

3. Results

Table I shows the obtained results for schemes ADAI, ADA II, and GRNN, proposed in [2]. GRNN was trained with Scheme I pattern.

Table 1
Performance of a GRNN against an ADA

Scheme	Hours Ahead			
	(PPB)	1	12	24
GRNN	RMSE	59,6	58,05	73,94
	MAE	34,67	43,72	53,1
ADA I	RMSE	19,92	38,84	140,00
	MAE	16,67	81,53	98,76
ADA II	RMSE	19,92	58,36	58,36
	MAE	16,67	36,13	36,13

The best results were achieved for the prediction of 1-hour ahead SO_2 concentrations in both GRNN and ADA networks, which agrees with results obtained by Mendoza [2] and Turias [5]. There is a significant improvement using the ADA II network since both MAE and RMSE errors are much lower than those obtained with GRNN. Results of 1-hr prediction are shown in figures 2 and 3.

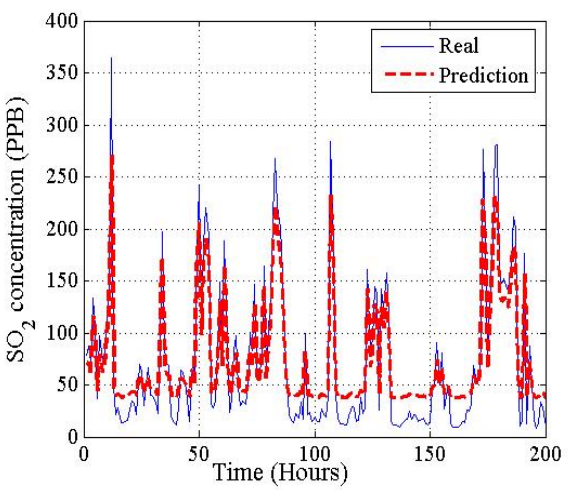

Fig 2. 1-hour forecasting with a GRNN

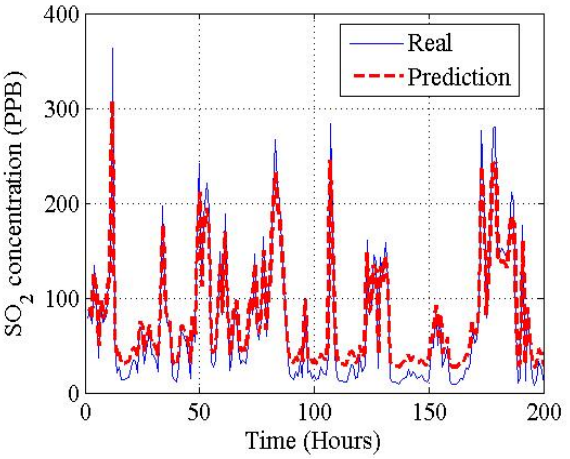

Fig 3. 1-hour forecasting with an ADA.

In figure 5, for the case of SO_2 levels prediction with 12 hours ahead with ADA I, the prediction apparently presents a time shift, which prevents getting satisfactory results. This is due to the patterns organization in this scheme.

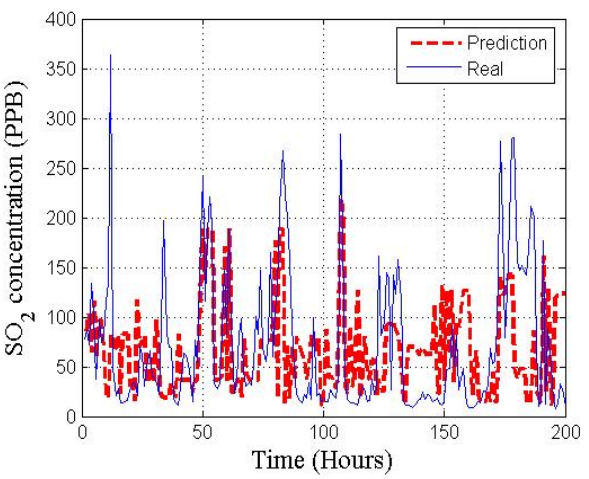

Fig 4. 12-hours forecasting with a GRNN.

Results obtained with ADA II were much better than those obtained with ADA I and GRNN, comparing them in Table 1, MAE and also RMSE error were reduced, and ADA II showed no time-shift for the prediction of SO_2. Figures 4, 5 and 6 show results for 12-hour ahead prediction for the different networks

that were used.

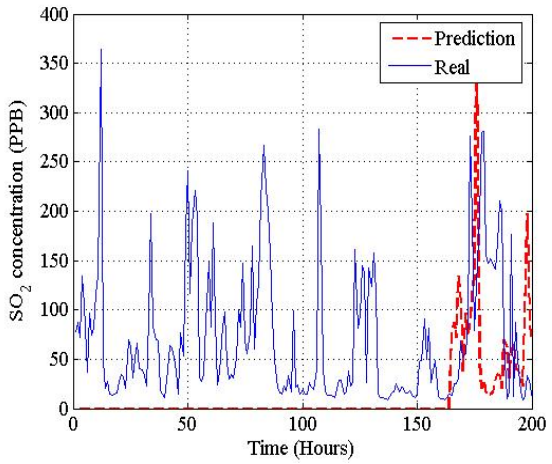

Fig 5. 12-hours forecasting with an ADA I.

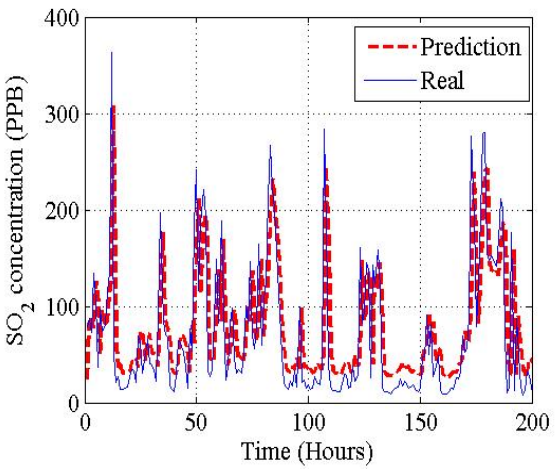

Fig 6. 12-hours forecasting with anADA II.

For the case of 24-hr prediction, again, ADA II scheme showed a better performance over the GRNN, and ADAI. Figures 7 and 9 show results for 24-hr prediction using GRNN and ADA II.

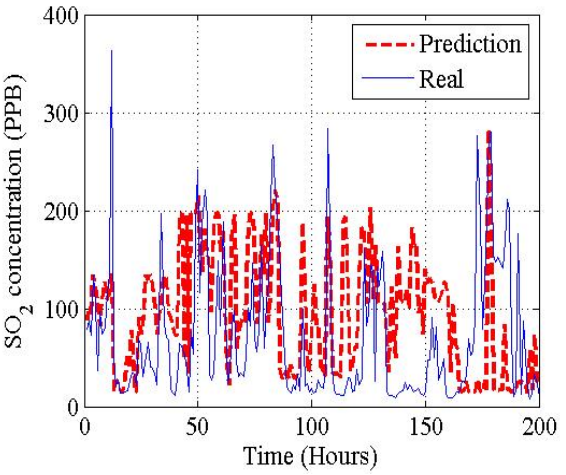

Fig. 7. 24-hours forecasting with a GRNN.

The results for ADA I are shown in the figure 8, where it is also time-shifted as results for 12-hr forecasting.

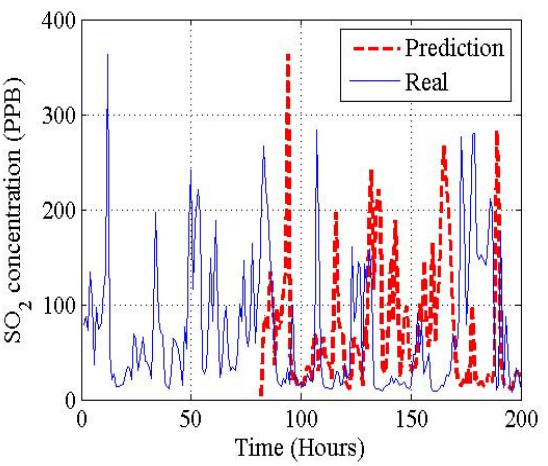

Fig. 8. 24-hours forecasting with ADA I.

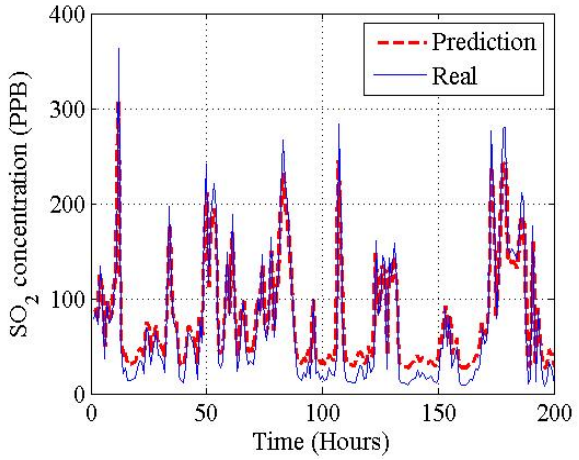

Fig. 9. 24-hours forecasting with ADA II.

Obtained results in SO_2 forecast concentration levels with ADA show that the scheme of patterns plays an important role for obtaining acceptable results.

4. Conclusions

This work shows the comparison of the performance of a Linear Regression Neural Network (ADALINE) and a Non-Linear Regression Network (GRNN) to forecast concentration levels of SO_2. One of the main differences is, that a linear regression network needs less parameters adjustment than a Non-linear regression network, thus facilitating its implementation, however, to obtain better results with a linear regression network, it is necessary to search for pattern scheme that allows us reduce the error in the SO_2 prediction of concentration levels. ADA II outperformed GRNN in all the cases, showing that an appropriate pattern Scheme must be used.

In both cases, error increases as the number of forecast hours is made increases. It has been shown that the use of a linear regression neural network improves the SO_2 prediction of concentration levels, reducing the error obtained with a Non-linear regression neural network.

References

[1] Instituto Estatal de Ecología. Calle Aldana N.12 esquina calle Republica, Pueblito de Rocha, c.p. 36040, Guanajuato, Gto.

[2] U.S. Mendoza-Camarena, F. Ambriz-Colin, D.M. Arteaga-Jauregui and A.Vega-Corona. SO_2 concentrations forecasting for different hours in advance for the city of Salamanca, Gto., Mexico. 2005.

[3] Madan M.. Gupta, Liang Jin, and Noriyasu Homma. Static and Dynamic Neural Networks, 2003.

[4] Emili Balguer Ballester, Emilio Soria Olivas, and Jose Luis Carrasco Rodríguez. Forecasting of surface ozone concentratioons 24 hours in advance using neural networks. *International conference o Neural Networks and Applications*, 2001.

[5] I. Turias, F.J. Gonzalez, and P.L. Galindo. Application of neural techniques to the modelling of time-series of atmospheric pollution data in the 'campo de Gibraltar' region. Neural Network Engineering Experiences, Procc. Of the 8 International conference on application of neural network: 9-16, 2003.