IL NUOVO CIMENTO VOL. 23 C, N. 6 Novembre-Dicembre 2000

Nowcasting of urban air pollutants by neural networks (*)

F. BENVENUTO and A. MARANI
Università di Venezia, Dipartimento Scienze Ambientali - Dorsoduro, 2137, I 30123 Venezia, Italy

(ricevuto il 27 Luglio 1999; revisionato il 21 Giugno 2000; approvato il 25 Luglio 2000)

Summary. — The modelling of urban air quality prediction is a difficult task because: i) the processes are controlled by complex chemical and physical mechanisms; ii) its state is ascertained by measuring too few parameters for a sufficient chemical picture; iii) sampling measurements are generally collected at too few points without consideration of scaling (for example CO is a local phenomenon while O_3 is regional and both are often measured by the same monitoring network); iv) balances of chemical species are often forced to work far from "local equilibria". In order to overcome these problems, Artificial Neural Networks (ANNs) were used here because they are model free and require very little knowledge about the underlying system structure. CO, NO_2, and O_3 concentrations at the time $(t + \Delta t)$ are variables that depend on their previous concentrations and of other external information, such as meteorological data, solar radiation, chemical precursors or vehicle traffic information. ANNs used in this work were able to explain over 90% of the variability of the pollutant concentrations considered at the next hour (CO, NO_2, and O_3) and over 80% of that of the next three-hour O_3 concentration. The forecasting of CO peaks exceeding a given value has been successfully performed by transforming original concentration time series into a probability series and processing the transformed data by an ANN. Sensitivity analysis has provided useful insight into the most important forecasting variables and their relevant links.

PACS 92.60.Sz – Air quality and air pollution.

1. – Introduction

Photochemical pollution is a complex and highly non-linear process involving hydrocarbons (HC), nitrogen oxides (NO_x) and sunlight and influenced by the local meteorological conditions because: a) precursor emissions are weather sensitive (particularly to air temperature and humidity); b) transport processes depend on mixing depth, wind speed and atmospheric stability; c) chemical reactions are initiated by the ultraviolet radiation available. Some attempts to separate the meteorological from the chemical signal have already been done showing the inadequacy of the deterministic models to simulate urban air pollution (Rao and Zurbenko, 1994; Flaum *et al.*, 1996; Derwent and Hov, 1988).

(*) The authors of this paper have agreed to not receive the proofs for correction.

In this framework, for the management of urban air quality the regulation of vehicle traffic flow is needed, which is the main source of primary pollutants, using both *forecasting* (one or two days) and *nowcasting* (few hours or less) methods. The former is suitable to alert the population and to reduce or to stop vehicle circulation in selected areas, the latter to real-time regulation of vehicle distribution on the road network. Classic approaches (statistical and deterministic) show various difficulties, essentially due to the complexity of area textures (Boznar *et al.*, 1993; Nunnari *et al.*, 1998), to the wide database required (Robeson and Steyn, 1990) and to the poorly satisfied hypotheses for statistical applications (Burrows *et al.*, 1995; Clark and Karl, 1982; Comrie, 1997; Ryan, 1995).

The *Artificial Neural Networks* (ANNs) can overcome these difficulties because they are model-free working under the only hypothesis that the input variables (experimental space) form an almost complete phase space. In this space the best auto-regressive (AR) model can become a particular case (the linear limit) of ANNs, whereas the deterministic mathematical models simulate some conceptualized (subjective) process sometime using a parameter space larger than the experimental one (see also the Appendix). In this work we try to make use of all the time and space information available in the data collected by the local networks of air pollution monitoring stations in the investigated area. In particular, for O_3, CO, and NO_2 (regulated by Italian legal standards for the *attention* and *alarm* levels) we aim to: 1) show the power of ANNs for the *nowcasting* to 1 or 3 hour-lag ahead; 2) investigate the contribution of single inputs and the persistence effects; 3) identify the meteorological variables with the most influence on the short-term concentrations; 4) evaluate the ANNs nowcasting skill for exceeding values of standard concentrations (*e.g.*, Gardner and Dorling, 1999).

ANNs provide a powerful tool for data analysis and data processing (Bishop, 1994; Maier, 1995) which have been used for a wide variety of environmental applications, *e.g.* discrimination of seismic signals (Romeo *et al.*, 1995), controlling chaotic dynamical systems (Alsing *et al.* 1994), grassland community forecasting (Maier *et al.*, 1998), solar radiation and water quality simulation (Elizondo *et al.*, 1994), classification of remotely sensed data (Blamire, 1996). For a general survey of ANNs and for further references see Bishop (1994), Denning (1992), Hertz *et al.* (1991), Kosko (1992a, b), and Niemann (1989).

2. – The experimental data sets

The data set used (Liguori, 1996; Liguori *et al.*, 1997a, b) consists of hourly concentrations of air pollutants and meteorological parameters recorded at different urban sites during 1995 in Mestre (Venice, Italy, see fig. 1). The monitoring network is described in table I and included meteorological parameters from a private monitoring network (*Ente Zona Industriale di Porto Marghera*) and data from the air-quality network of the Venice Municipality. The large database (more than forty hourly time series, the shorter one with 7000 values), integrated with information about hourly motor vehicle traffic flows, allowed a broad statistical analysis preliminary to the modelling phase. However, in the literature ANNs have been used also for different data sets with a time range of much less than one year (Comrie, 1997; Nunnari *et al.*, 1998).

The Pearson (r's) correlation matrix among all the variables, used as a screening factor (and not as a statistical explanatory measure) is summarised in tables II a, b, c),

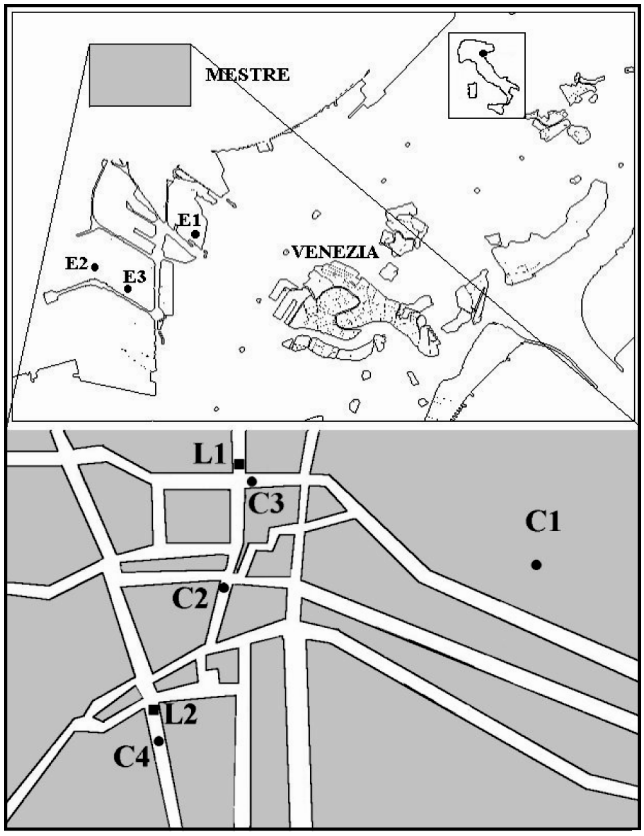

Fig. 1. – The Mestre urban area surrounding Venice are the monitoring networks, listed in table I.

where only the values of $|r|$ exceeding 0.7 are reported (the reference value $|r| = 0.7$ is not too far from the perfect correlation $|r| = 1$ and can make a sufficiently wide data set available). This matrix addresses the choice of relationships between variables and constraints for the neural structures and allows some qualitative observation:

– CO and NO are positively correlated with *Non Methanic Hydrocarbons* (NMHCs) detected at each station.

– O_3 is positively correlated with temperature, solar radiation, and wind velocity and negatively with temperature lapse rate.

– the presence of buildings and canyons probably induces different wind effects and little turbulence: so the indication of wind direction in a composite and complex structure like that one of a urban site is fairly meaningless.

– C3 station has been excluded from table II because its data are poorly correlated with data recorded at the other stations.

– Among the stations, the data detected at station C1 are correlated with C2 data just as data detected at C2 are correlated with C4 data, but C1 data are not correlated with those of C4 (on account of the fact that the roughness of the urban sites interferes with the complex structure of local turbulence).

TABLE I. – *Parameters measured at the monitoring stations.* THC = *total hydrocarbons*, NMHC = *non-methane hydrocarbons*, PM10 = *particulate suspended matter with diameter of* 10 μm *or less. The measurement units are the following*: °C *for temperature*, m/s *for wind velocity, degree from the north direction for wind direction*, hPa *for pressure*, W/m^2 *for the global solar radiation, the humidity is expressed as percentage of saturated air, precipitation in* mm, $\mu g/m^3$ *for the various concentrations, traffic flux is according to hourly data, heights are measured in meters* (*above the ground*).

Network	Station	Parameters	Height (meters, above the ground)
Ente Zona Industriale di Porto Margera	E1	Temperature and Wind (Velocity and Direction)	10
	E2	Wind (Velocity and Direction)	40
	E3	Temperature	10, 70, 140
		Pressure	10
		Global Radiation and Humidity	4
		Precipitation	0
Venice Municipality	C1 - Parco Bissuola	SO_2, NO, NO_2, CO, O_3, THC, NMHC, PM10	2–8
	C2 - Piazzetta Matter	SO_2, NO, NO_2, CO, O_3, THC, NMHC, PM10	2–8
	L1 - Crossing via Da Verrazzano-viale Garibaldi	traffic flux measuring station	—
	L2 - Crossing via Piave-via Carducci	traffic flux measuring station	—
	C3 - via Da Verrazzano	NO, NO_2, CO, O_3	2–8
	C4 - via Piave	NO, NO_2, CO, O_3, PM10	2–8

TABLE II. – a) *Cross correlation matrix among pollutants. In general, the values reported are only those greater than* 0.7 *in modulus* (*in* a), b), c) *the symbol — means value less than* 0.7).

Station		C1					C2					C4			
	Parameter	NO_2	NO_2	CO	O_3	NMHC	NO	NO_2	CO	O_3	NMHC	NO	NO_3	CO	O_3
C1	NO	1	—	—	—	0.78	0.80	—	—	—	—	0.73	—	—	—
	NO_2	—	1	—	—	—	—	0.74	—	—	—	—	—	—	—
	CO	—	—	1	—	0.76	—	—	0.76	—	—	—	—	—	—
	O_3	—	—	—	1	—	—	—	—	0.90	—	—	—	—	0.81
	NMHC	0.78	—	0.76	—	1	0.74	—	0.79	—	0.74	—	—	—	—
C2	NO	0.80	—	—	—	0.74	1	—	0.85	—	0.73	0.84	—	0.70	—
	NO_2	—	0.74	—	—	—	—	1	—	—	—	—	0.70	—	—
	CO	—	—	0.76	—	0.79	0.85	—	1	—	0.82	0.79	—	0.78	—
	O_3	—	—	—	0.90	—	—	—	—	1	—	—	—	—	0.82
	NMHC	—	—	—	—	0.74	0.73	—	0.82	—	1	—	—	—	—
C4	NO	0.73	—	—	—	—	0.84	—	0.79	—	—	1	—	0.83	—
	NO_2	—	—	—	—	—	—	0.70	—	—	—	—	1	—	—
	CO	—	—	—	—	—	0.70	—	0.78	—	—	0.83	—	1	—
	O_3	—	—	—	0.81	—	—	—	—	0.82	—	—	—	—	1

TABLE II. – b) *Cross correlation matrix among pollutants and meteorological parameters.* (These last two matrices are not written symmetrically purely for the sake of brevity.)

Station	C1							C2				
Parameter	WV		T10		ΔT21	Rad (*t*)	Rad (*t* − 2)	WV		T10		Rad (*t*)
	E1	E2	E1	E3				E1	E2	E1	E3	
NO	—	—	—	—	—	—	—	—	—	—	—	—
NO_2	−0.73	−0.74	−0.70	−0.71	—	−0.71	—	−0.75	−0.74	−0.70	−0.71	−0.72
CO	−0.70	−0.73	−0.76	−0.76	—	0.73	—	−0.71	−0.70	−0.72	−0.71	—
O_3	0.79	0.69	0.85	0.86	−0.72	0.71	0.89	0.73	0.69	0.78	0.79	0.74
NMHC	—	—	—	—	—	—	—	—	—	—	—	—

TABLE II. – c) *Cross correlation matrix among pollutants and meteorological parameters.* (These last two matrices are not written symmetrically purely for the sake of brevity.)

Station	C4				
	WV		T10		Rad (*t*)
	E1	E2	E1	E3	
NO	—	—	0.69	0.70	—
NO_3	−0.74	−0.73	−0.72	−0.73	−0.75
CO	−0.75	−0.71	−0.79	−0.75	−0.74
O_3	0.70	0.71	0.79	0.80	0.70

– The temperature at 140 m measured at E3 is correlated with temperatures at 10 m and at 70 m. The high correlation between the temperatures at 140 m and 10 m (and the simultaneous low correlation of air pollutants with the thermal lapse rate between 70 m and 10 m) can indicate that photochemical smog is linked to the ground temperature and not so much to the atmospheric stability.

– The ground temperature appears relevant for O_3 measured at C1 station, located in an open green park, and a little less for the remaining stations; probably because the C2 and C4 sites are more directly influenced by traffic flow.

The correlations among traffic fluxes and air pollutants were high only at the stations C2 and C4, the closest to the two traffic flux detectors L1 (measuring fluxes in the west↔east direction) and L2 (measuring fluxes in the south↔north direction), where the more significant values are those with CO.

Considering bi-directional fluxes (west↔east and south↔north direction), air pollution showed three peaks corresponding to three daily traffic peaks (7.30–9 am, 12 am–1 pm and 5–7 pm) and a more intense weekly presence of this phenomenon on Friday, when the traffic is higher (18% weekly total).

The persistence of CO, NO_2 and O_3 proved significant only for a few hours (less than three hours lag for CO and four hours lag for O_3 and NO_2 when the r value drops below 0.4). High auto-correlation with lags longer than 24 hours depends on the periodicities of meteorological and vehicle traffic effects.

On the whole, the observed temporal and geographical distributions of the air pollutants showed the expected behaviour of primary and secondary pollutants (Seinfeld, 1986). In particular, with wind calm: i) O_3, the earliest tropospheric oxidant measured (Roscoe and Clemitshaw, 1997) and the main index of photochemical smog (Chin *et al.*, 1994; Russell *et al.*, 1995), reached the highest concentrations in suburban areas (Pfeffer, 1994) during summer with intense solar radiation; and ii) CO and NO_2 concentrations are directly related to traffic flows (Karim and Matsui, 1995) only in winter when a stable atmosphere exhibits ground thermal inversion.

Before use, the data set has been made "uniform" by removing days with errors or missing data (approximately 5% of the data set) and by normalisation between 0.2 and 0.8 (in order to improve network sensitivity) through the relation (I_j = observed value,

$I_{\min}$ and $I_{\max}$ = respectively, the minimum and maximum value of the data set):

$$v_j = 0.2 + 0.6 \cdot (I_j - I_{\min}) / (I_{\max} - I_{\min}) . \tag{1}$$

Before comparisons with experimental data, the ANN outputs were rescaled by inverting eq. (1) (Comrie, 1997). These procedures provide comparable data ranges and avoid the asymptotes of the sigmoid function used as neurone activator.

The data were organised as sets of independent vectors consisting of values of input and output variables. These sets were divided, via random drawing without replacement, into two subsets: one for the training phase and the other, the complementary one, for the testing phase. The extraction procedures were repeated many times to have an objective set of parameters for checking the statistical quality of *neural* architectures.

3. – The neural network architectures

The neural architecture (SNNS, 1998) used in this work is the *feed-forward back-propagation* (*) with two layers of multiple neurones (the input and hidden layer) and an output layer with a single neurone. The input layer is the only one made up of linear neurones. This architecture, where every non-linear neurone (of the hidden and output layer) is connected with every neurone of the previous layer by weighted links and is activated by the sigmoid transfer function $f(x) = 1/(1 + e^{-x})$, is quoted as the most suitable for pollution prediction problems (Boznar *et al.*, 1993).

The performance of this type of ANN seems insensitive to the choice of the activation function so that the multi-layer feed-forward architectures have the potential of being universal learning machines (Hornik, 1990) and can give an arbitrarily accurate approximation to any arbitrary function (Hornik *et al.*, 1990). The feed-forward back-propagation learning procedure is driven by optimum criteria so that these ANNs are the most suitable for describing past experience without copying (Ruiz-Suârez *et al.*, 1995). The input sensitivity analysis of the ANNs used has been performed by retraining the networks without some selected inputs and observing the changes in the network outputs.

The choice of neural architectures is conditioned by available data and the tasks to be accomplished. In this work we tried to combine modelling efficiency and architectural simplicity. To do this, a single hidden layer and only one output neurone (pollutant concentration at time $t + 1$ or $t + 3$) were fixed whereas a correlation criterion guides the selection of input neurones, and the number of hidden neurones were chosen subjectively. As screening criterion of networks performances, Pearson's correlation index and other simple statistical tests were used (see table A).

When results looked good, more accurate statistical analyses have been carried out (see table B) based on the contingency tables.

As input data the O_3, CO, and NO_2 have been considered with some meteorological parameters: air temperature, thermal lapse rate, wind velocity and solar radiation

(*) The word *feed-forward* refers to the propagation of the information through the net (from the input layer to the output neurone); the word *back-propagation* refers to the learning algorithm which, with the aim of decreasing the error, proceeds by iterations from the output neurone back to the input layer.

TABLE A. – *Quality indices for screening tests of* ANNs *performances* (O_i = *observed values,* $\underline{O}$ = *average of observed values,* P_i = *predicted values,* N = *number of data*). *The threshold denotes the variability of the various parameters: for good performances each of them, excluding Willmott's indices, should be as small as possible* (*).

Name	Formula	Threshold
Mean absolute Error:	$\mathrm{ME} = N^{-1} \sum_{i=1}^{N} \lvert P_i - O_i \rvert$	$\geqslant 0$
Linear regression value:	$\widehat{P}_i = a + bO_i$ (a = intercept, b = slope of the linear regression between predicted and observed values)	
Root-Mean-Square Error:	$\mathrm{RMSE} = N^{-1}\left[\sum_{i=1}^{N} (P_i - O_i)^2\right]^{1/2}$ ($\mathrm{RMSE}=\mathrm{RMSE_s}+\mathrm{RMSE_u}$)	$\geqslant 0$
Unsystematic Root-Mean-Square Error:	$\mathrm{RMSE_u} = N^{-1}\left[\sum_{i=1}^{N} (P_i - \widehat{P}_i)^2\right]^{1/2}$	$\geqslant 0$
Systematic Root-Mean-Square Error:	$\mathrm{RMSE_s} = N^{-1}\left[\sum_{i=1}^{N} (\widehat{P}_i - O_i)^2\right]^{1/2}$	$\geqslant 0$
Willmott's index of agreement:	$d_j = 1 - \left[\sum_{i=1}^{N} \lvert O_i - P_i \rvert^j \Big/ \sum_{i=1}^{N} (\lvert P_i - \underline{O} \rvert + \lvert O_i - \underline{O} \rvert)^j\right] \quad j = 1, 2$	$0 \leqslant d_j \leqslant 1$

(*) ME and RMSE represent the mean value of the prevision (ME is less sensitive than RMSE to the extreme values of the differences between observed and predicted values); RMSE can be decomposed as $\mathrm{RMSE_s}$ and $\mathrm{RMSE_u}$ in order to give an insight of the influence of systematic errors on the model (the model can be considered satisfactory as much as $\mathrm{RMSE_s}$ is close to zero) (Willmott, 1982; Willmott *et al.*, 1985). Willmott's indices, varying from 0 (poor model) to 1 (perfect model), represent the ratio between the Mean-Square Error and the "Potential Error" (defined as the sum of squared absolute values of the distances from P_i to O_i and represent the largest value that can be attained for each observation/model-simulation pair). The advantage of d_1 (also known as the modified index of agreement) is that errors and differences are given their appropriate weighting, not inflated by their squared values. Squaring in statistics is useful because squares are easier to manipulate mathematically than are absolute values, but use of squares forces an arbitrarily greater influence by the larger values on the statistics. Experience using both d_2 and d_1 shows that in general $d_2 > d_1$ for the range of most values, although this relationship does not hold for extremely low values of both statistics (Legates and McCabe, 1999).

(which proved to have a great impact on the formation of O_3 at two time steps beforehand). For each measuring station, input variables with higher persistence (one, two, and/or three time-lags) and stronger correlation with the pollutant concentration to be forecast were chosen. The neural architectures implemented are similar to that of fig. 2. The utility of using the longest available times series, with a great number of variations and peak values, and without missing values of the involved variables, oriented the choice of the pollutant to be forecast in each monitoring station: O_3 at C1, NO_2 at C2, CO at C4. We have chosen to illustrate in this paper only the better results for each monitoring station and consequently to avoid showing various unsatisfactory forecasting attempts relative to all the three air pollutants measured at all the stations.

For the ozone nowcasting, two different neural networks were utilised: the first one ($1O_3$) for next-hour concentration $O_3(t+1)$ and the second one ($3O_3$) for the three-hour lag $O_3(t+3)$ concentration. The first network (fig. 2) utilised 10 neurones in the input layer selected by the aforementioned statistical analysis, 8 in the hidden layer (number

TABLE B. – *Formulas for testing the forecasting skill based on standard contingency tables at different thresholds* (*Wilks*, 1995). *A = number of simulated and observed events, B = numbers of observed but not stimulated events, C = number of simulated but not observed events, D = number of not simulated and not observed events.*

Name and Variability Range	Formula
Probability of Detection (measures the percentage of pollutant events that were correctly forecast). Range 0–1	$\mathrm{POD} = A/(A + B)$
Miss Rate (measures the rate at which pollutant events occurred but which failed to be forecast). Range 0–1	$\mathrm{MISS} = 1 - \mathrm{POD} = B/(A + B)$
False Alarm Rate (measures the tendency of the pollutant forecast to overpredict pollutant occurrences). Range 0–1	$\mathrm{FAR} = C/(C + A)$
Correct Null Forecast (measures the forecast skill at predicting "clean" days, it is the skill at which non-events are forecast). Range 0–1	$\mathrm{CNULL} = D/(D + C)$
Critical Success Index or Threat Score (combines forecast occurrences without regard to successful null forecast). Range 0–1.	$\mathrm{CSI} = A/(A + B + C)$
True Skill Score or Hanssen Kuipers Skill Score (includes the success of null forecast in the form of a ratio of observed skill to perfect forecast skill, this measure is not dependent on the relative frequency of occurrence and non-occurrence or the number of trials. If all forecasts are correct TSS = 1, if all forecast are incorrect TSS = − 1). Range − 1–1.	$\mathrm{TSS} = \mathrm{POD} + \mathrm{CNULL} - 1$
Heidke Skill Score (is often used as a measure of the skill of a set of forecasts compared to the skill of a random forecast). Range 0–1	$S = \dfrac{2(AD - BC)}{[(B^2 + C^2 + 2AD + (B + C)(A + D)]}$

chosen after various empirical tentatives) and 1 in the output (fixed *a priori*). The 10 input neurones were: four neurones representing the present concentration of ozone and that of three hours earlier ($O_3(t)$ and the auto-regressive component with longer persistence $O_3(t-3)$, $O_3(t-2)$, $O_3(t-1)$); two neurones representing two chemical variables involved in the photochemical smog mechanism ($NO_2(t)$ and non-methanic hydrocarbons NMHC(t) (Azzi *et al.*, 1995; Russell *et al.*, 1995; Marsili-Libelli, 1996)); and those meteorological variables for which the preliminary statistical analysis has shown more influence on O_3 concentration measured at C1: wind velocity WV(t), ground air temperature $T(t)$, thermal lapse rate $\Delta T(t)$ and solar radiation $\mathrm{Rad}\,(t-2)$ (this is due to a two-hour delay observed between ozone and solar radiation time series).

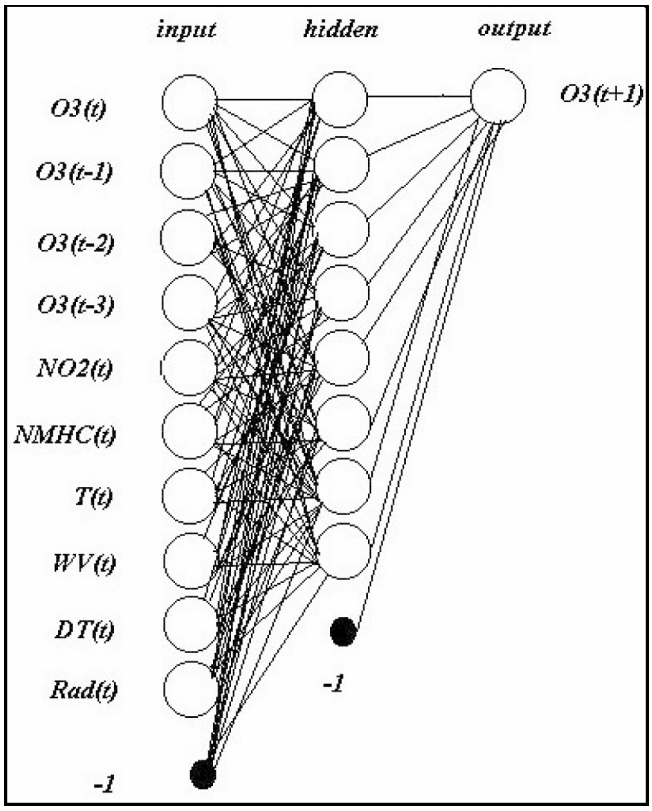

Fig. 2. – An example of the neural network architecture utilised for the 1-hour "nowcasting" O_3 model.

For the three-hour nowcasting $O_3(t+3)$, the network input was enriched with the chemical variable NO(t) recorded separately in the C2 monitoring station (the one more closely correlated with C1). In this way advantage has been taken of the same spatial information recorded in the dataset.

For CO, the network architecture (1CO) utilised for next-hour $CO(t+1)$ concentration nowcasting had 11 input, 8 hidden and 1 the output neurones. The CO time series utilised was the one relative to the C4 monitoring station, because it has the highest number of acute episodes. The model input was composed of: four neurones representing the CO concentrations at present time and at the three former hours (the auto-regressive component), one neurone representing the chemical variable NO(t) (at the same station, *i.e.* the one in which the correlation with CO proved closest), three meteorological variables highly correlated with CO (WV(t), $T(t)$, Rad(t)) and three neurones representing the traffic flow (Tr(t), *i.e.* the transiting vehicle number recorded at L2), the day of the week $d(t+1)$ and time of the day $h(t+1)$. $d(t+1)$ and $h(t+1)$, being well determined functions, are contemporaneous values of the output: it could be done assigning values $n = 1, 2, \ldots, 7$ to the neurone representing the day of the week and values $n = 1, 2, \ldots 24$ to the neurone representing the hour of the day. In this way it is possible to introduce into the models two variables controlling the

association between time intervals of more intense traffic flows and transportation of human beings for daily/weekly activities.

Concerning NO_2, the time series used for the 1-hour nowcasting was the one relative to the C2 monitoring station, due also to the great number of available data for this pollutant. Three different network architectures were tested, all with 7 input neurones, 4 of which were always WV(t), $T(t)$, Rad (t), $NO_2(t)$. The first architecture used as input also the NO_2 concentrations at the three previous times (the auto-correlation was significant only since lag 3) and 6 neurones in the hidden layer. The second one still used 6 hidden neurones but substituting as input the three NO_2 neurones with three new ones: the last three backward discrete time derivatives d of $NO_2[d(t), d(t-1), d(t-2)]$ with the aim of exploring the sensibility of the network to the increasing/decreasing local trend of NO_2. The third architecture used the same input neurones of the first one, but with two hidden layers, each composed by three neurones.

3. – The results

Descriptive statistics of the $1O_3$ and $3O_3$ model outputs are given in tables IIIa) and IV. The input-output correlation coefficient of the $1O_3$ model is very satisfactory and a bit lower than the one for the $3O_3$ model. The better results of the next-hour nowcasting with respect to the three-hour one is due to the difficulty in simulating the ozone behaviour at a long time-lag ahead. The results were also confirmed by contingency tables and by the classical forecast skills (tables V and VI) which concern nowcasting of peak events near the air quality standards of the Italian Law (180 μg m^{-3} for the "*attention*" level and 360 μg m^{-3} for the "*alarm*" level). However, the low presence of high concentrations negatively influenced the network ability to predict those exceeding the 180 μg m^{-3} and 360 μg m_1^{-3} thresholds because the output neuron is not trained enough. The $1O_3$ network is able to reproduce 4 of the 8 values exceeding 180 μg m^{-3} and in the remaining cases 50% of the concentrations reach values greater than 120 μg m^{-3}. Considering as new threshold a concentration of 120 μg m^{-3}, the number of concentrations exceeding it and correctly recognised by the $1O_3$ model was more than 70% (see table VI). In the case of three-hour nowcasting the $3O_3$ network recognised only 1 of the 2 values exceeding the threshold of 180 μg m^{-3}, and considering the threshold of 120 μg m^{-3}, the network recognised 45% of them in the training set and 53% in the testing set, as table VI shows (for the threshold 180 μg m^{-3} the POD is 50%, but there were too few acute episodes to be able to judge this case). The sensitivity analysis (SA) of the inputs of two neural architectures confirmed the complexity of O_3 nowcasting and showed the role of local meteorology in the dynamic of this pollutant. In particular, SA showed the similar relevance as input variables in the implemented neural models of O_3 persistence and of solar radiation on O_3 formation, a lower sensitivity of the models to wind velocity and to thermal lapse rate, and also similar input weights for NO_2 and NMHC. The residual auto-correlation was: for $1O_3$ model 0.01 at lag 4, and with all the other values less than 0.04, for the $3O_3$ model − 0.053 at lag 2 and 0.069 at lag 6 and 0.114 at lag 14.

For CO nowcasting, the last three variables introduced (Tr (t), $d(t+1)$, $h(t+1)$) allowed an improvement of the network performances: CO is, in fact, a primary pollutant whose main source is traffic flow. The input neurone Tr (t) gives important information for the CO prediction, and in this framework the two neurones $d(t+1)$ and

TABLE III. – *Descriptive statistics for the output testing sets of* O_3 (*part* a) *and* CO (*part* b) *networks: mean, standard deviation, the confidence intervals for the mean at* 95% *of significance, sample median, minimum and maximum value, standard error of the mean.*

a)

O_3 model with $\Delta t = 1$ h, "1O_3" network				O_3 model with $\Delta t = 3$ h, "3O_3" network		
$\mu g/m^3$	observed	simulated	residuals	observed	simulated	residuals
Mean	63.0	63.8	0.86	63.7	63.9	− 0.2
Stand deviation	43.1	41.1	13.5	44.7	42.8	24.7
Confidence interval (95%)	59.2–66.8	60.2–67.4	− 0.32–2.05	59.3–68.1	59.7–68.1	− 2.6–2.2
Median	57.0	57.6	2.0	56.0	55.3	− 2.0
Min	0.0	2.4	− 74.1	0.0	1.1	− 82.7
Max	191	189.4	70.0	191.0	229.7	78.8
Standard error	1.93	1.84	0.60	2.2	2.1	1.2

b)

CO model with $\Delta t = 1$ h						
"1CO" network				"1COAP" network		
$\mu g/m^3$	observed	simulated	residuals	observed	simulated	residuals
Mean	1.81	1.78	0.03	0.0023	0.0018	0.005
Stand deviation	1.43	1.14	0.77	0.0342	0.0385	0.0137
Confidence interval (95%)	1.7–1.92	1.7–1.87	− 0.03–0.09	$(-5–41)\cdot 10^{-4}$	$(-2–38)\cdot 10^{-4}$	$(-2–12)\cdot 10^{-4}$
Median	1.50	1.52	− 0.04	0	0	0
Min	0.10	0.37	− 2.85	0	0	− 0.3284
Max	14.60	10.92	6.97	0.9993	0.9997	0.2426
Standard error	0.06	0.04	0.30	0.0009	0.0010	0.0004

TABLE IV. – *Quantitative measures of agreement in the output testing sets of the neural network models for the three air pollutants considered. The terms R, b, d_i, R are dimensionless, while the remaining terms have the unit* μg m^{-3}.

	O_3 model with Δt=1 h: "1O_3" network	O_3 model with Δt=3 h: "3O_3" network
ME	±7.5	±28.8
Pearson correlation coefficient R	0.95	0.84
Linear regression:	$a = 6.7$ $b = 0.91$	$a = 12.6$ $b = 0.81$
RMSE	13.5	24.7
$RMSE_s$	4.0	8.5
$RMSE_u$	12.9	23.1
d_1, d_2	$d_1 = 0.87$ $d_2 = 0.97$	$d_1 = 0.76$ $d_2 = 0.92$
	CO model with Δt=1 h	
	"1CO" network	"1COAP" network
ME	±0.77	±0.014
Pearson correlation coefficient R	0.85	0.94
Linear regression:	$a = 0.56$ $b = 0.67$	$a = 0.0006$ $b = 1.1$
RMSE	0.77	0.0137
$RMSE_s$	0.47	0.0018
$RMSE_u$	0.61	0.00135
d_1, d_2	$d_1 = 0.70$ $d_2 = 0.91$	$d_2 = 0.86$ $d_2 = 0.96$

$h(t+1)$ have been introduced to weight the potential accumulations of CO, whose phenomenology is just related to the time of the day and the day of the week. The descriptive statistics and performance of 1 CO network are shown in tables IIIb) and IV and the results were also confirmed by the white-noise behaviour of the residuals' auto-correlation. Concerning nowcasting of peak values, the tests of tables V and VI are not satisfactory (*e.g.*, in the training set only one of the five concentrations greater than the "attention" level of 15 μg m^{-3} established by the Italian Law have been predicted) and the results were even worse in three-hour forecasting. These unsatisfactory results suggested the idea of transforming the original time series into another one measuring the probability of exceeding a fixed threshold with which to train a new neural network. For that it has been assumed that the measured data $\{C_i\}$ represent the most probable values of many repeated observations. If these observations are normally distributed (with density function $f_X(x, C_i, \sigma)$, where C_i and σ denote, respectively, expected value and standard deviation), then the probability of exceeding a threshold value C_A is given by

$$P_i^{C_A} = 1 - \int_{-\infty}^{C_A} f_X(x, C_i, \sigma)\, \mathrm{d}x \,; \tag{2}$$

the threshold value C_A could be, *e.g.*, the attention or the alarm level.

With $\sigma = 2.03$ μg m^{-3} (supposed constant and obtained from the entire data set) the time series of the probability $\{P_i\}$ of exceeding the CO attention level is shown in part b)

TABLE V. – *Verification of the pollutants forecast skill in the tesing set utilising standard contingency tables at different thresholds. A = number of simulated and observed events, B = number of observed but not simulated events, C = number of simulated but not observed events, D = number of not simulated and not observed events.*

1 $O_3 \geq 120$ μg/m^3	Forecast Yes	Forecast No
Observed Yes	42 (*A*)	17 (*B*)
Observed No	15 (*C*)	426 (*D*)
1 $O_3 \geq 180$ μg/m^3	Forecast Yes	Forecast No
Observed Yes	1 (*A*)	1 (*B*)
Observed No	0 (*C*)	498 (*D*)
3 $O_3 \geq 120$ μg/m^3	Forecast Yes	Forecast No
Observed Yes	29 (*A*)	26 (*B*)
Observed No	11 (*C*)	334 (*D*)
3 $O_3 \geq 180$ μg/m^3	Forecast Yes	Forecast No
Observed Yes	1 (*A*)	1 (*B*)
Observed No	2 (*C*)	396 (*D*)
CO ≥ 9 μg/m^3	Forecast Yes	Forecast No
Observed Yes	2 (*A*)	4 (*B*)
Observed No	0 (*C*)	652 (*D*)
CO ≥ 12 μg/m^3	Forecast Yes	Forecast No
Observed Yes	0 (*A*)	3 (*B*)
Observed No	0 (*C*)	661 (*D*)
COAP ≥ 0.5 (CO ≥ 15 μg/m^3)	Forecast Yes	Forecast No
Observed Yes	2 (*A*)	0 (*B*)
Observed No	0 (*C*)	1398 (*D*)
for $NO_2 \geq 200$ μg/m^3)	Forecast Yes	Forecast No
Observed Yes	(*A*) 45	(*B*) 16
Observed No	(*C*) 16	(*D*) 626

of fig. 3 (O_3 and NO_2 were not considered in this framework because the original data set already allowed to reach satisfactory results). It shows that the curve of the alarm probability (red) is simplified with respect to that of the original time series (blue) because all the values far from alarm are assimilated to "*zero events*" (*i.e. no events*). In this way a lot of *noise* has been removed and the peak values are emphasised in the probability series. The problem of a shortage of peak values could still be present, but suitable threshold values could compensate it. This method was applied with a network (COAP) implemented to nowcast the CO next-hour probability of exceeding the CO "attention" level established by the Italian Law (15 μg m^{-3}). COAP has the same architecture as the previous 1CO, but with the values of CO concentrations replaced by the related probabilities of crossing the attention level. Tables IIIb) and IV show the better performances obtained in 1-hour nowcasting. These results were confirmed by the white-noise behaviour of the residuals' auto-correlation ($COAP_{observed} - COAP_{simulated}$) always less than 0.13 with only a peak of -0.23 at lag 24 and by the characteristics of probability nowcasting shown in tables V and VI. It is important that the neural architecture has not been modified: this in order to put in evidence the advantage of the transformation (2). The sensitivity analysis has shown the primary

TABLE VI. – *Statistical analysis for* O_3, CO, *and* NO_2 (*the neural model with one hidden layer*) *on the testing sets based on the contingency tables.*

	$1O_3 \geq 120\,\mu g/m^3$	$1O_3 \geq 180\,\mu g/m^3$	$3O_3 \geq 120\,\mu g/m^3$	$3O_3 \geq 180\,\mu g/m^3$
POD	0.71	0.50	0.53	0.5
MISS	0.29	0.50	0.47	0.5
FAR	0.26	0.00	0.28	0.67
CNULL	0.97	1.00	0.97	0.99
CSI	0.57	0.50	0.44	0.25
TSS	0.68	0.50	0.50	0.5
S	0.69	0.67	0.56	0.4
	$CO \geq 9\,\mu g/m^3$	$CO \geq 12\,\mu g/m^3$	$1COAP \geq 0.5$	$NO_2 \geq 200\,\mu g/m^3$
POD	0.33	0	1	0.73
MISS	0.67	1	0	0.23
FAR	0	—	0	0.26
CNULL	1	1	1	0.97
CSI	0.33	0	1	0.58
TSS	0.33	0	1	0.70
S	0.5	0	1	0.71

behaviour of CO due to the high sensitivity of the COAP network to the traffic flow input variable (linked also with the time of the day and the day of the week), the important weight of persistence in the CO nowcasting, and the importanceof wind velocity.

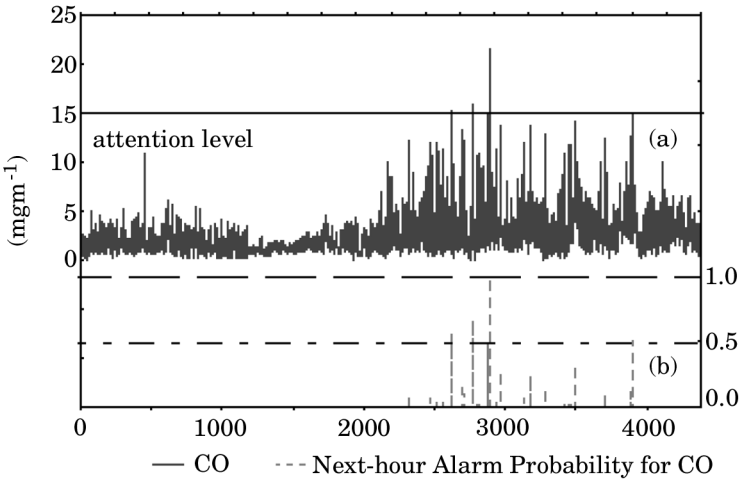

Fig. 3. – a) Original CO time series and the attention level threshold (15 μm m^{-3}); b) nowcasting of next-hour probability of exceeding the attention level. In formula (3) of the main text for C_A = 15 μg m^{-3}, C_i = 15 μg m^{-3} and σ = 2.03 μg m^{-3} the corresponding probability is 0.5. Values of the original time series (part a) exceeding the attention level correspond in part b) to probability peaks exceeding the horizontal line at 0.5.

As regards NO_2, the results obtained with the three different networks were generally quite similar: the r's coefficients for the three cases were $r_1 = 0.91$, $r_2 = 0.90$, $r_3 = 0.90$ and the corresponding mean errors were $ME_1 = \pm 11.5$, $ME_2 = 11.6$, $ME_3 = 11.7$. The satisfactory forecast skills of the first neural model are shown in tables V and VI, with the threshold of 200 $\mu g\ m^{-3}$. The first model is a little bit better than both the second and the third one and the classical statistical analysis, the tests on forecast skill (performed in the same way as for O_3 and CO) and the white-noise residual auto-correlation have confirmed it. The SA showed the significant role of temperature and the non-linear role of wind velocity in the NO_2 formation and transport in this area: among all the meteorological parameters considered, they proved the most closely linked to NO_2. It also showed the role of NO_2 short-term persistence which proved quite similar to its variation rate expressed by the time derivative at the same lags.

5. – Conclusions

The main results obtained in this work can be summarised in the following points:

a) ANNs are a good forecaster of hourly O_3 concentration at one time lag when the following inputs are used: i) present O_3 data and two time-lags earlier (the three-hour persistence is probably related to ozone production reactions being slow enough to produce delayed effects not longer than three hours); ii) present CO and NO_2 data (which take into account traffic fluxes and oxidant atmosphere capacity with CO, and traffic and ozone precursors with NO_2); iii) wind velocity; iv) air temperature; v) thermal lapse rate; vi) and two-hour lag before solar radiation;

b) in order to forecast CO peaks, it is better to introduce a probability measure of crossing a fixed threshold, thus defining the meaning of the word "peaks";

c) the ANNs implemented in this study are able to explain more than 90% of O_3, CO, NO_2 next-hour concentrations and more than 80% of next-three-hour O_3 concentration;

d) the use of more complex phase spaces (*e.g.*, utilising NMHC for CO and NO_2 and air humidity for O_3 forecasting) do not greatly improve nowcasting because the information added by other variables is already included in the input set used in the present work (*e.g.*, by auto-regressive data);

e) the short persistence of the pollutants considered does not allow long-range forecasting (one day or more) because their concentration is determined by meteorological variables which do not allow long-range forecasting windows. This is due both to sunlight periodicity (*i.e.* cloud variability), to low wind, rain, pressure, atmospheric humidity persistence and to their mutual independence which makes their combination less persistent than the parameters considered singly;

f) links among various chemical and meteorological parameters of the atmosphere are largely non-linear, but the preliminary treatment by Pearson's r is enough to detect a phase space in which the non-linear ANNs can work well;

g) results obtained with ANNs have not been compared with those with ARMAX models because these latter are a particular case of ANNs built on the same inputs. The difference between ANNs and ARMAX models operating in the same phase space lies in the deviation from the linearity produced by the transfer function used for neurones

activation. In this work, a good implementation of ANNs with *sigmoid* transfer function has been obtained utilising the interval [0.2–0.8] in which the shifting from the linearity (and consequently from ARMAX models) is approximately of 10% at the upper and lower limit.

These results confirm the ability of the ANNs for nowcasting air pollutant concentrations both in industrial and in urban areas. From this point of view, an ANN architecture *driven* by physical processes confers intelligence on a neural model which generally works as a "*grey box*". We deem that the use of ANNs as a predictor of air pollutants and the method here proposed could be considered a good basis for controlling the urban air quality by operating both in real time for "regulatory" purposes, and "*a posteriori*" for planning purposes.

* * *

The authors are grateful to Ente Zona di Porto Marghera, to the Venice Municipality, to Dr. Renzo Biancotto and Maria Rosa (PMP-ULSS 12 Mestre/Venezia) for providing the data. Special thanks are due to Dr. Francesca Liguori (ARPA-Padova) for her useful comments and suggestions.

Appendix

ANNs can be considered a product of artificial intelligence which, imitating the natural neurone networks, allow expert systems with learning skills. In this respect they are alternative to the expert systems based on choice trees driven by "*if ... then ...*" relationships. Actually, the literature suggests a high ANNs variety, often devoted to specific functions such as pattern recognition and ranking. The most suitable ANNs to interpret environmental processes (and, between them, the pollution one) are the ones known as *feedforward back propagation* because they map inputs to outputs non deterministically (like natural processes, they describe past experience without copying it) by optimising forecasting of the learning section. Avoiding complex mathematical proofs, it is possible to reach evidence of the ANNs ability to generalise classical models (*i.e.* the wide scenarios of auto-regressive ones) by writing input/output analytical transformation of a $(1, 1, 1, 1)$ network (see fig. 4a) with one input neurone, two hidden layers with one neurone, and one output neurone; and of another network $(1, 2, 1, 1)$ (see fig. 4b) with one input neurone, two hidden layers with, respectively two and one neurones, and one output neurone.

Case of fig. 4a)

$$\left(Y' = vX,\ Y = \frac{1}{1 + e^{-vX}},\ Z' = wY\right) \Rightarrow Z(x;\ v,\ w,\ u) = \frac{u}{1 + e^{-wY}}.$$

Case of fig. 4b)

$$\left(Y_1' = v_1 X,\ Y_2' = v_2 X,\ Y_1 = \frac{1}{1 + e^{-v_1 X}},\ Y_2 = \frac{1}{1 + e^{-v_2 X}},\ Z' = w_1 Y_1 + w_2 Y_2\right) \Rightarrow$$

$$\Rightarrow Z = \frac{u}{1 + e^{-(w_1 Y_1 + w_2 Y_2)}},$$

where $0 \leqslant u \leqslant 1$ was introduced in order to normalise the output Z. For example, in fig. 5 it is possible to compare the different outputs for different values of four weight

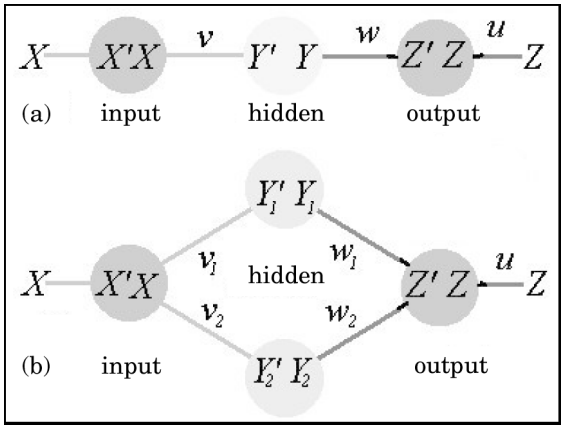

Fig. 4. – Two very simple neural architectures: a) (1, 1, 1, 1) one input neurone, one hidden layer with just one node, one output neurone; b) (1, 2, 1, 1) is a similar architecture but with two neurones in the first hidden layer and only one in the second hidden layer.

parameters of the (1, 2, 1, 1) network. This example shows the great flexibility of neural architectures, much wider than the linear models, for reproducing different data behaviours. These simple arguments allow the following considerations to be drawn regarding neural architectures:

i) the general non-linearity of input-output connections (including the linear limit as a special case),

ii) the great importance of hidden neurones as active neurones able to do both non-linear and linear transformations via the transfer function,

iii) the functional complexity of a neuronal architecture is related with the number of weights.

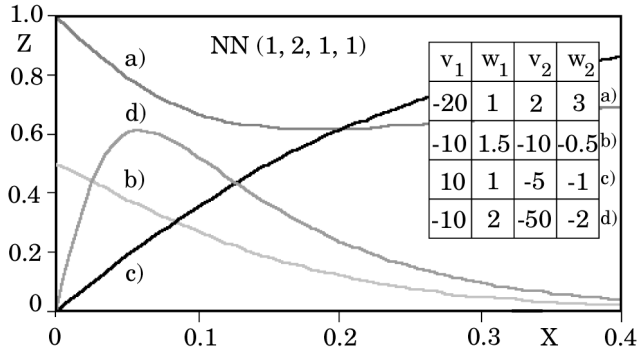

Fig. 5. – An example of the great variety of possible outputs that can be obtained by a (1, 2, 1, 1) network with some different weight values.

REFERENCES

ALSING P. M., GAVRIELIDES A. and KOVANIS V., *Using Neural Networks for controlling chaos. Phys. Rev. E*, **49** (1994) 1225.

AZZI M., JOHNSON G. M., HYDE R. and YOUNG M., *Prediction of* NO_2 *and* O_3 *concentrations for* NO_x *plumes photochemically reacting in urban air, Mathl. Comput. Modelling*, **21** (1995) 39.

BISHOP C. M., *Neural networks and their applications*, *Rev. Sci. Instrum.*, **65** (1994) 1803.

BLAMIRE P. A., *The influence of relative sample size in training artificial neural networks, Int. J. Remote Sensing*, **17** (1996) 223.

BOZNAR M., LESJAK M. and MLAKAR P., *A Neural Network based method for short-term predictions of ambient* SO_2 *concentrations in highly polluted industrial areas of complex terrain, Atmosph. Env. B*, **27** (1993) 221.

BURROWS W. R., BENJAMIN M., BEAUCHAMP S., LORD E. R., MCCOLLOR D. and THOMSON B., *Cart decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada, J. Appl. Met.*, **34** (1995) 1848.

CHIN M., JACOB D. J., MUNGER J. W., PARRISH D. D. and DODDRIDGE B. G., *Relationship of ozone and carbon monoxide over North America, J. Geophys. Res.*, **99** (1994) 14565.

CLARK T. L. and KARL T. R., *Application of prognostic meteorological variables to forecasts of daily maximum one-hour ozone concentrations in the North-Eastern United States, J. Appl. Met.*, **21** (1982) 1662.

COMRIE A. C., *Comparing Neural Networks and regression models for ozone forecasting, J. Air & Waste Manage Assoc.*, **47** (1997) 653.

DENNING P. J., *Neural Networks, Am. Sci.*, **80** (1992) 426.

DERWENT R. and HOV Ø., *Application of Sensitivity and Uncertainty Analysis Techniques to a Photochemical Ozone Model, J. Geophys. Res.*, **93** (1988) 5185.

ELIZONDO D., HOOGENBOOM G. and MCCLENDON R. W., *Development of a Neural Network Model to Predict Daily Solar Radiation, Agric. Forest Met.*, **71** (1994) 115.

FLAUM J. B., RAO S. T. and ZURBENKO I. G., *Moderating the influence of meteorological conditions ambient ozone concentrations, J. Air & Waste Manage. Assoc.*, **46** (1996) 35.

GARDNER M. W. and DORLING S. R., *Neural network modelling and prediction of hourly* NO_x *and* NO_2 *concentrations in urban air in London Atmospheric Environment, Atmosph. Env.*, **33** (1999) 709.

HERTZ J., KROGH A. and PALMER R. R., in *Introduction to the Neural Computation* (Addison Wesley) 1991, pp. 89-188.

HORNIK K., *Approximation Capabilities of Multilayer Feedforward Networks, Neural Networks*, **4** (1990) 251.

HORNIK K., STINCHOMBE M. and WHITE H., *Universal Approximation of an unknown mapping and its derivatives using Multilayer Feedforward Networks, Neural Networks*, **3** (1990) 551.

KARIM MD. M. and MATSUI H., in *Effects of wind speed on the dispersion of pollutant concentration in a road of Nagoya city*, in *4th International Conference on Computers in Urban Planning and Urban Management, Melbourne, Australia, 1995.*

KOSKO B. (a) In *Neural Networks for Signal Processing* (Prentice Hall, NJ) 1992, pp. 164, 166, 175, 180-182.

KOSKO B. (b) *Neural Networks and Fuzzy Systems; a dynamical approach to machine intelligence* (Prentice Hall, NJ) 1992, pp. 1995 181, 188, 196-212, 222, 327.

LEGATES D. R. and MCCABE G. J., *Evaluating the use of "Goodness of Fit" measures in hydrologic and hydroclimatic model validation, Water Resources Res.*, **35** (1999) 233.

LIGUORI F., *Inquinamento di atmosfere urbane: interpretazione dei dati e nowcasting con reti neuroneali*, Degree Thesis in Environmental Sciences (1996); web address http://www.ivsla.unive.it

LIGUORI F., MARANI A. and BENVENUTO F., *Nowcasting con reti neuroneali, AER: Meteorologia/ /Climatologia/Agrometeorologia/Ambiente*, **6** (1997, a) 9. URL http://motap.arpa.emr.it/Ravenna

LIGUORI F., BENVENUTO F., DAL BO' G. and MARANI A., *Quality control of meteorological time series with neural network, MAP Newsletter*, **7** (1997, b) 92.

MARSILI-LIBELLI S., *Simplified kinetics of tropospheric ozone, Eco. Mod.*, **84** (1996) 233.

MAIER R. H., *A review of Artificial Neural Networks, Research Report No. R 131*, Department of Civil and Environmental Engineering, The University of Adelaide (1995).

MAIER R. H., DANDY G. C. and BURCH M. D., *Use of artificial neural networks for modelling cynobacteria Anabaena spp. in the River Murray, South Australia, Eco. Mod.*, **105** (1998) 257.

NIEMANN H., *Pattern Analysis and Understanding* (2nd edition) (Springer, New York) 1989, pp. 199-205.

NUNNARI G., NUCIFORA A. F. M. and RANDIERI C., *The application of neural techniques to the modelling of time-series of atmospheric pollution data, Eco. Mod.*, **111** (1998) 187.

PFEFFER H. U., *Ambient air concentrations of pollutants at traffic-related sites in urban areas of North Rhine-Westphalia (Germany), Sci. Total Envir.*, **146-147** (1994) 263.

RAO S. T. and ZURBENKO I. G., *Detecting and tracking changes in ozone air quality, J. Air & Waste Manage. Assoc.*, **44** (1994) 1089.

ROBESON S. M. and STEYN D. G., *Evaluation and comparison of statistical forecast models for daily maximum ozone concentration, Atmosph. Env. B*, **24** (1990) 303.

ROMEO G., MELE F. and MORELLI A., *Neural Networks and discrimination of seismic signals, Computer & Geosci.*, **21** (1995) 279.

ROSCOE and CLEMITSHAW, *Measurement technique in gas-phase troposheric chemistry: a selective view of the past, present, and future, Science*, **276** (1997) 1065.

RUIZ-SUÂREZ J. C., MAYORA-IBARRA O. A., TORRES-JÌMÈNEZ J. and RUIZ-SUÂREZ L. G., *Short-Term Ozone Forecasting by Artificial Neural Network, Advances in Engineering Software*, **23** (1995) 143.

RUSSELL A., MILFORD J., BERGIN M. S., MCBRIDE S. M., MCNAIR L., YANG Y., STOCKWELL W.R. and CROES B., *Urban Ozone Control and Atmospheric Reactivity of Organic Gases, Science*, **269** (1995) 491-495.

RYAN W. F., *Forecasting severe ozone episodes in the Baltimore Metropolitan Area, Atmosph. Env.*, **29** (1995) 2387.

SEINFELD J. H., *Atmospheric Chemistry and Physics of Air Pollution* (Wiley & Sons, New York) 1986, pp. 3-194.

SNNS, *Stuttgart Neural Network Simulator*, Institute for Parallel and Distributed High-Performance Systems (IPVR), University of Stuttgart, 70565 Stuttgart (Germany), 1998.

WILKS D. S., *Statistical Methods in the atmospheric sciences: an introduction*, edited by Dmowska and Holton (Academic Press, International Geophysics Series San Diego, CA) 1995, pp. 233-258.

WILLIMOTT C. J., *Some comments on the evaluation of model performances, Bull. Am. Met. Soc.*, **63** (1982) 1309.

WILLMOTT C. J., ACKLESON S.G., DAVIS R. E., FEDEMMA J. J., KLINK K. M., LEGATES D. R., O'DONNELL J. and ROWE C. M., *Statistics for the evaluation and comparison of Models, J. Geophis. Res.*, **90/C5** (1985) 8995.