NEURAL NETWORKS FOR ENVIRONMENTAL PROBLEMS: DATA QUALITY CONTROL AND AIR POLLUTION NOWCASTING

F. BENVENUTO[1,]* and
A. MARANI[1,2]

[1] *Università di Venezia - Dip. Scienze Ambientali*
Dorsoduro 2137, I 30123, Venezia - Italy
[2] *Istituto Veneto di Scienze - Lettere ed Arti*
Campo Santo Stefano 2945, I 30124, Venezia - Italy

* to whom all correspondence should be addressed:
e-mail: fabiob@unive.it

ABSTRACT

This work illustrates the use and some related results of Artificial Neural Networks (ANNs) for data quality control of environmental time series and for reconstruction of missing data. ANNs are applied to the following problems: *i)* short and medium-term predicting of air pollutant concentrations in urban areas, *ii)* interpolating and extrapolating daily maximum temperature, *iii)* replacing time distribution with spatial distributed information (pollutant concentrations at different measuring sites). Observed versus predicted data are compared to test the efficacy of ANNs in simulating environmental processes. Results confirm ANNs as an improvement of classical models and show the utility of ANNs for restoration of time series.

INTRODUCTION

A great variety of environmental data written in paper format are at present available and it often seems useful to recover them by digitisation. In fact data referring to the past, if transferred on a digital support, can furnish information contents not repeatable in the future. The problem of checking data recorded by hand has never been studied before both because it has arisen recently and because it is different from those related to the automatic data acquisition.

These problems concern the missing data and the outliers of time series (and, symmetrically, spurious errors) since an uninterrupted time series gives far more information than two separate ones with a missing interval. As for the outliers, although they provide important information as they represent exceptional events, they can (also) lead to blunders if their exceptionality is due to spurious errors.

Data comparison techniques deal with important scaling problems, especially when the data sampling involves different scales.

It is in turn evident that each datum is related with its error and it is really impossible to make corrections of the data individually by hand: an automatic checking procedure is needed to get an insight into data *representativeness* and quality.

The missing data, the outliers and the quality problems call for the availability of a model whereby one can generate simulations and compare them with real data. By such a model one can produce almost true events that make up for the missing ones and also control the exceptional events. Recreating the mechanism of the process is unnecessary because it is irrelevant how the model produces data.

NEURAL NETWORK APPROACH

Artificial Neural Networks (*ANNs*) are a suitable model for the aforementioned purposes, provided that efficient architectures are available. An *ANN* can be viewed as a computer system that is made up of several simple and highly interconnected processing elements (McClelland, 1986) which process information by their dynamic state response to inputs. They provide a powerful tool for problems difficult to solve by traditional approaches, and frequently many of them have been addressed with neural networks: e.g. analysis of seismic signals (Romeo *et al.*, 1995), control of chaotic dynamical systems (Alsing *et al.*, 1994), predicting grassland community (Tan and Smeins, 1996) or solar radiation (Elizondo *et al.*, 1994), classification of remotely sensed data (Blamire, 1996; Côte and Tatnall, 1995), air quality control (Nunnari *et al.*, 1998; Gardner and Dorling,1999; Soia and Soia, 1999).

The neural architecture used in this work is the *feed-forward back-propagation*[1] consisting of only two layers of multiple neurons (the input and hidden layer) and of a single neuron in the output layer. 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 $f(x) = 1/\left(1 + e^{-x}\right)$, is cited 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 every 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. *ANNs* are particularly useful for investigations on large data sets and for problems with input/output relationships only partially known. In fact, *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 an *ANN*, whereas the deterministic mathematical models simulate some conceptualized (subjective) process sometimes using a parameter space larger than the experimental one.

The approach adopted in this work makes use of the time and space information collected by the monitoring networks and generalises other classical methods (the class of *ARMAX* models, see Appendix). For example an *ANN* with a linear activation function and with no hidden layers is equivalent to a multiple linear regression (Marzaban and Stumpf, 1996), moreover it is possible to show that a feed-forward back-propagation *ANN* with only one output neuron and with linear transfer function is a linear multivariate function of input variables. *ANNs* are *"grey box"* models if they account for the mechanisms of controlling processes by incorporating selected working conditions: this may be done by combining weight analysis, input-output sensitivity analysis, changes of network architecture, and physical information about the data set available (Benvenuto and Marani, 2000). The *ANNs* implemented in the present work, with a single hidden layer, were able to explain over *90%* of the variance of the environmental phenomena investigated.

---

[1] 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.

APPLICATIONS AND RESULTS

This work presents an overview of problems (and related findings) concerning the possibilities of using *ANNs* both as risk warning (nowcasting and forecasting) and as data interpolator, which take account of the data preceding and following the present datum. Moreover, topics regarding *ANNs* utilised as territorial homogeneity selectors have been considered, which could be useful, for example, to determine the boundaries of aerological basins.

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 normalization between *0.2* and *0.8* (in order to improve network sensitivity) thorough the relation (I_j=observed value, I_{min} and I_{max}=respectively the minimum and maximum value of the data set):

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

Before comparisons with experimental data, the *ANN* outputs was rescaled by inverting the equation (1) (Comrie, 1997); these procedures provide comparable data ranges and avoid the asymptotes of the sigmoid function.

The data were organized as sets of independent vectors consisting of values of input and output variables. These sets were divided, via random drawing without replacement, in two subsets: the former for the training phase and the other, the complementary one, for the testing phase (even not less than 40% of the entire data set). The extraction procedures were repeated several times to have an objective set of parameters for checking the statistical quality of neural architectures.

Forecasting

A problem arising from time series analysis is to forecast (medium/long term) or to nowcast (short term: *1* or *3* hours) the system's evolution. Predicting of photochemical smog is an example of complex data modelling because the processes involved are detected by measuring at only a few ground sites chemical indexes which depend on partially known chemical mechanisms, on poorly understood emission fields and on uncertain turbulent mixing and transport phenomena. The data sets used in this work (Liguori, 1996) consist
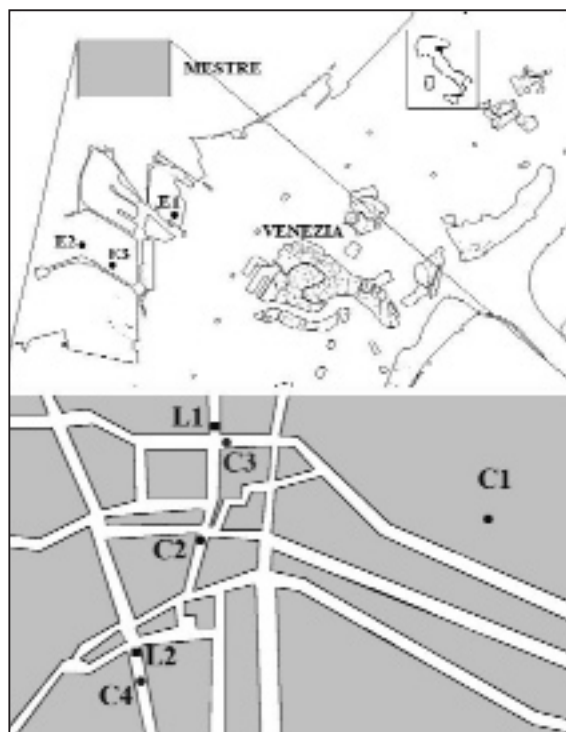

Figure 1. The urban area of Mestre surrounding Venice with indicated the monitoring networks listed in Table 1.

of hourly mean concentrations of air pollutants and meteorological parameters recorded at different urban sites during *1995* in Mestre (Venice, Italy - Figure 1).

The monitoring network is described in Table 1 and included meteorological parameters from a private monitoring network *(Ente Zona Industriale di Porto Marghera)*, data from the air-quality network of the Venice Municipality, and data on vehicle flow rates (Liguori, 1996). The large database of hourly time series (the shorter one with *7000* values) allowed preliminary broad statistical analysis. The *ANNs* implemented have been selected trying to achieve both modelling efficiency and architectural simplicity. The Pearson's correlation index with other simple statistical tests were used (Devore, 1990) as quick screening criterion of network performances and, only with the best results, more accurate statistical analyses were performed (systematic and unsystematic mean square error (Devore, 1990); Willimot indexes of agreement, (Willimot, 1982); probability of detection, missing rate, false alarm

Table 1.   Parameters measured at the monitoring stations, THC = total hydrocarbons, NMHC = non methane hydrocarbons, PM_{10} = particulate suspended matter with diameter of $10 \mu m$ or less. The measurement units are the following: ˚C for temperature, m s^{-1} for wind velocity, degree from the northern 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 as mm, μg m^{-3} for the various concentrations, vehicle flow rates were hourly data, heights are measured in meters (above the ground).

Network	Station	Parameters	Height
Ente Zona Industriale di Porto Margera	E1	Temperature Wind (velocity and direction)	10
	E2	Wind (velocity and direction)	40
	E3	Temperature Pressure Global Radiation Humidity Precipitation	10, 70, 140 10 4 4 0
Venice Municipality	C1	SO_2, NO, NO_2, CO, O_3, THC, NMHC, PM_{10}	2-8
	C2	SO_2, NO, NO_2, CO, O_3, THC, NMHC, PM_{10}	2-8
	L1	Vehicles flow rates measuring station	–
	L2	Vehicles flow rates measuring station	–
	C3	NO, NO_2, CO, O_3	2-8
	C4	NO, NO_2, CO, O_3, PM_{10}	2-8

rate, correct null forecast, critical success index, true skill score, Heidke skill score, (Wilks, 1995). An *ANNs* model for nowcasting hourly O_3, CO, NO_2 concentrations at one-hour lag (and, only for O_3, also at three-hours lag) has been implemented (by SNNS software [SNNS@SNNS Group, 1994]), similar to that of Figure 2, where the network for O_3 is reported. The present application highlights the efficiency of the method and its possibility of gaining insights into the mechanisms of the phenomena. The scheme considers each pollutant concentration at the time *(t+Δt)* as dependent variable of its past concentrations and of external components such as meteorological data (Table 1), chemical precursors and, in one case, also vehicles traffic information[2]. *ANNs* architectures similar to that of Figure 2 were able to explain over *90%* of the variability of the next hour concentrations of the three pollutants considered.

The network for O_3 (Figure 2) utilised 10 neurons in the input layer selected by statistical analysis, 8 in the hidden layer (number chosen after various empirical attempts) and 1 in the output (fixed a priori). The *10* input neurons were: four neurons representing the actual and the three former-hours ozone concentration [$O_3(t)$ and the auto-regressive component with longest persistence $O_3(t-3)$, $O_3(t-2)$, $O_3(t-1)$]; two neurons representing two chemical variables involved in the photo-chemical smog mechanism [$NO_2(t)$ and non-methanic hydrocarbons *NMHC(t)*, (Azzi *et al.*, 1995; Russel *et al.*, 1995)]; and those meteorological variables for which the preliminary statistical analyses have shown more influence on O_3 concentration [wind velocity *WV(t)*, ground air temperature *T(t)*, solar radiation *Rad(t-2)*[3] and thermal lapse rate *ΔT(t)*]. The good results were also confirmed by the contingency tables and by the

---

[2] This last variable measured at *L1* and *L2* (Figure 1), resulted highly significative only for *CO* nowcasting and for O_3 spatial distribution analysis; due to this reason it was omitted from the other basis neural architectures.

[3] This is due to a two hours delay observed between ozone and solar radiation time series.
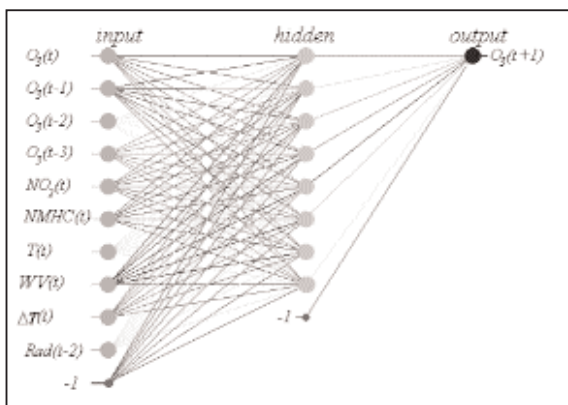
Figure 2. An example of the neural network architecture utilised for the 1-hour and 3-hour "nowcasting" O_3 model.

classical forecast skills (Table 2 and Table 3), which concern nowcasting of peaks events near the limits established by the Italian Legislation *(180μg m^{-3} is the attention level and 360μg m^{-3} is the alarm level)*. However, the low presence of high concentrations influenced negatively the network ability to predict the exceeding of the Italian Low thresholds because the output neuron is not trained enough. Considering the concentration threshold of *180μg m^{-3}*, the percentage of concentration exceeding the threshold and correctly recognised by the *1-hour O_3 model* was more than *70%*. In the case of *3-hours O_3 model*, results were obviously worse: the output of the network recognised only *1 of 2* values exceeding the threshold of *180μg m^{-3}*. The sensitivity analysis *(SA)* of the inputs of two neural architectures relative to O_3 nowcasting at *1 or 3* hours ahead (architecture of Figure 2) 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 relevance of solar radiation on O_3 formation and the low sensitivity to wind velocity and to thermal lapse rate, and similar weights for *NO_2* and *NMHC*. The residual auto-correlation was: for the *1-hour O_3* model *0.01* at lag *4*, and with all the other values less than *0.04*; for the *3-hours O_3* model *-0.053* at lag *2*, and *0.069* at lag *6*, and *0.114* at lag *14*.

Table 2. Verification of the pollutants forecast skill in the testing set utilising standard contingency tables (Devore, 1990; Wilks, 1995) at different thresholds; A= number of simulated events and observed, B= number of observed but not simulated, C= number of simulated but not observed, D= number of not simulated and not observed

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

Table 3. Statistical analysis for O_3, CO, and NO_2 ($1O_3$ means 1-hour lag, $3O_3$ means 3-hours lag) on the testing sets based on the contingency tables: (i) the Probability of Detection [POD=A/(A+B)], measures the percentage of pollutant events that were correctly forecast; (ii) the miss rate [MISS=1-POD=B/(A+B)] measures the rate at which pollutant events occurred but failed to be forecast; (iii) the false alarm rate [FAR=C/(C+A)] measures the tendency of the pollutant forecast to over-predict pollutant occurrences; (iv) the correct null forecast [CNULL=D/(D+C)] measures the forecast skill at predicting "clean" days, it is the skill at which non events are forecast; (v) the critical success index [CSI=A/(A+B+C)] or threat score combines forecast occurrences without regard to successful null forecast; (vi) the true skill score [TSS=POD+CNULL-1] 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); (vi) the Heidke skill score {$S=2(AD–BC)/[B^2+C^2+2AD+(B+C)(A+D)]$} is often used as a measure of the skill of a set of forecasts compared to the skill of a random forecast (Wilks, 1995)

	$1O_3 \geq 180$ $\mu g\ m^{-3}$	$3O_3 \geq 180$ $\mu g\ m^{-3}$	$CO \geq 9$ $\mu g\ m^{-3}$	$CO \geq 12$ $\mu g\ m^{-3}$	$COAP \geq 0.5$	$NO_2 \geq 200$ $\mu g\ m^{-3}$
POD	0.71	0.53	0.33	0	1	0.74
MISS	0.29	0.47	0.67	1	0	0.26
FAR	0.26	0.28	0	-	0	0.26
CNULL	0.97	0.97	1	1	1	0.98
CSI	0.57	0.44	0.33	0	1	0.58
TSS	0.68	0.50	0.33	0	1	0.72
S	0.69	0.56	0.5	0	0.5	0.71

The forecasting of O_3 daily maximum values has also been considered: in fact for human health what is really important to know is whether the threshold maximum values could be exceeded, more than the medium values forecasted. In the network input patterns hourly concentration has been replaced by daily maximum concentrat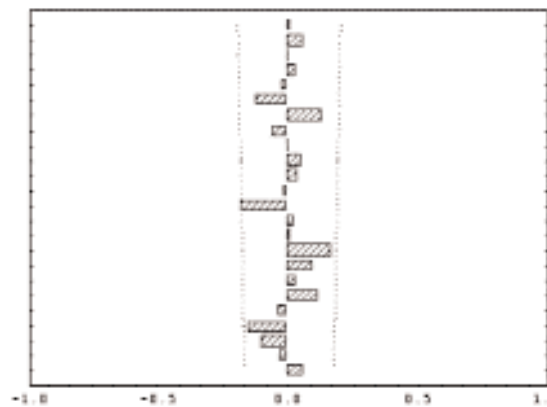ions, and the values of the others parameters measured at the same time of the daily maximum have also been included. The architecture used is almost the same as the one for O_3 *1-hour* and *3-hours* now-casting, but in order to take into account the spatial variability of the phenomenon, the maximum values measured in the remaining monitoring stations (different from that of reference) have been included in the input patterns: in this way an eventual scarcity of data could be bypassed. Figure 3 shows the white noise behaviour of the residual auto-correlation of daily O_3 forecasting (observed-forecasted values) and Table 4 the statistical forecast skill of the network.

Figure 3. Residual (observed-forecasted values) auto-correlation of O_3 daily maximum values: white noise behaviour is evident.

Table 4. Statistical skill of O_3 daily maximum values forecasting

POD	0.75
MISS	0.25
FAR	0.38
CNULL	0.88
CSI	0.52
TSS	0.63
S	0.58

Regarding the forecasting of *CO* peak values exceeding a fixed threshold, it has been successfully performed by transforming the concentrations time series into another one representing the probabilities of exceeding the attention threshold and then processing the transformed data by an *ANN*. This method was applied with a network (*COAP*) implemented to nowcast the *CO* next-hour probability of exceeding the attention level *(15 μg m⁻³)*. *COAP* has a similar architecture to the previous ones: *11* input, 8 hidden and 1 the output neurones. The model input was composed of: four neurones *PCO(t)*, *PCO(t-1)*, *PCO(t-2)*, *PCO(t-3)* representing the *CO* probabilities of crossing the attention level at present time and at the three former hours (the auto-regressive component), three neurones representing meteorological variables highly correlated with *CO* [*WV(t)*, *T(t)*, *Rad(t)*], by one neurone representing the chemical variable *NO(t)* and by three neurones representing the traffic flow [*Tr(t)*, i.e. the hourly transiting vehicle number]. 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 by assigning values *n=1,2,..,7* to the neurone representing the day of the week and values *n=1,2,...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. Table 2 and Table 3 show the performances this network obtained in *1*-hour nowcasting (*0.5* is the corresponding probability of exceeding the attention level, *15 μg m⁻³*), these results were confirmed by the white noise behaviour of the residual auto-correlation (*COAP_{observed}* - *COAP_{simulated}*) always less *0.13* with only a peak of *-0.23* at lag *24*. Figure 4 shows: (part a) original *CO* time series and the attention level threshold; (part b) nowcasting of next-hour probability of exceeding the attention level: even in this case, all the statistical tests on the model outputs gave excellent results with respect to the forecasting of the original concentration time series (threshold at *9 μg m⁻³* and *12 μg m⁻³*, see Table 2 and Table 3). It has confirmed the advantage of emphasising highest values in order to predict the peaks more accurately. The sensitivity analysis has shown the primary behaviour of *CO*
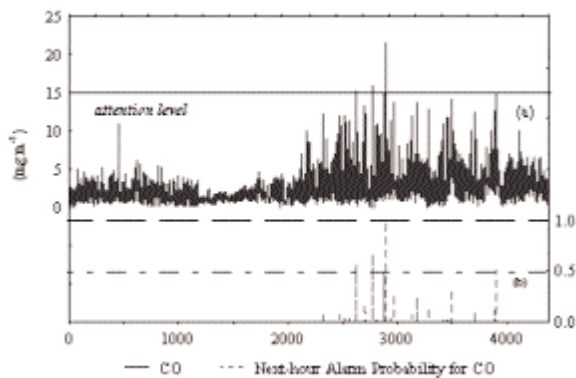

Figure 4. (a) original *CO* time series and the attention level threshold *(15 μg m⁻³)*; (b) nowcasting of next-hour probability of exceeding the attention level.

due to the high sensitivity of the *COAP* network to the traffic flow input variable.

Concerning *NO₂*, three different network architectures were tested, all with *7* input neurones, *4* of which were always *WV(t)*, *T(t)*, *Rad(t)*, *NO₂(t)* measured at the same station of *NO₂*. The first architecture used as input also the *NO₂* concentrations at the three previous times (the auto-correlation was significant only up to lag 3) and 6 neurones in the hidden layer. The second one still used 6 hidden neurones but substituting as input the three *NO₂* neurones with three new ones: the last three backward discrete time derivatives *d* of *NO₂ [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 *NO₂*. The third architecture used the same input neurones as the first one, but with two hidden layers, each composed by three neurones. The performances of the three different networks were quite similar: the r's coefficients for the three cases were $r_1=.91$, $r_2=.90$, $r_3=.90$ and the corresponding mean errors were $ME_1=\pm11.5$, $ME_2=\pm11.6$, $ME_3=\pm11.7$. The satisfactory forecast skills of the first neural model are shown in Table 2 and Table 3, with the threshold of *200 μg m⁻³*. The first model is a little bit better than both the second and the third one and the classical 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
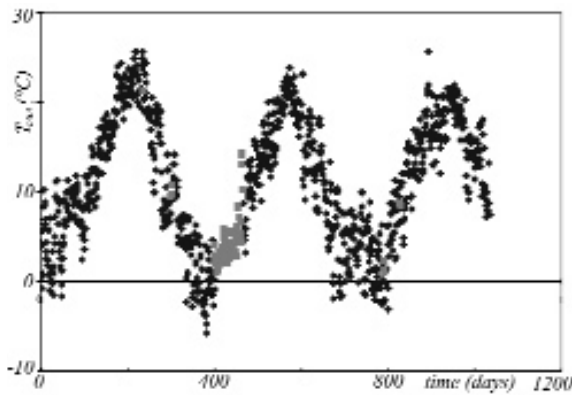
Figure 5.   Missing daily minimum temperatures and the recreated ones by means of the neural network: the original data are coloured blue, the missing reconstructed values are coloured pink

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.

Interpolation and data quality control

The application of *ANNs* for data interpolation oriented to generate or reconstruct missing data has been proved useful. As an example we describe here the good results reached in a set of daily minimum temperatures (Figure 5) including *1035* current data having isolated missing data and a *61*-day interruption. The network was set up in connection with the gap to be filled: at the first step the one-day gaps were filled, than the two-day gaps and so forth. At last, to fill in the *61*-day gap, subsystems with *60*-day lags were used. Results obtained showed that also the *61*-day patch was absorbed in an acceptable manner by the neural network: it implies that the data availability deriving from neighbouring stations might have improved results significantly. Obviously the larger the lag the larger the network error, anyway in the described case the Pearson's correlation coefficient *r* between measured versus reconstructed time series resulted *r=0.98*.

Spatial Analysis

Another successful application of *ANNs* concerned the spatial analysis of photochemical pol-

lutants using O_3 concentrations measured at five monitoring stations in the Provincia of Vicenza (approximately *40 km* NW of Venice - Italy): i) Bassano del Grappa [BG], ii) Montecchio Maggiore [MM], iii) Schio [SC], iv) Valdagno [VL], v) Western Vicenza [WV]. The data-base included ground measurements of NO_2 and O_3 concentrations, meteorological parameters and traffic flows (i.e. the hourly transiting vehicle number) (Pastore, 1998). For every station an *ANN* with *7* input neurons has been realised: they are the three meteorological parameters (solar radiation at two hours before, temperature and humidity) at the same station and the O_3 concentrations at the other four. These *non-local ANNs* replace then the local information (O_3 concentration at previous lags) with spatially distributed information (O_3 concentrations at different measuring stations).

The results obtained were better than those related to the isolated monitoring stations and the performed SA of non-local *ANNs* allowed the compiling of a *distance-ranking list* (Table 5) of the five different measuring sites. It can be used as an objective criterion of *connectivity* between two different sites of a geographic area, which allows the defining of boundaries of *homogenous aerological basins*. The linear regression relative to the five stations of the Provincia di Vicenza are shown in Figure 6: the Pearson's coefficients, respectively, were always over *r=0.97*.

CONCLUSIONS

In the present work it has been shown that some problems difficult to solve by traditional approaches can be addressed with *ANNs*: in particular *ANNs* can be used both to forecast air pollutant concentrations in urban areas as functions of some chemical variables and of local meteorological parameters, and to interpret their spatial behaviours, reaching also better results than in the case where only temporal information is available. Results obtained confirm the utility of similar neural architectures for predicting air quality pollutant concentrations both in industrial and in urban areas, as shown in the literature (Nunnari *et al.*, 1998; Garner and Dorling, 1999). *ANNs* are suitable for investigations working on large data sets and for problems in which the inputs and corresponding output values are known but the relationships between the inputs and the outputs are
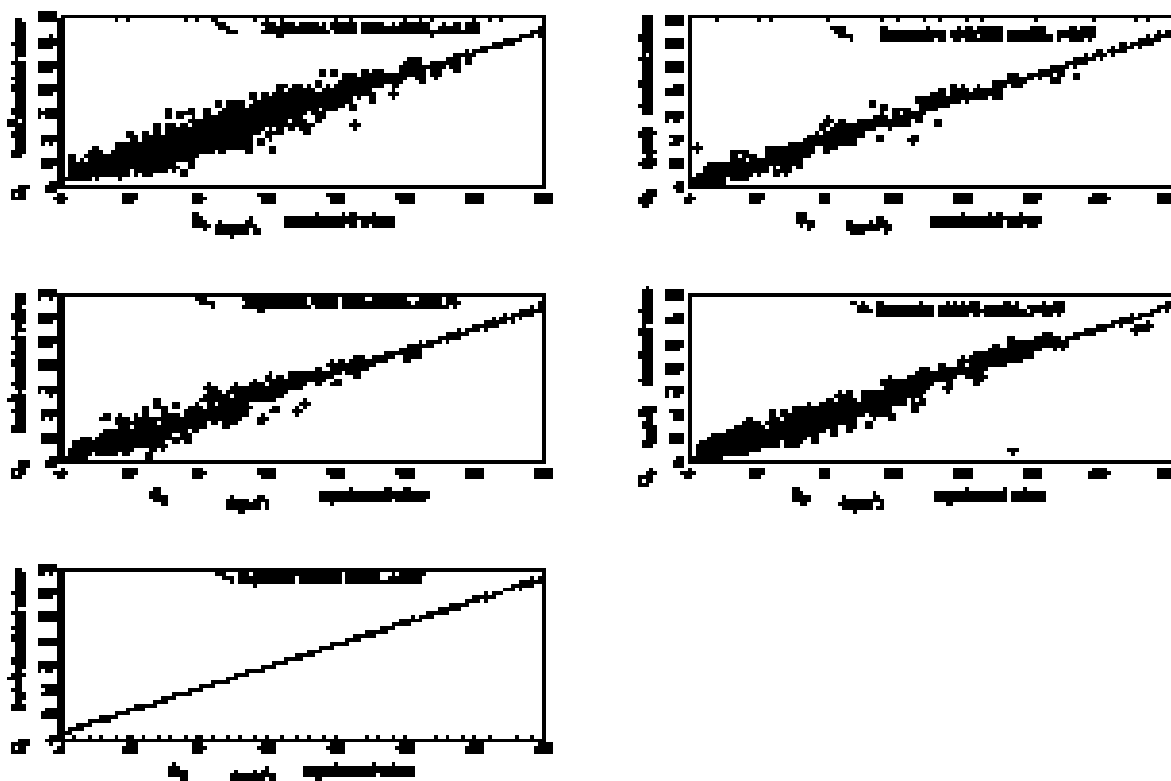
Figure 6. Linear regressions of O_3 measured versus O_3 simulated relative to the five monitoring stations of Bassano del Grappa, Montecchio Maggiore, Schio, Valdagno, Western Vicenza.

difficult to understand with usual analysis techniques. These conditions are commonly found in many air quality applications, owing to the relationships between physical and chemical processes in the environmental systems. *ANNs* do not eliminate the preparatory work on the data sets: indeed the data, before being processed, need a very careful statistical analysis to discover links between variables in order to avoid redundant patterns and improve accuracy.

The approach described will be a component for an air monitoring network implementation in the urban area around Venice designed to evaluate regulatory programs in improving ambient air quality. For example, *ANNs* could be utilised for controlling urban air quality pollution by regula-

Table 5. ANNs for O_3, performed considering the response to perturbations applied to the different input neurons, allowed the compiling of a distance-ranking list between the five different measuring sites. The value in each box denotes the sensitivity coefficients of each monitoring station with respect to the remaining four. In the first row and first column are listed the input sites (BG=Bassano del Grappa, MM=Montecchio Maggiore, SC=Schio, VL=Valdagno, WV=Western Vicenza)

	BG	*MM*	*SC*	*VL*	*WV*
Bassano del Grappa (BG)	–	1.36	1.52	0.52	0.86
Montecchio Maggiore (MM)	1.59	–	0.54	0.29	2.46
Schio (SC)	2.14	0.56	–	1.68	0.60
Valdagno (VL)	1.23	0.07	1.97	–	0.94
Western Vicenza (WV)	0.34	2.22	0.04	0.81	–

tion of traffic flows during severe weather conditions. *ANNs* were confirmed as a computational approach improving classical models in which the solution is learned from a set of examples. It is important to note that *ANNs* are pattern recognition diagnostic techniques, with some predictive skill, allowing assimilation of high quantities of data. Finally, the following considerations can be drawn: a) *ANNs* can provide considerable support in conducting quality controls on experimental data; b) results can be better when data from geographically distributed areas are available; c) in order to replace missing data or to assess outliers, it is advisable to conduct series training and forecasting on the same sets; d) the weights and SA provide useful insight into the most important forecasting variables and their relevant links; e) *ANNs* are particularly suitable whenever a large quantity of data is available (since networks train by example, the more data they have experienced, the better the network will perform) and whenever there is no simple solution by traditional technology.

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.

APPENDIX

The *ANNs* are product of the artificial intelligence, which miming the natural neurons 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 an high *ANNs* variety, often devoted to specific functions such as pattern recognition and ranking. The most suitable *ANNs* to interpret environmental processes (and, among them, the pollution one) are those known as *feed-forward back propagation* because they map inputs to outputs non-deterministically (like the natural process, they describe
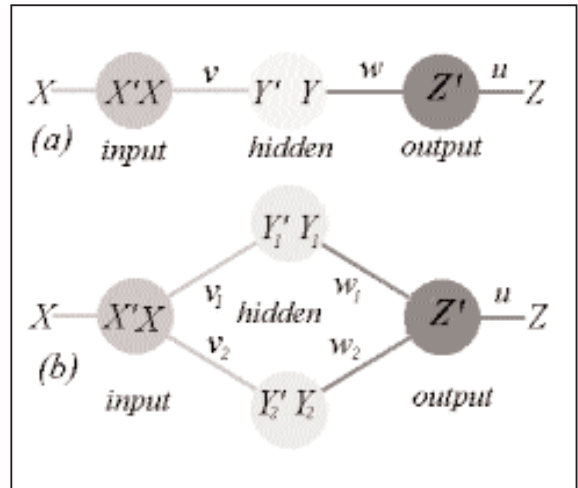

Figure 7. Two very simple neural architectures: a) *(1,1,1,1)* one input neurone, one hidden layer with just one node, one output neurone; b) *(1,2,1,1)* is a similar architecture but with two neurones in the first hidden layer and only one in the second hidden layer

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 by writing input/output analytical transformation of a *(1,1,1,1)* network (Figure 7a) with one input neuron, two hidden layers with one neuron, and one output neuron; and of another network *(1,2,1,1)* (Figure 7b) with one input neuron, two hidden layers with respectively two and one neurons, and one output neuron.

$$(Y' = vX, Y = \frac{1}{1 + e^{-vX}}, Z' = wY) \Rightarrow$$

$$\Rightarrow Z(x; v, w, u) = \frac{u}{(1 + e^{-wY})}$$

$$(Y' = v_1 X, 'Y'_2 = v_2 X,$$

$$Y_1 = \frac{1}{(1 + e^{-v_1 X})},$$

$$Y_2 = \frac{1}{1 + e^{-v_2 X}}),$$

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

where *0≤u≤1* was introduced in order to normalise the output Z. For example, in Figure 8 it is possible to compare the different outputs for dif-

ferent values of four weight parameters of the *(1,2,1,1)* network of Figure 7b. 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 consideration to be drawn regarding neural architectures:

i) the general non linearity of input-output connections, including as a special case the linear limit (the best *ARMAX* model can become a particular case of an *ANN*),

ii) the great importance of hidden neurons as active neurons capable of doing non linear transformations via transfer function,

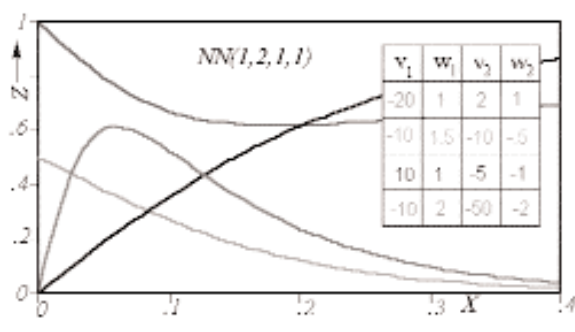iii) the functional complexity of a neuronal architecture is related to the number of weights.

Figure 8. 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. (1994), Using Neural Networks for Controlling Chaos, *Physical Review E*, **49**, 1225-1231.

Azzi, M., Johnson, G.M., Hyde, R. and Young, M. (1995), Prediction of NO_2 and O_3 concentrations for NOx plumes photochemically reacting in urban air, *Mathematical Computation Modeling*, **21**, 39.

Benvenuto, F. and Marani, A. (2000), Nowcasting of urban air pollutants by neural networks. *Nuovo Cimento,* **23C***,* 567-586.

Blamire, P.A. (1996), The influence of relative sample size in training artificial neural networks, *International Journal of Remote Sensing*, **17**, 223-230.

Boznar, M., Lesjak, M. and Mlakar, P. (1993), A Neural Network based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain, *Atmospheric Environment*, **27B**, 221.

Comrie, A.C. (1997), Comparing Neural Networks and regression models for ozone forecasting, *Journal of Air & Waste Management Association*, **47**, 653.

Côte, S. and Tatnall, A.R.L (1995), A Neural Network-based method for tracking features from satellite sensor images, *International Journal of Remote Sensing*, **16**, 3695-3701.

Devore, J.L. (1990), Probability and Statistics for Engineering and the Sciences, Duxbury Press.

Elizondo, D., Hoogenboom, G. and McClendon, R.W. (1994), Development of a Neural Network Model to Predict daily Solar Radiation, *Agricultural and Forest Meteorology*, **71**, 115-132.

Gardner, M.W. and Dorling, S.R. (1999), Neural network modelling and prediction of hourly NOx and NO_2 concentrations in urban air in London, *Atmospheric Environment*, **33**, 709-719.

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

Hornik, K., Stinchombe, M. and White, H. (1990), Universal approximation of an unknown mapping and its derivatives using Multilayer Feedforward Networks, *Neural Networks*, **3**, 551.

Liguori, F. (1996), Inquinamento di atmosfere urbane: interpretazione dei dati e nowcasting con reti neuronali, Graduation thesis in Environmental Sciences, University of Venice.

Marzaban, C. and Stumpf, G.J. (1996), A Neural Network for Tornado Prediction Based on Doppler Radar-Derived Attributes, *Journal of Applied Meteorology*, **35**, 617-626.

McClelland, J.L., Rumelhart, D.E. and the PDP Research Group (1986), Parallel Distributed Processing Explorations in the Microstructure of Cognition, MIT Press, Cambridge MA., I: Foundations.

Nunnari, G., Nucifora, A.F.M. and Randieri, C. (1998), The application of neural techniques to the modelling of time-series of atmospheric pollution data, *Ecological Modeling*, **111**, 187-205.

Pastore, A. (1998), Modelli neuronali dello smog fotochimico a scala di bacino, Graduation thesis in Environmental Sciences, University of Venice.

Romeo G., Mele, F. and Morelli, A. (1995), Neural Networks and Discrimination of Seismic Signals, *Computer & Geoscience*, **21**, 279-288.

Ruiz-Suârez, J.C., Mayora-Ibarra, O.A., Torres-Jimenez, J. and Ruiz-Suarez, L.G. (1995), Short-Term Ozone Forecasting by Artificial Neural Network, *Advances in Engineering Software*, **23**, 143-149.

Russel, A., Milford, J., Bergin, M.S., McBride, S.M., McNair, L., Yang, Y., Stockwell, W.R. and Croes, B. (1995), Urban Ozone Control and Atmospheric Reactivity of Organic Gases, *Science*, **269**, 491-495.

Soia, G. and Soja, A.M. (1999), Ozone indices based on simple meteorological parameters: potentials and limitations of regression and neural network models, *Atmospheric Environment*, **33**, 4299-4307.

SSNS©SNNS GROUP (1994), Stuttgart Neural Network Simulator, Institute for Parallel and Distributed High-Performance Systems (IPVR), University of Stuttgart, 70565 Stuttgart, Germany.

Tan, S.S. and Smeins, F.E. (1996), Predicting Grassland Community Changes with an Artificial Neural Network Model, *Ecological Modeling*, **84**, 91-97.

Wilks, D.S. (1995), Statistical Methods in the atmospheric sciences: an introduction, Academic Press, International Geophysics Series.

Willimott, C.J. (1982), Some comments on the evaluation of model performances, *Bulletin of American Meteorological Society*, **63**, 1309-1313.