Radial Basis Function and Elman Networks for Pollutant's Parameter Prediction in the Region of Annaba Algeria

Mohamed Tarek Khadir
University of Badji Mokhtar, Dep. of Computer Science
Laboratoire de Gestion Electronique de Documents (LabGED)
Annaba, Algeria
khadir@labged.net

Sabri Ghazi
University of Badji Mokhtar, Dep. of Computer Science
Laboratoire de Gestion Electronique de Documents (LabGED)
Annaba, Algeria
ghazi@labged.net

Abstract— **This paper describes the development of air pollutants concentration prediction models of five different pollutants (O3, PM10, SO2, NOx, COx), using Radial Basis Function, and Elman Networks, two neurocomputing paradigms. Each Artificial Neural network (ANN) predicts, therefore the concentration of the five different pollutants. These models are developed in order to give 12 hours ahead prediction for the region of Annaba, northeast of Algeria (north of Africa). Receiving the measurement of air pollutant concentration and the metrological parameters (wind speed, temperature and humidity) at time t; the models are designed to predict air pollutant concentration at t+12 hours. Once predicted pollutant concentrations are obtained, and the validity of each ANN model is proven, the performances of both ANN models are comprehensively compared and assessed. Conclusions are finally drawn and the use of a particular ANN network over another is justified on the light of the obtained results.**

Keywords- ***Pollutants concentration prediction, Artificial Neural Network, Elman Network, Radial Basis Function, neurocomputing.***

I. INTRODUCTION

All air pollution generated by rapid urbanization, population growth and industrialization has taken alarming dimensions, and is one the greatest evil that mankind has to face in the coming years. To prevent the further decline of air pollution, scientific planning of analysis methods and pollution control are required. Within this framework it is necessary to implement air quality forecasting tools in order to take needed measures, such as reducing the effect of a predicted pollution peak on the surrounding population and eco-system. Many factors influence air pollutants concentrations. Among the most important are: metrological conditions, topology and population density. This makes air pollution difficult to model. Many air pollution prediction models have been studied [1] such as, mathematical emission models, linear models; Artificial Neural Networks-based models and hybrid models, in order to design air quality prediction systems, moderates air pollution and limit the influence of peaks periods by informing community by taking the necessary precaution. Air pollutant can be chemical such as: (SO_2: dioxide of sulfur, O_3: ozone, NO_x: oxide of nitrogen, CO_x: oxide of Carbon), or solid such as PM10 (Particulate Matter with an aerodynamic diameter of 10 micrometer). PM10 are the sum of all solid and liquid particles suspended in air, many of which are hazardous. This complex mixture includes both organic and inorganic particles, such as dust, pollen, soot, smoke, and liquid droplets. These particles vary greatly in size, composition, and origin. Particles in air are either: directly emitted, for instance when fuel is burnt and when dust is carried by wind, or indirectly formed, by a photochemical interaction between gaseous pollutants in the air space.

The paper is organized as follows. Section II introduces the specifications of the studied area and the nature and amount of used data. Section III explains the usage of Artificial Neural Networks as pollutants prediction tool. Section IV details the design of the two ANN approaches, i.e., RBF and Elman. In the light of the available data, chosen networks topologies are justified. Section V presents all obtained results and establishes a comparison study. Finally, Section VI concludes the research paper.

II. STUDIED AREA AND AVAILLABLE DATA

A. Studied Area

In Algeria, respiratory infections remain the leading cause of child mortality after measles and diarrhea [2]. Chronic bronchitis, lung cancer and asthma, among other diseases are also caused by pollution. This is especially apparent in some unventilated and highly industrialised regions; Annaba may be counted among these. In the region, the prevalence of asthma is higher than the national rate, 55% of asthmatics have more than one crisis per month and 42% of patients were hospitalized at least once during the year. Annaba region is located in the Eastern part of Algerian coast (600 km of Algiers); its town is constituted

Fig. 1 Geographic map of Annaba, northeast of Algeria (Google Earth ©).

Fig.2: Topology of Annaba (Google Earth ©)

of a vast plain bordered in the South and West, of a mountainous massive in North, and by the Mediterranean Sea in the East. Its bowl-shaped topography favors air stagnation and the formation of temperature inversions. These situations allow the pollutants accumulation and the increase in concentration that result. Sea and land breezes contribute to slope transport of clouds of pollutants. Indeed, the clouds of pollutants are carried away by land breeze to the sea. These clouds of pollutants return to the city as a result of sea breezes along the Seraidi Mountain. The clouds look over the city in a form of circle. The pollutants are deposited slowly by gravity and therefore there is pollution affecting the three receivers (sea, land and air) [2]. Contaminants Air emissions are distributed differently depending on industry. The industry is the engine of growth and environmental degradation in Annaba and its surrounding area, where the most industrial sites are located nearby, such as the complex of phosphate and nitrogen fertilizers, Asmidal and the Metalsteel complex of El-Hadjar. These industrial activities are the main source of particulate matter and sulfur oxides, while emissions of carbon monoxide, nitrogen and lead are mainly due to the transport sector. Emissions from road traffic are the main pollutants of the atmosphere. Air pollution has increased with the increasing number of vehicles (an annual increase of 5% in Algeria) and the absence of emissions control. Meanwhile, the treatment of waste (domestic, industrial, hospital, toxic), which is to put them in wild sites, is also a source of air pollution to such an extent, that they are incinerated in free air.

B. *Data Pre-processing*

The dataset used in this work covers the period 2003-2004 and was provided by SAMASAFIA network center [3] on a continuous basis of 24 hours. Air pollutants monitored continuously include the concentrations of: nitric oxide (NO), carbon monoxide (CO), ozone (O3), particulate matter (PM10), nitrogen oxides (NOx), nitrogen dioxide (NO2) and sulfur dioxide (SO2). The used dataset includes also three meteorological parameters: Wind Speed (WS), Temperature (T) and relative humidity (H).

The available data is an hourly measurements of pollutants concentration of two years, form 1/1/2003 to 31/12/2003 and form 1/1/2004 to 31/12/2004, in Annaba. The 2003 dataset was used for training and the one of 2004 for validation; this will help us to efficiently asses and then get the performance of the model. The pollutant concentration measurements are in microgram/ m3. To adapt the data with our model, we have applied normalization using equation (1). Negative values, resulting from faulty measurements where replaced using (2), this will make the input ranges between [0 1], and in the output we will obtain a percentage indicating the degree of the peak alarm.

$$Vp = \frac{Vp}{(max(Vp) - min(Vp))} \tag{1}$$

where V_p is a vector of parameter, *min* and *max* are function that returns the maximum and minimum of the vector.

$$T = T + (T_{max} - T_{\min}) \tag{2}$$

where T is vector of temperatures, *min* and *max* being the maximum and the minimum values of temperature.

III. ARTIFICIAL NEURAL NETWORKS FOR AIR POLLUTANT'S CONCENTRATION

Artificial Neural Networks (ANNs) are an alternative to classical statistics methods used for prediction in air quality monitoring stations [1]. Large amount of available data from monitoring stations can be very useful in order to design efficient ANN predictions models. Relationship between meteorological parameter and air pollutant concentration is very difficult to model. Mlakar and Boznar [4] present a Multi Layered Perceptron (MLP) based model for SO_2 concentration prediction, it receives the meteorological and emission parameters as inputs, the model shows an efficient predictions and outperform the ARMA (Autoregressive Moving Average) based model. Mlakar and Joseph [5] present a pattern recognition inspired approach to select the parameter for the air pollution prediction model, in order to optimize the training time and increases model performance.

To get the most adaptable model for air pollution prediction Jorquera *et al.* [6] present a comparison between three models: ANN based model, linear model and Fuzzy logic based model. These models have been applied to predict Ozone concentration in Santiago Chile, the Fuzzy and ANN model has shown an efficient prediction and outperform the linear model. Gardner and Droling [7] present NO_2 prediction model based on MLP topology, compared with linear regressor using the same input data and parameter, the MLP based model shows better accuracy. In Gardner and Droling [8], the long-term tendency of Ozone concentration in London has been studied using ANN based model. The prediction of PM2.5 (Particulate matter with diagonal smaller the 2.5 micrometer) concentration has been well studied in Perez *et al.* [9], using ANN based model which is able to predict short-term concentration of PM2.5. Perez and Reyes [10] focused on the prediction of PM10.

This is to cite but a few, other research work may be found in the literature such as: Foxall *and al.* [11] for ozone, Corani [12] for Ozone and PM10, Bianchini *et al.* [13] for NO_2, etc.

IV. Model Development

A. Radial Basis Function (RBF)

Radial Basis Function networks has been proposed and used in many studies [14][15], a detailed presentation can be found in Chen *et al.* [16]. As illustrated in Figure 3, RBF networks consist in three layers. The first layer is composed by *n* input neurons connected to each input. The input neurons (or processing before the input layer) standardizes the range of the values by subtracting the median and dividing by the interquartile range. The input neurons then feed the values to each of the neurons in the hidden layer. The neurons of the hidden layer have the radial basis function (3) as activation function,

$$f(x) = e^{(-x-M)^2/2\delta^2} \tag{3}$$

where: M and δ are two parameters respectively the input variable mean and standard deviation.

The outputs of the nodes are combined linearly to give the final network output.

The main advantages of radial basis network may be: The time taken in designing a radial basis network is often less when compared to the training a sigmoid/linear networks; and the number of neurons required for designing the network is considerably less when compared to standard back propagation network [17].

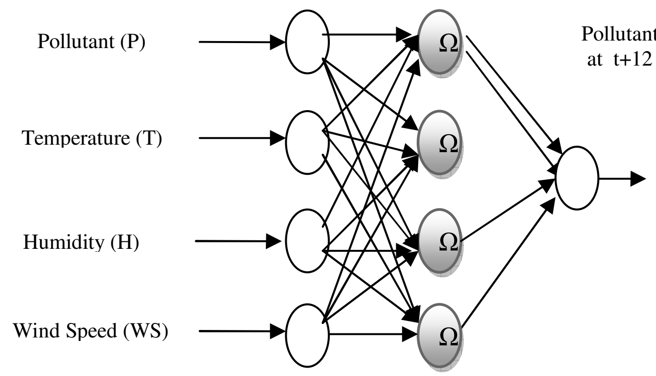

Figure 3: RBF network architecture.

The training algorithm, used iteratively, creates a blank RBF network (no neurons at first), and sequentially adding nodes by including one neuron at a time. Neurons are added to the network until the Sum of Squared Error (SSE) reaches a set target or the maximum numbers of neurons are reached. At each iteration the input vector, resulting in lowering the network error, is used to create a radial basis neuron. The distance between the connecting weights determines the output of hidden neurons and input vector, which is further, multiplied by bias, an additional scalar quantity being added between neuron and fictitious neuron. The output is propagated in a feed forward direction to output layer neurons.

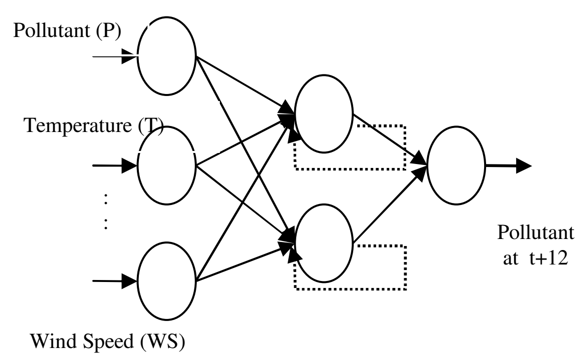

Figure 4: Elman's network architecture.

These, will give a response if the connection weights are close to input signal. This output is, then compared to a target vector. If the error reaches the error goal (a set value given the desired difference between the actual and modelled output), then training is terminated, otherwise the

 ISBN: 978-1-61208-101-4

next neuron will be added. The complete training procedure is implemented in the function *newrb* available in the neural network toolbox supported by MATLAB.

B. Elman networks

Elman networks [18] are considered as recurrent neural networks, because they have feedback connections in the neurons of the hidden layer as illustrated in Figure 4. Elman networks are called buffered networks, as they take in consideration the previous output value and use it to calculate the next step output. The network leaves a trace of its behavior and keeps a memory of its previous states which may enhance the accuracy of the predictions.

V. RESULTS AND DISCUSSION

A. Measurement indeces

The performance of both modeling ANN approaches are evaluated in terms of Index of Agreement (IA) and the Mean Squared Error (MSE). The formulation of IA and MSE are given in equation (4) and (5) respectively.

$$IA = 1 - \frac{\overline{(Cp - C0)^2}}{\overline{\left|Cp - \overline{C0}\right|^2 + \left|C0 - \overline{C0}\right|^2}} \quad (4)$$

$$MSE = \sum_t^n \frac{(C_{pt} - C_{0t})^2}{n} \quad (5)$$

where C_p is the predicted value and C_0 is the measured value of pollutant concentration.

B. Obtained results

Table 1 presents the performance and the architecture of the RBF based models, the topology of the final selected networks as well as the performance of the models in terms of Index of Agreement (IA) and the Mean Squared Error (MSE) as performance measures.

TABLE 1. PERFORMANCE OF THE RBF MODELS

Pollutant	*Topology*	*IA*		*MSE*	
		Training	*Validation*	*Training*	*Validation*
PM10	10-320-1	0.9999	0.9987	0.0004	0.5508
O_3	10-180-1	0.9999	0.6922	0.0229	1.2571
NO_X	10- 105-1	0.9966	0.4866	0.1358	2.7515
SO_2	10-90-1	0.9976	0.5930	0.0995	2.9806
CO_X	10-45-1	0.9900	0.9716	0.1090	1.0513

It can be seen that RBFs show good performances for PM10 and Cox. For O_3, it gives medium accuracy, but for NO_x, it has the worst performance.

The training mechanism for Elman networks is similar to the one used for MLPs and is based on the back propagation algorithm. The overtraining phenomena affect EN in the same manner than it affect MLPs [19]. Table 2 shows the obtained results for all five pollutants.

TABLE 2: THE PERFORMANCE OF THE TESTED ELMAN NETWORKS

Pollutant	*Topology*	*IA*		*MSE*	
		Training	*Validation*	*Training*	*Validation*
SO_2	10-14-1	0.93362	0.92574	0.012658	1.9379
PM10	10-11-1	0.9947	0.89638	0.75991	4.3207
CO_2	10-10-1	0.99861	0.88857	0.039862	1.7877
NO_2	10-11-1	0.9927	0.94002	.88212	2.9005
O_3	10-13-1	0.90107	0.74896	0.37845	0.27922

Figures 5 shows the model performances over a range of 12h for both ANN paradigms (RBF and Elman's Networks) for PM10 predictions. Similarly, Figure 6 shows the results for O_3, Figure 7 for NO_X, Figure 8 for CO_X and Figure 9 for SO_2.

According to these figures and the summarizing Tables 1 and 2, RBF models seem the best adapted to predict efficiently pollutant concentration, regardless of their easier training and validation procedure. Elman models shows good generalization and predicts well pollutants concentration at the exception of SO_2, where this type of models shows the worst performances due to the existence of several peaks of this pollutant within the used data.

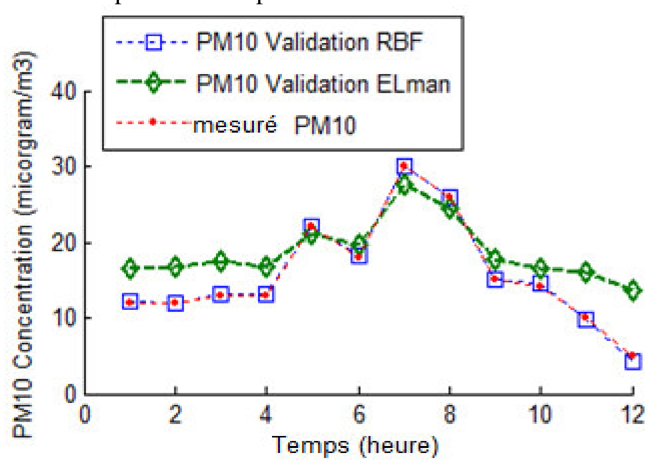

Figure 5 : Predicted and measured values for PM10

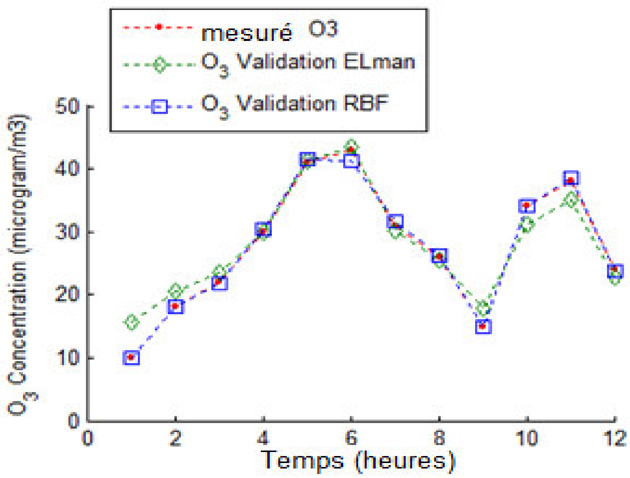

Figure 6 : Predicted and measured values for O_3

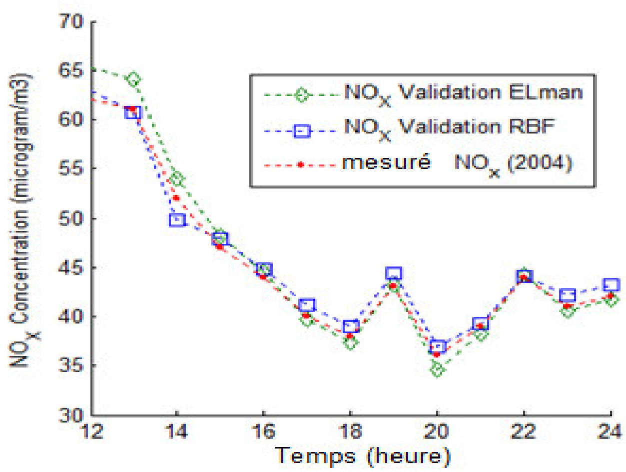

Figure 7 : Predicted and measured values for NO_x

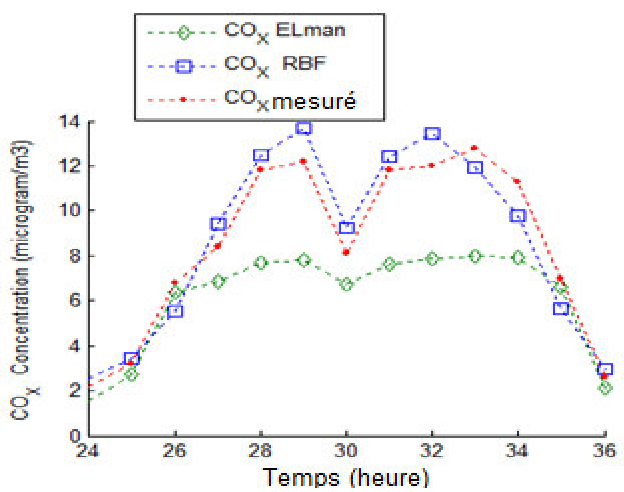

Figure 8 : Predicted and measured values for CO_x

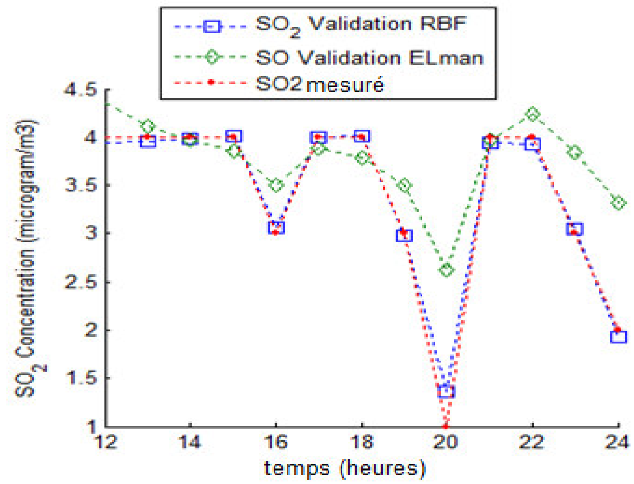

Figure 9 : Predicted and measured values for SO_2

VI. CONCLUSION

Two types of ANN models have been designed, validated and tested for the prediction of five air pollutants. RBF models have shown the best performances followed closely by Elman Network for single step prediction. However these types of models required longer training time and were computationaly effort consumming, even if no much design effort were taken for the choice of the topology number of neurons and feed back loops as in Elman networks..

Elman's networks overestimate or under estimate some pollutants peaks. This is seen for PM10 and CO_x. The addition of emission parameters related to cars and industries may improve prediction quality.

Finally, these results are specific to the region of Annaba. The use these models for others region must comply with data specific proprieties and training must be performed in order to identify the vest network topology.

REFERENCES

[1] Mlakar P. and Boznar M. (2007), Air Pollution Modeling and Its Application XIV : Artificial Neural Network-Based Environmental Models, Springer, USA.

[2] Mebirouk H. And Mebirouk-Bendir F. (2007). Principaux acteurs de la pollution dans l'agglomération d'Annaba. Effets et développements, Colloque International sur l'Eau et l'Environnement, Alger.

[3] www.samasafia.dz\journaux (access 10/10/2009).

[4] Mlakar P. and Boznar M. (1994). Short-term air pollution prediction on the basis of artificial neural network. 2nd International Conference on Air pollution, Barcelona, Spain, pages 545-552.

[5] Mlakar P. and Josef S. (1997) Determination of Features for Air Pollution Forecasting Models, Proceedings of 3rd ed. International Conference on Intelligent Information Systems (IIS '97), Grand Bahama Island, Bahamas, pages 345-356.

[6] Jorquera H., Pérez R., Cipriano, A., Espejo, A., Letelier M.V., and Acuna, G. (1998). Forecasting ozone daily maximum levels at Santiago, Chile, Atmospheric Environment, Vol 32(20), pages 3415-3424.

[7] Gardner, M.W. and Dorling S.R. (1999). Neural network modeling and prediction of hourly NOx and NO2 concentration in urban in London, Atmospheric Environment, vol. 33, pages 709-719.

[8] Gardner M. and Dorling S. (1999). Meteorologically adjusted trends in UK daily maximum ozone concentrations, Atmospheric Environment, Vol. 34(2), pages 171-176.

[9] Pérez P., Trier, A., and Reyes, J. (2000). Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago, Chile. Atmospheric Environment, Vol. 34, pages 1189-1196.

[10] Perez P. and Reyes J. (2002). Prediction of maximum of 24-h average of PM10 concentrations 30 h in advance in Santiago, Chile, Atmospheric Environment Vol. 36, pages 4555-4561.

[11] Foxall R., Krcmar I., Cawley G., Dorling, S., and Mandic, D. (2001). Nonlinear modeling of air pollution times series, IEEE International Conference On Acoustics, Speech, And Signal Processing, Salt Lake City, Utah, USA, vol.6 Pages 3505– 3508.

[12] Corani G. (2005). Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning, Ecological Modeling, Vol. 185, Issues 2-4, Pages 513-529.

[13] Bianchini M., Di Iorio F., Maggini M., Mocenni C., and Pucci A. (2006). A Cyclostationary Neural Network Model for the Prediction of the NO2 Concentration, ESANN'2006 proceedings – European Symposium on Artificial Neural Networks, Bruges, Belgium,.

[14] Chen, S., Cowan, C.F., and Grant, P. M. (1991) Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks, IEEE Transactions on Neural Networks, Vol. 2(2), pages 302-309.

[15] Govindarajan L. and Sabarathinam P.L. (2006). Prediction of Vapor liquidEquilibrium Databy Using Radial Basis Neural Networks, Chem. Biochem. Eng. Q. Vol. 20 (3), pages 319–323.

[16] Cybenko, G. (1989). Approximation by superposition of sigmoidal function, Math. Cont. Sig. syst., vol. 2, pages 303-314.

[17] Costa, M., Pasero, E., Piglione F., and Radasanu D. (1999). Short term load forecasting using a synchronously operated recurrent neural network, Proceedings of the International Joint Conference on Neural Networks, vol.5, pages 3478-3482, Washington, DC, USA.

[18] Elman J. (1990). Finding structure in time. Cognitive Science vol 14, pages 179-211.

[19] Sjoberg, J. and Ljung, L. (1992). Overtraining, regularisation, and searching forminimum in neural networks. In 10th IFAC Symposium on System Identification, Copenhagen, Vol 2, pages 49-72.

 ISBN: 978-1-61208-101-4