Daily Urban NOx Peak Forecasting Using Recurrent Neural Network

S. Aronica[1], U. Brunelli[1], V. Piazza[1], L. Pignato[1], F. Sorbello[2], S. Vitabile[3]

[1]Dipartimento di Ricerche Energetiche ed Ambientali, Università di Palermo,
Viale delle Scienze, Edificio 9, 90128, Palermo, Italy
[2]Dipartimento Ingegneria Informatica, Università di Palermo,
Viale delle Scienze, Edificio 6, 90128, Palermo, Italy
[3]Dipartimento di Biotecnologie Mediche e Medicina Legale –Sezione di Radiologia
Università di Palermo Via del Vespro 90127 Palermo, Italy

ABSTRACT

The previsions of air pollutant trends have received great attention in the last years because the urban air pollution is an increasing problem. In particular the previsions of NOx, as an indicator of motor vehicle exhaust, is a useful tool to monitor pollution levels; it allows local authorities to manage the urban traffic to prevent the air pollution concentration values that are dangerous for human health. In the cities with heavy vehicular traffic it is possible to have high concentrations of air pollutants. This model can be useful for administrations to rationalise the preventive intervention to reduce the occurrence of high pollutants concentrations and to avoid useless annoyances to people. In this paper, the authors developed a recurrent neural network (RNN) model to forecast the maximum daily nitrogen oxides (NOx) concentrations in Palermo. The authors have compared the neural network results with those obtained from a stochastic model (AR2). The RNN model is different from the statistic models because it utilizes a pattern recognition approach. The research is based upon collected data from six monitoring sites during the period 2003-2004. The model developed is a potential tool for the predictions air quality parameters and it is superior to the traditional stochastic model.

Keywords: recurrent neural network, time-series forecasting, air pollution.

1. INTRODUCTION

The prediction of air pollution, in the urban areas, is one of the principal problems of air pollution quality research. Vehicular traffic is a major source of air pollutants, such as carbon monoxide (CO), benzene, nitrogen oxides (NOx) and polycyclic aromatic hydrocarbons. They are the principal causes of undesirable effects on health and even premature deaths. Grazuleviciene et al. [1] studied that long-term exposure to nitrogen dioxide (NO_2), an indicator of motor vehicle exhaust; increases the risk of myocardial infarction. This phenomenon seems to be most relevant in urban areas like Palermo (Fig.1), where peculiar orographic and atmospheric conditions can lead to pollutants accumulations.

Fig. 1. Map of Palermo

The increased sensitivity of national residents to environmental problems obliged the European Parliament to emit laws like Directive 2001/81/EC which sought to establish, for the first time, national emission ceilings for four pollutants – nitrogen oxide (NOx), sulphur dioxide (SO2), volatile organic compounds (VOC) and ammonia (NH3) – which cause acidification, eutrophication and ozone formation present at low altitudes. In urban areas Italian government has established to adopt as control parameters for the most polluted towns two level of concentration "attention and alarm ceilings". When they exceeded for some days the attention ceilings the administrations are obliged to take emergency measure for their reduction under the established limits. Many administrations try to reduce pollutant concentrations by limiting the vehicular traffic for some days (i.e. alternative circulation of even and odd number plates and "Sundays on foot"). In this situation models forecasting at lead-time tend to be an adequate method to plan a health warning system. When the hourly air pollution concentrations exceed the imposed values, the use of forecasting models can provide the best time to suggest that regulations are enforced. This would prevent unnecessary annoyances to the city residents. A great variety of operational warning systems based on statistical models have been developed. The stochastic models, well described by Box and Jenkins [2] and used by many researchers, forecast future values of a time series from current and past values. In our case we have compared the neural network results with those obtained from a stochastic model (AR2) and the results of the simulation show that the employment of a neural network is more efficient than the stochastic model. An artificial neural network has more flexibility than the stochastic models Kukkonen et al [3]. Neural networks have recently become an alternative to conventional methods and in the next features they are going to become the single most important instrument in air pollution distribution models Nagendra and Khare [4]. Viotti et al. [5], use a multy-layer perceptron neural network to forecast short and middle long-term concentrations levels for O3, NOx, NO2, CO. Hooyberghs et al. [6], describe the development of multi layer perceptron neural network to forecast the daily average PM10 concentrations in Belgian urban areas one day ahead. Zhang and San [7] use a wavelet neural network to model hourly NOx and NO2 concentrations of variance of emission source. Maqsood et al. [8] examine the applicability of the Hopfield model for hourly weather forecasting in a Canadian region. Model performance was compared with a standard MLP neural network, with an Elman recurrent network and with a Radial Basis Function Neural Network. The paper reports the results of the hourly forecasting in four typical days: one in winter, one in spring, one in summer and one in fall. The MLP and Elman networks have been labelled as the best forecasters. Elman neural networks have been successfully used in other forecasting applications and time series prediction. Luk et al. [9] evaluated three alternative ANN models for rainfall forecasting: a Multi-Layer Perceptron Neural Network, an Elman Neural Network and a Time Delay Neural Network (TDNN). The study eventuates that the three approaches have comparable performance as long as the complexity of the network is variable. In details the Elman network had the simplest structure, but was complex at the same time. As seeing in the literature review, artificial neural network represent the best model for air pollution statistical prediction; in particular multi-layer perceptron, with inherent static memory structure, is the preferred neural network in air pollution concentrations forecasting. In this paper the authors have used a recurrent neural network, this network has a dynamic memory. The outputs of the hidden and output layer are allowed to feedback onto themselves through a buffer layer, called the context layer. This feedback allows recurrent neural networks to learn, recognise and generate temporal patterns, as well as spatial patterns. The aim of this research is to develop a recurrent neural network (RNN) model to forecast the maximum daily nitrogen-oxides (NOx) concentrations in Palermo and comparing the results obtained with those obtained by a traditional stochastic model. The paper is organized as follows: in section 2 a theoretical description of the area and of the data used, in section 3 a description of the stochastic model used, in section 4 the background on the ANN model is provided, in section 5 and 6 the structures of Recurrent Neural Network model used are shown and the experimental results obtained.

2. DATA GATHERING

The ambient monitoring net (Fig.2), at Palermo, is managed by the AMIA (Municipal company Hygiene and Atmosphere), which at present consists of:

- 18 monitoring points, 10 of which are monitoring stations and 8 of which are mini-plants;
- 89 survey equipment items for chemical and physical air parameters;
- An elaboration, collection and registration centre, (CRED) Fig.3;
- Two information broadcast points.

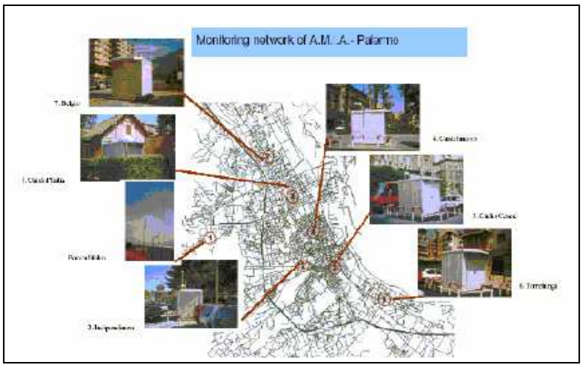

Fig. 2. Palermo monitoring network

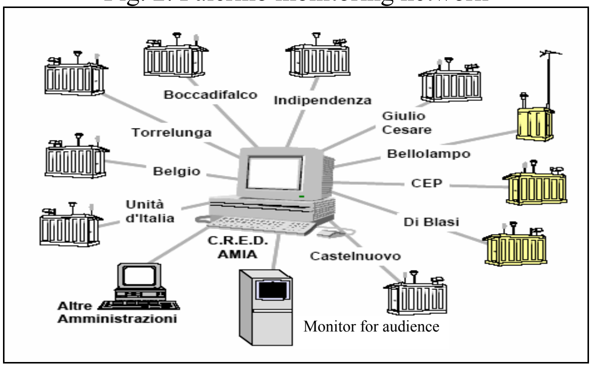

Fig. 3. Monitoring network structure

The meteorological monitoring stations are three: Bellolampo, Boccadifalco and Castelnuovo, they are located, respectively, at three different heights in respect to the sea level. The meteorological data includes the following parameters: wind direction and intensity, barometric pressure, humidity, solar radiation and ambient temperature. The monitoring stations are located in avenues where there is an important volume of vehicular traffic for every day of the week, which is more intense during morning hours. In these monitoring stations hourly concentrations of the following air pollutions were measured: carbon monoxide (CO), nitrogen oxides (NO_X), methane (CH_4), nitrogen dioxide (NO_2), ozone (O_3) suspended particulate (PM_{10}) and sulphur dioxide (SO_2). Until 2001, authorities were not required to respect particular laws about the emission ceilings. The situation changed in 2001 when the European Union set a national emission ceiling with Directive 2001/81/EC.The aim of this Directive is to limit emissions of acidifying and eutrophying pollutants and ozone precursors. The Directive has also the long-term objective not to exceed critical levels by establishing national emission ceilings, taking the years 2010 and 2020 as benchmarks (Official Journal 309). For Italy the emission ceiling is reported in Table1 and they are referred to Ministerial Decree 04/02/2002 nr 60 and to Legislative Decree 05/21/2004 nr 183.

AIR POLLUTANT	**SO_2 ($\mu g/m^3$)**		**CO (mg/m^3) average 8h**	**NO_2 ($\mu g/m^3$) average 1h**	**O_3 ($\mu g/m^3$)**		**PM10($\mu g/m^3$) average 24h**
	max 1h	**24h**			**max 1h**	**max 8h**	
Alarm Ceiling				$400^{(1)}$	240		
Attention Ceiling					180		
Limit for the protection of the human health	350	125	10	$250^{(2)}$		120	50
Exceeded maximum number for calendar year	24	3		18			35

Table.1. Emission ceiling of air pollutions.

(1) On three hours consecutive
(2) Valid until 12/31/2005, every year diminishes of 10 $\mu g/m^3$ until the final limit value of 200 $\mu g/m^3$ from 01/01/2010 in then.

The authors trained six recurrent neural networks for each model of RNN, with the data of nitrogen oxides (NO_X) concentrations of six stations in the Palermo area: Indipendenza, Belgio, Torrelunga, Unità d'Italia, Giulio Cesare and Di Blasi. Each input pattern is composed of twelve (hourly) values: wind direction and intensity, barometric pressure and ambient temperature, respectively at the stations of Bellolampo, Boccadifalco and Castelnuovo; they were recorded from January 1st 2003 to December 31st 2003 and are referred to a daily maximum nitrogen oxides (NO_x).

3. THE STOCHASTIC MODEL

In order to process the data presented in this research, the authors have employed an autoregressive model (AR2). This is one of the most frequently used approaches as far as air pollution forecast is concerned. In this model the history of atmospheric nitrogen oxides (NO_x) determines the NO_x concentration in the following sequence as the current value is expressed as a finite, linear aggregate of previous values of the process and a shock α_t. Thus:

$$\overline{z}_t = \phi_1 * \overline{z}_{t-1} + \phi_2 * \overline{z}_{t-2} + \cdots \phi_p * \overline{z}_{t-p} + a_t \tag{1}$$

with:

$$\overline{z}_t = z_t - \mu \tag{2}$$

where:

μ is the mean of values of time series, ϕ_i are the weights, and a_t is a white noise called shock, extracted from a fixed distribution, usually assumed to be normal and having mean zero and variance σ^2_α.

This method is called Autoregressive Process of order p, and it contains p+2 unknown parameters μ, ϕ_1, ϕ_2.......ϕ_π, σ^2_α, which in practice have to be estimated from the data. The additional parameter σ^2_a is the variance of the white noise process α_τ. The weights (ϕ_i) are obtained from the Yule-Walker equations which estimate autocorrelation with the Jenkins and Box technique. In our case the order of the process is two, so the weights are:

$$\phi_1 = (1 - r_2)/(1 - r_1^2) \tag{3}$$

$$\phi_2 = ((r_2 - r_1) * r_1)/(1 - r_1^2) \tag{4}$$

where r_1 and r_2 are the autocorrelation functions.

The time series was recorded between January 1st 2003 and December 31th 2003 and the forecast is referred to the whole year.

4. THE NEURAL MODEL

Artificial neural network (ANN) is composed by a great number of units joined together in a pattern of connectionism. In the network the units are shared in three groups: input units which received information to be processed, output units where the results of the processing are produced and in between hidden units. Each input unit has an activation value that represents some characteristic external to the network. The neural network operates in the following way: each input unit sends its activation values to each of the hidden units to which it is connected, then each hidden unit calculates its own activation value depending on the activation values that it received from the input units. This signal is then passed on to the output layer or to another layer of hidden units. The pattern of activation values is determined by the weights, or strength of connectionisms between the units. The weights may be either positive or negative. A negative weight represents the inhibition of the receiving unit by the activity of a sending unit. For each receiving unit, his activation value is calculated according an activation function. This function sums together the contributions of all sending units. A more realistic and plausible neural network model than the feed forward neural network is the RNN. The RNN is a neural network with feedback connections not unlike the human brain (see Fig.4). Its many layers of hidden units are called the Context Layer, and it includes recurrent connections that send the signals back from higher to lower levels. Such recurrence is necessary in order to describe such cognitive character as short-term memory. Feed forward neural networks have been successfully used to solve problems that require the computation of a static function whose output depends solely upon the current input. However, in a realistic case, problems can't be solved by a static learning function, as the function changes with each input received.

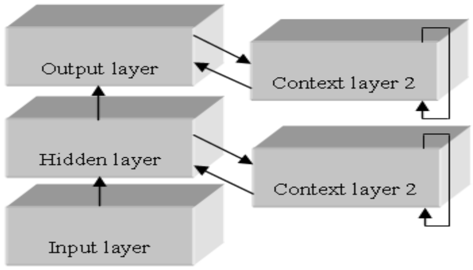

Fig. 4.The Elman network architecture

The recurrent neural network used in this research is the Elman type Elman, J. L. [10]. In this network, the outputs of the hidden and output layer are allowed to feedback onto themselves through a buffer layer, called the context layer. This feedback allows Elman networks to learn, recognise and generate temporal patterns, as well as spatial patterns. Every hidden neuron is connected to only one context layer neuron through a constant weight of value one. Hence the context layer virtually constitutes a copy of the state of the hidden layer one instant before. The number of context neurons is consequently the same as the number of hidden neurons. Optionally, every neuron of the output layer can be connected to only one neuron of a second context layer through a constant weight of value one. In our experiments, the sigmoid activation function for the hidden layer and the linear function for the output layer were always used.
The training algorithm is the Resilient Back Propagation (RProp), it is a local adaptive scheme, performing fast and robust supervised batch learning in neural networks M Riedmiller and H Braun [11], Hannan, J.M. & Bishop, J.M. [12], Vollmer and A. Strey [13]. The basic principle of RProp is to eliminate the harmful influence of the size of the partial derivative on the weight step. As a consequence, only the sign of the derivative is considered to indicate the direction of the weight update. Therefore each weight has its own adaptive step size Δji rather that the learning rate of the standard Back-Propagation algorithm:

$$\Delta_{ji}(t) = -\Delta_{ji}(t) * \text{sign}\ (g_{ji}(t)) \tag{5}$$

where g(t) is the gradient function and Δji is calculated as:

$$\Delta_{ji} = \begin{cases} \eta^{+} * \Delta_{ji}(t-1) & \text{if} \quad g_{ji}(t) \cdot g_{ji}(t-1) > 0 \\ \eta^{-} * \Delta_{ji}(t-1) & \text{if} \quad g_{ji}(t) \cdot g_{ji}(t-1) < 0 \\ \Delta_{ji}(t-1) & \text{otherwise} \end{cases} \tag{6}$$

In our trials the values of the parameters used to learn the neural network are: η^{+}=1,2, η^{-}=0.1, Δ_0=0.5 (fixed starting value for Δ_{ji}), Δ_{max}=50 (the upper limit for Δ_{ji}).

5. DATA PROCESSING AND NETWORK STRUCTURE
Each input pattern is composed of twelve (hourly) values: wind direction and intensity, barometric pressure and ambient temperature respectively from three different meteorological stations (Bellolampo, Boccadifalco, Castelnuovo).

Monitoring Station	Torrelunga	Unità D'Italia	Belgio	Di Blasi	Giulio Cesare	Indipendenza
Parameters Number	139	184	138	126	188	229

Table 2. Test set dimension.

The data were shared in two different sets: a training data set and a testing data set. Respectively, these are composed by the all data of the 2003-year and by the all data of the year 2004. The data sets were submitted after a scaling process, they were eliminated on the dates in which the instruments reported an error or when they were out of order. The test data set is composed by the data reported in Table2.
In addition, each value in the neural network was normalised in the range [-1, 1] using the following linear transformation:

$$X' = (X - V_m)/(V_{max} - V_{min}) \tag{7}$$

where X' is the new normalized value, X is the old value, V_{max} is the maximum of the considered data set, V_{min} is the minimum of the considered data set and V_m is the average value of the considered data set. The topology of neural network is a problem that depends on various factors. It is important to determine the appropriate network architecture in order to obtain the best results. Several artificial neural network topologies were implemented by changing the number of layers, and the number of the hidden and context units. Neural model simulations were performed using the Stuttgart Neural Network Simulator (SNNS) v. 4.1 [14]. The connection weights were initialised to zero-mean random values with adequate upper and lower bounds of (-1, 1). To define the optimum ANN structure using the common trial and error method. Different structures of the ANN have been tested with various different hidden nodes. It was found that thirty hidden nodes are the optimum for both ANN, in this experiment. The primary aim of developing an ANN is to generalise the features of the processed time series. A popular technique to achieve generalisation, avoiding over fitting, is the early stopping method presented by Sarle [15]. In the conducted experimental trials, training epochs were set to 150 for neural network model. To evaluate the model performance, the authors selected three parameters:

- MAE are defined by the following expressions:

$$MAE = \sum_{i=1}^{N} |O_i - P_i| / N \tag{8}$$

- Index of agreement (d):

$$d = 1 - (\sum (O_i - P_i)^2 / \sum (|P_i - O_i| + |O_i - O_m|)^2 \tag{9}$$

- Linear Correlation Coefficient (r):

$$r = 1 - (\sum (O_i - P_i)^2 / \sum (P_i - P_m)^2) \tag{10}$$

where O_i is the observed value at time i, P_i is the predicted value at time i, N is the total number of observations, O_m is the average of the observed values, P_m is the average value of the observed values.

6. EXPERIMENTAL RESULTS

The experimental results were described adopting the *d* and *r* index values. As shown in Fig.5 the *d* and *r* values are very close to 1, confirming the effectiveness of the proposed approach. In Fig.6 the MAE related to both RNN and stochastic model are shown.

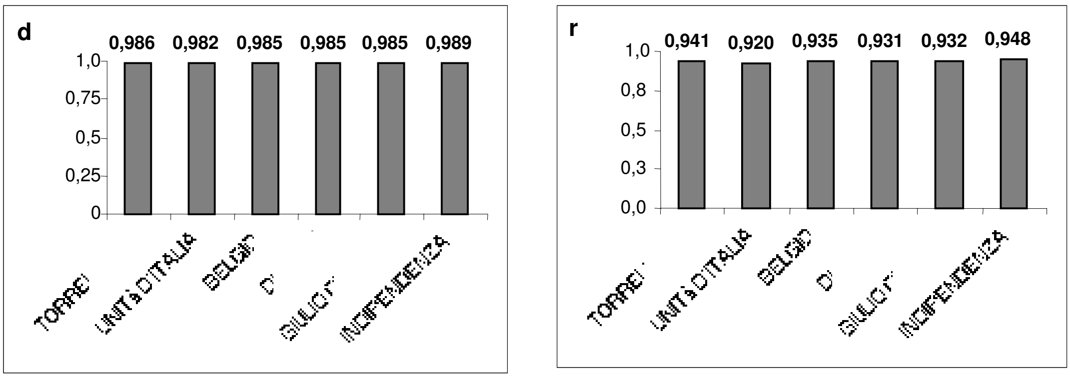

Fig. 5. Statistic Indexes for Recurrent Neural Network.

It has been demonstrated that The Recurrent Neural Network has a better behaviour in the prediction task, giving a better MAE for each monitoring station. Stochastic methods are not able to follow the trend of NO_x concentrations. Finally, Fig. 7 depicts both the observed and measured NOx concentration for the *Belgio* monitoring station. The average MAE is about 20 μg/m^3 considering a test time series of 138 days. Similar behaviours have been obtained for the other monitoring stations. An important aspect of the conducted experimental trials has been the difficulty for peak values forecasting. The success obtained in this type of experiment suggests that the application of modelling and forecast techniques for the prediction of complex natural phenomena with RNA deserves further attention and studies.

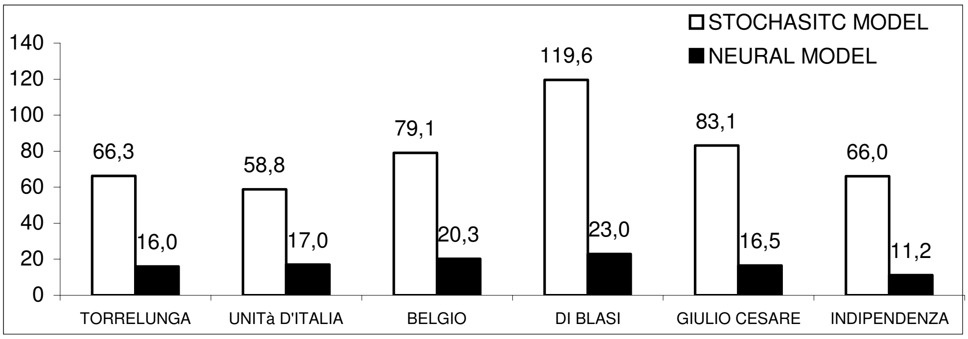

Fig. 6. Obtained MAE [μg/m^3] using RNN (black rectangle) and stochastic model (white rectangle).

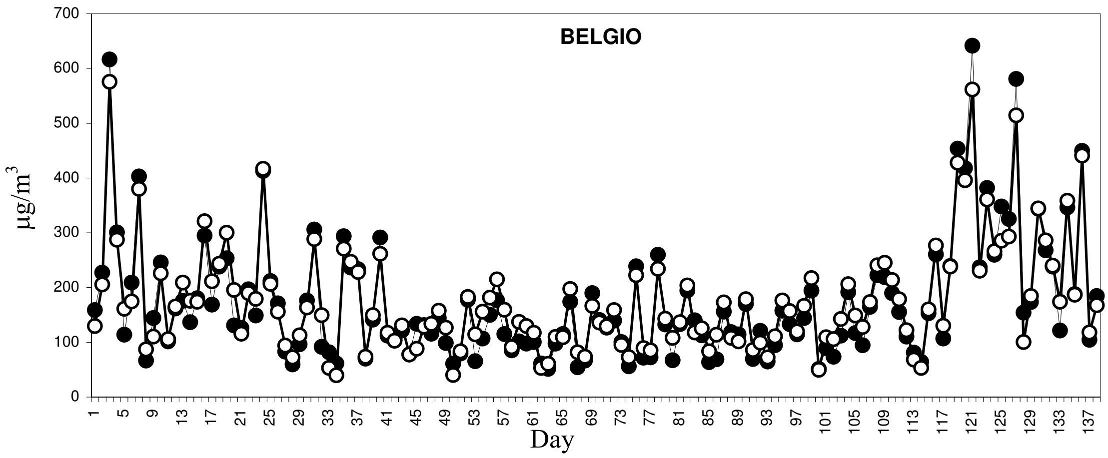

Fig. 7. NOx forecasted (white points) and measured (black points) concentration

7. CONCLUSION

The prevision of air pollutant trends has received great attention in the last years as urban air pollution is an increasing problem. In particular, the prevision of NO_x, as an indicator of motor vehicle exhaust fumes, is a useful in the monitoring of pollution levels. In this paper an Elman neural network forecaster was outlined and tested. The authors trained six recurrent neural networks for each model of RNN, with the data of nitrogen oxides (NO_X) concentrations of six stations in the Palermo area (Indipendenza, Belgio, Torrelunga, Unità d'Italia, Giulio Cesare e Di Blasi). Each input pattern is composed by twelve (hourly) values: the wind directions and intensity, barometric pressure and ambient temperature, respectively by the stations of Bellolampo, Boccadifalco and Castelnuovo; they were recorded from January 1st 2003 to December 31st 2003 and are referred to a daily maximum nitrogen oxides (NO_x). Experimental results show that the *d* and *r* values are very close to 1, that the Recurrent Neural Network has shown a better behaviour in the prediction task, giving a better MAE for each monitoring station. Stochastic methods are hence not able to follow the trend of NO_x concentrations. Finally, considering the observed and measured NOx concentration for the *Belgio* monitoring station, the average MAE is about 20 $\mu g/m^3$ considering a test time series of 138 days. Similar behaviours have been obtained for the other monitoring stations.

8. REFERENCES

[1] Nicolas L. Gilbert, Mark S. Goldberg, Bernardo Beckerman, Jeffrey R. Brook, Michael Jerrett. Assessing Spatial Variability of Ambient Nitrogen Dioxide in Montreal, Canada, with a Land-Use Regression Model. J. AIR & Waste Manage. Assoc.55:1059-1063.

[2] Box G E.P. and Jenkins G.M. Time Series Analysis: Forecasting and Control. Holden-Day. S. Francisco. ISBN 0-8162-1104-3.

[3] Jaakko Kukkonen, Leena Partanen, Ari Karppinen, Juhani Ruuskanen, Heikki Junninen, Mikko Kolehmainen, Harri Niska, Stephen Dorling, Tim Chatterton, Rob Foxall and Gavin Cawley.Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki. Atmospheric Environment, Volume 37, Issue 32, October 2003, Pages 4539-4550.

[4] S.M. Shiva Nagendra and Mukesh Khare. Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. Ecological Modelling, In Press, Corrected Proof, Available online 11 July 2005.

[5] P. Viotti, G. Liuti and P. Di Genova. Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia. Ecological Modelling, Volume 148, Issue 1, 1 February 2002, Pages 27-46.

[6] Jef Hooyberghs, Clemens Mensink, Gerwin Dumont, Frans Fierens and Olivier Brasseur. A neural network forecast for daily average PM10 concentrations in Belgium. Atmospheric Environment, Volume 39, Issue 18, June 2005, Pages 3279-3289.

[7] Zhiguo Zhang Ye San. Adaptive Wavelet Neural Network For Prediction Of Hourly NOx And NO2 Concentrations. Proceedings of the 2004 Winter Simulation Conference

[8] Jorge Henriques, Paulo Gil†, António Dourado and H. Duarte-Ramos. Application Of A Recurrent Neural Network In Online Modelling Of Real-Time Systems. ESIT ´99 European Symposium on Intelligent Techniques, June 3-4, 1999, Orthodox Academy of Crete, Greece with Introductory Fuzzy Control Internet Course, May 31 - June 2, 99.

[9] Imran Maqsood, Muhammad Riaz Khan, Ajith Abraham (2002) “Weather Forecasting Models Using Ensembles of Neural Networks” Third International Conference on Intelligent Systems Design and Applications, Intelligent Systems Design and Applications, Advances in Soft Computing, Springer Verlag, Germany, pp. 33-42, 2003.

[10] K.C. Luk, J.E. Ball, A. Sharma (2000) “A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting” Journal of Hydrology n.227 (2000) pp.56–65.

[11] Elman, J. L. (1990), “Finding structure in time”. In Cognitive Science, n.14 pp. 179-211.

[12] Vollmer and A. Strey, Experimental study on the precision requirements of RBF, RPROP and BPTT training, ICANN99. Ninth International Conference on Artificial Neural Networks, IEE Conf. Publ. N.470, IEE, London, UK, 239--44, 1999.

[13] M Riedmiller and H Braun. A direct adaptive method for faster BackPROPagation learning: The RPROP algorithm. In Proceedings of the IEEE International Conference on Neural Networks 1993 (ICNN 93), 1993.

[14] SNNS-Stuttgart Neural Network Simulator, url: http://www-ra.informatik.uni-tuebingen.de/SNNS/.

[15] W.S. Sarle, Stopped training and other remedies for over fitting, Proc. of the 27th Symp. On the Interface of Computing Science and Statistic, 352-360, (1995).