

A neural network model for three-hours-ahead prediction of ozone concentration in the urban area of Palermo

U. Brunelli, V. Piazza & L. Pignato

*Dipartimento di Ricerche Energetiche ed Ambientali,
Università di Palermo, Palermo, Italy*

Abstract

The purpose of this study is to use a recurrent neural network (Jordan model) to forecast ozone concentrations (O_3) with a short lead-time (1-3h) in the lower atmosphere. The network has been trained using a time series that was recorded between January 1st 2003 to December 31st 2003 and at two monitoring stations in Palermo (Italy). Each input pattern is composed of twelve (hourly) values: wind direction and intensity, barometric pressure, and ambient temperature; respectively gathered at the meteorological stations in Bellolampo, Boccadifalco and Castelnuovo. Ozone predictions are notoriously complex when using either deterministic or stochastic models which explains why this model was developed using a Neural Network. Neural Networks possess the ability to learn about non-linear relationships between the variables used. The model developed is a potential tool for the predictions air quality parameters and it is superior to the traditional stochastic model.

Keywords: emission sources, neural network, elemental concentrations.

1 Introduction

Vehicular traffic is a major source of air pollution, such as carbon monoxide (CO), benzene, nitrogen oxides (NO_x) and polycyclic aromatic hydrocarbons. They are the principal causes of undesirable health effects and even premature deaths. This phenomenon seems to be most relevant in urban areas like Palermo where specific orographic and atmospheric conditions can generate pollutants' accumulations. In this paper, the authors developed a statistical model to forecast the ozone (O_3) concentrations three hours ahead. Ozone is a colourless gas with a



strong odour and, if inhaled, can damage or kill living cells. It is normally created in the upper atmosphere by the action of sunlight on normal oxygen. Ozone is a secondary pollutant, meaning it is not directly emitted into the atmosphere, but is instead generated by chemical reactions in the atmosphere. The pollutants that react to produce a secondary pollutant are called “precursor compounds”. In the case of ozone, there are two key precursors: volatile organic compounds (VOCs) and nitrogen oxides (NO_x). Automobiles emit both of these pollutants. Nitrogen compounds are made when fuel is burned, such as in power plants and motor vehicles. In the upper atmosphere, ozone has a beneficial effect by absorbing the harmful ultraviolet rays of sunlight. At the earth’s surface, ozone is harmful to crops, forests, building materials and the health of humans and animals. Warm, sunny days with little wind and no rainfall help to form ozone. When these conditions occur over several days, ozone can continue to build up to unhealthy levels, especially in cities. Ozone can be blown long distances downwind of urban areas. Ozone can irritate the respiratory system at levels often found in many cities during the summer months. It not only affects people with existing breathing problems, but also can affect healthy children and adults. All this has obliged the European Parliament to emit laws as the Directive 2001/81/EC [10], which sought to establish, for the first time, national emission ceilings for four pollutants – nitrogen oxide (NO_x), sulphur dioxide (SO_2), ozone (O_3), volatile organic compounds (VOC) and ammonia (NH_3) – causing acidification, eutrophication and ozone formation present at low altitudes. In urban areas with high concentrations of traffic, Italian government and Local Administrations search to reduce pollutants’ concentrations. This is achieved by using cleaner fuels, electric transport and sometimes by ordering the closure to vehicular traffic for some days. In this situation lead-time forecasting models tend to be an adequate method when planning a health warning system. A great variety of operational warning systems based on empirical, deterministic, statistical, stochastic and hybrid models have been developed to allow preventive action when high pollutants’ concentrations take place. Lately, the increasing development of IT allows the use of artificial neural network (ANN) models and their development in air quality forecasting applications. As air dynamics vary on a seasonal basis, artificial neural networks can better represent air pollution distributions when compared to the general statistical and deterministic models. This has also been observed by Kukkonen et al [1] an artificial neural network has more flexibility than the statistical and deterministic models. The neural networks have recently become an alternative to conventional methods and in the next features they are going to be an important instrument in modelling air pollution distributions, Nagendra and Khare [2]. Viotti et al. [3], use a multi-layer perceptron neural network to forecast short and middle long-term concentrations levels for O_3 , NO_x , NO_2 , CO Hooyberghs et al. [4], describe the development of multi layer perceptron neural network to forecast the daily average PM_{10} concentrations in Belgian urban areas one day ahead. As seen in the literature review, artificial neural network represent the state-of-the-art for air pollution statistical prediction; in particular multi-layer perceptron, with inherent static memory structure, is the neural network mostly used in the air pollution



concentrations forecasting. In this paper the authors have used a recurrent neural network. This network has dynamic memory as the outputs of the hidden and output layer are allowed to feedback onto themselves through a buffer layer, called the context layer.

The aim of this research is to develop a recurrent neural network (RNN) model to forecast the ozone concentrations (O_3) with a short lead-time (1-3h) in the lower atmosphere. The paper is organised as follows: in section 2 a theoretical description of the area and the data used, in section 3 the background on the ANN model is provided, in section 4 and 5 the structures of Recurrent Neural Network model used and the experimental results obtained.

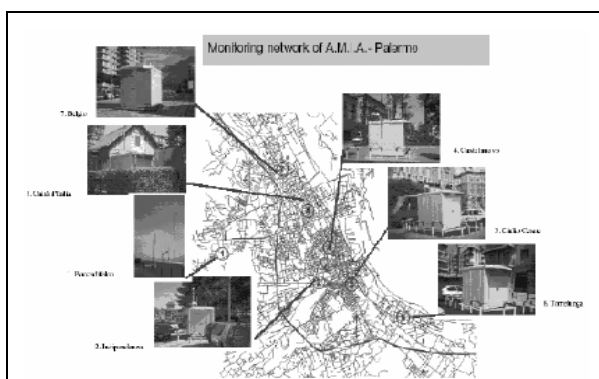


Figure 1: Palermo monitoring network.

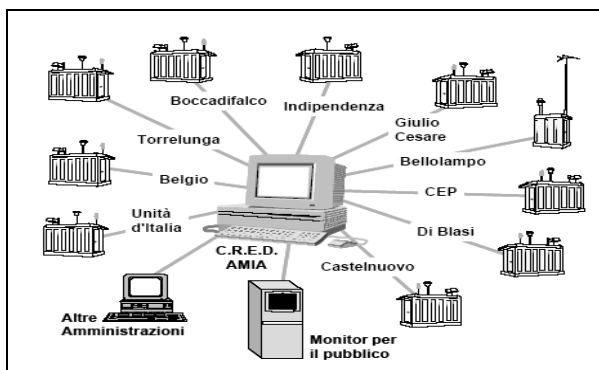


Figure 2: Monitoring network structure.

2 Area and data descriptions

The ambient monitoring net (Figure 1), at Palermo, is managed by the AMIA (Azienda Municipalizzata per l'Igiene e l'Atmosfera), it is at present composed from:

- 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) Figure 2;
- Two information broadcast points.

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 (CH₄), nitrogen dioxide (NO₂), ozone (O₃) suspended particulate (PM₁₀) and sulphur dioxide (SO₂). 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.

Table 1: Emission ceiling of air pollutions.

AIR POLLUTANT	SO ₂ (µg/m ³)		CO (mg/m ³)	NO ₂ (µg/m ³)	O ₃ (µg/m ³)		PM10(µg/m ³) average 24h
	max 1h	24h	average 8h	average 1h	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

(1) On three hours consecutive.

(2) Valid until 12/31/2005, every year diminishes by 10 µg/m³ until the final limit value of 200 µg/m³ from 01/01/2010 in then.



The authors trained two recurrent neural networks, with the data of ozone (O_3) concentrations of two stations (Boccadifalco, Castelnuevo). Each input pattern is composed of twelve (hourly) values: the wind directions and intensity, barometric pressure, and ambient temperature, respectively by the stations of Bellolampo, Boccadifalco and Castelnuevo; they were recorded from January 1st 2003 to December 31st 2003 and are referred to the ozone concentrations (O_3) with a short lead-time (1–3h).

3 Recurrent neural network

The artificial neural networks (ANN) can be grouped into two great categories: not recurrent networks, called also Feed Forward networks and recurrent networks, called also networks to Feedback. The Feed Forward neural networks operate as a combinatory system, where second output pattern and input pattern are associated in conformity with the relationship:

$$y(t) = f[x(t)] \quad (1)$$

On the other hand, recurrent networks operate as a sequential system, i.e. they associate output pattern to input pattern, with a dependency from a core state that evolves in the time as a function of the input introduced into the neural network. By representing the core state with a $Z(t)$ vector, we can define the operation of the network according to the following equations:

$$z(t+1) = f[z(t), x(t)] \quad (2)$$

$$y(t+1) = g[z(t), x(t)] \quad (3)$$

These equations define a dynamic system, in general not of linear type. This neural network model is more realistic and plausible than the feed forward neural network, as it is a neural network with feedback connections not unlike the human brain as seeing in Figure 3.

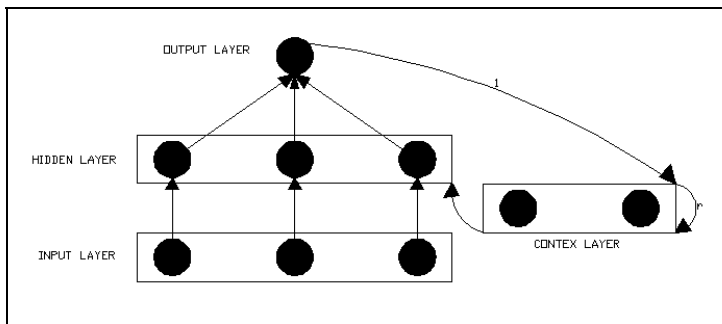


Figure 3: Sample of a recurrent neural network (Jordan).

As can be seen in figure 3, the neural network includes many layers of hidden units called contexts, and recurrent connections that send the signals back from higher to lower levels. Such recurrence is necessary to describe this cognitive character as short-term memory. Feed forward neural network have been

successfully used to solve problems that require the computation of a static function whose output depends only upon the current input, however, in a realistic case, because the function changes with each input received, the problems can't be solved by a learning static function. The recurrent neural network used in this research is the Jordan [1986] type.

The Jordan network is represented in Figure 3. It can be seen as a variation on a normal feed forward neural network, where feedback is added between the output and hidden layers, from which, in turn, a new layer is generated. This layer is called the context layer.

In particular, the context units receive the feedback signal at the input from the units of the output layer; the unit outputs of the context layer are distributed as inputs to the unit of the hidden layer. Connections between input layer and hidden layer, between hidden layer and output layer, and between context layer and hidden layer all have adaptive weight. On the other hand, output layer/context layer connections and self connections all have a fixed weight, as their sole function is to copy a fraction of their activation value. Usually such fixed value is one, apart for the auto-connections, for which an r -value, smaller than one, is prefixed, which is called recency constant. All the activation functions operate on the weighed sum of the input; they are: linear for the layers of input and context, while the sigmoid functions are used, typically, for the hidden and output layers. In this case however, the authors use the RM activation function for the output unit:

$$a_j(t) = \begin{cases} 0.85a_j(t-1) + 0.15net_j(t) \\ (1 - a_j(t-1)) \xrightarrow{\text{if}} net_j(t) > 0 \\ 0.85a_j(t-1) + 0.15net_j(t) \\ (1 + a_j(t-1)) \xrightarrow{\text{if}} net_j(t) \leq 0 \end{cases} \quad (4)$$

for the hidden units the BAM activation function:

$$a_j(t) = \begin{cases} 1 \xrightarrow{\text{if}} net_j(t) > 0 \\ a_j(t-1) \xrightarrow{\text{if}} net_j(t) = 0 \\ -1 \xrightarrow{\text{if}} net_j(t) < 0 \end{cases} \quad (5)$$

where $a_j(t)$ is the activation of unit j in step t , net_j is the input in unit j in step t . The training algorithm is the Resilient Back Propagation (RProp) algorithm. It is a local adaptive scheme, performing fast and robust supervised batch learning in neural networks, Riedmiller and Braun [6], Hannan and Bishop [5], Vollmer and Strey [7]. 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 Δ_{ji} rather than the learning rate of the standard Back-Propagation algorithm:

$$\Delta_{ji}(t) = -\Delta_{ji}(t) * \text{sign}(g_{ji}(t)) \quad (6)$$

where Δ_{ji} is calculated as:

$$\Delta_{ji} = \begin{cases} \eta^+ \Delta_{ji}(t-1) & \text{if} \\ & g_{ji}(t) \cdot g_{ji}(t-1) > 0 \\ \eta^- \Delta_{ji}(t-1) & \text{if} \\ & g_{ji}(t) \cdot g_{ji}(t-1) < 0 \\ \Delta_{ji}(t-1) & \text{otherwise} \end{cases} \quad (7)$$

where $g(t)$ is the gradient function. In our case the values of the parameters used to learn the neural network are: $\eta^+=1,2$, $\eta^-=0.1$, $\Delta_0=0.5$ (with a fixed starting value for Δ_{ji}), $\Delta_{\max}=50$ (the upper limit for Δ_{ji}).

4 Data processing and network structure

Each input pattern is composed of twelve (hourly) values: wind directions and intensity, barometric pressure and ambient temperature respectively from three different meteorological stations (Bellolampo, Boccadifalco, Castelnuevo). The data were shared in three different sets:

- A training data set
- A validation data set
- A testing data set

Respectively composed by 3792, 455, and 24 data gathered throughout 2003. The testing set is composed by the data of the day 7/01/2003 for the two stations. The data sets were submitted by a process of scaling, i.e. all the dates in which the instruments reported an error or were out of order have been removed. In addition, each value in the neural network was normalised in the range [-1, 1] using the following linear transformation:

$$X' = (X - V_m) / (V_{\max} - V_{\min}) \quad (8)$$

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[8]. The connection weights were initialized 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 [9]. In the conducted experimental trials, training epochs were set to 150 for the neural network model. To evaluate the model performance, the authors selected two parameters:

- Index of agreement (d):

$$d = 1 - \left(\sum (O_i - P_i)^2 / \sum (|P_i - O_i| + |O_i - O_m|)^2 \right) \quad (10)$$

which varies between 1 and 0. A computed value of 1 indicates perfect agreement between the observed and predicted values, while 0 connotes complete disagreement.

- Linear Correlation Coefficient (r):

$$r = 1 - \left(\sum (O_i - P_i)^2 / \sum (P_i - P_m)^2 \right) \quad (11)$$

The linear correlation coefficient measures the degree to which two variables are linearly related. 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, P_m is the average value of the predicted values, O_m is the average of the observed values.

Table 2: Statistical index.

Station	r	d
Boccadifalco	0,915	0,958
Castelnuovo	0,967	0,982

5 Conclusion

The results obtained from the RNA were described by the index values shown in Table 2. As can be seen, the results are very good; the values of d and r are very close to 1. A remarkable aspect in the experiments was the difficulty in predicting peak values. In point of fact, the network is not trained to recognise peak concentration values, as peaks do not appear in great numbers in the training set. The peak values tend to take place more in the validation and testing sets. Thus the neural network has not been trained to recognise such high concentrations. The success obtained in this type of experiment suggests that the

application of modelling and forecast techniques of complex natural phenomena with RNA deserves further attention and studies. The results of our model are described in Figure 4.

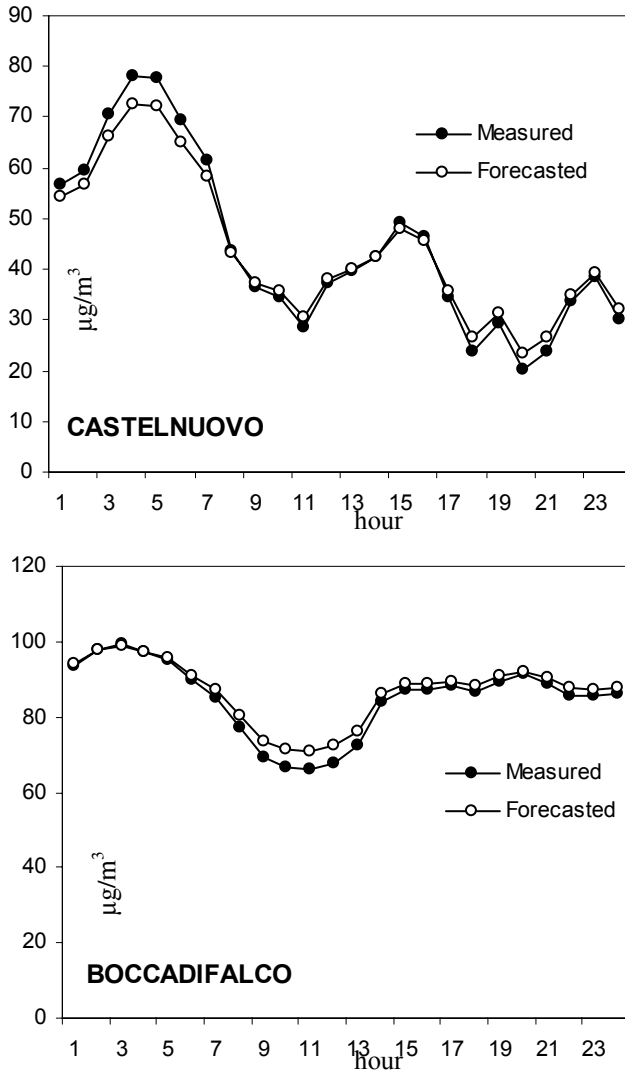


Figure 4: Results using the Jordan neural network – forecasted (white point) and measured (black point) O_3 ($\mu g/m^3$) concentration.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Hannan, J.M. & Bishop, J.M., (1997), A comparison of fast training algorithms over two real problems, *Proc. 5th Int. Conf. ANN*, pp: 1-6, Cambridge, UK.
- [6] 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.
- [7] 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.
- [8] SNNS-Stuttgart Neural Network Simulator, url: <http://www-ra.informatik.uni-tuebingen.de/SNNS/>.
- [9] 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).
- [10] Directive 2001/81/EC. Official Journal 309 of 27/11/2001 of The European Communities.

