Finzi, G. and G. Nunnari 2005. *Air Quality Forecast and Alarm Systems*. Chapter 16A of *AIR QUALITY MODELING - Theories, Methodologies, Computational Techniques, and Available Databases and Software. Vol. II – Advanced Topics* (P. Zannetti, Editor). Published by The EnviroComp Institute (http://www.envirocomp.org/) and the Air & Waste Management Association (http://www.awma.org/).

Chapter 16A

Air Quality Forecast and Alarm Systems

Giovanna Finzi [1] and Giuseppe Nunnari [2]

[1] *Dipartimento di Elettronica per l'Automazione, Università di Brescia, Brescia (Italy)*
finzi@ing.unibs.it
[2] *Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi, Università di Catania, Catania (Italy)*
gnunnari@dees.unict.it

Abstract: The following chapter reports a review of different stochastic and statistical modelling approaches, and the results obtained by their application to actual case studies both in urban and in industrial areas. The assessment of the results, in terms of daily and hourly forecast performance indexes and statistical indicators, is presented, compared and discussed. The structure of the chapter is the following. After a preliminary section summarizing the most significant stochastic and statistical modelling techniques (Section 1), some literature results are reported in Section 2. A more detailed mathematical description of the main techniques considered for air quality modelling is given in Section 3. Section 4 gives guidelines on how to build a model for air quality forecast and evaluate its performances. Some case studies concerning the modelling of tropospheric ozone concentrations, both in urban and industrial areas, are given in Section 5. In Section 6 the application of the selected techniques to the implementation of an operational decision support system (DSS) is described. Also, the performance of the system for two different metropolitan areas in the Northern part of Italy (namely Brescia and Milan) is evaluated. Finally, a short survey of the available software packages to implement the modelling techniques described is given in Appendix.

Key Words: stochastic models, air quality forecast, tropospheric ozone, decision support system.

1 Introduction

Stochastic and statistical models are based on semi-empirical relationships between available data and measurements. The distinction between stochastic and statistical models is not always clear in literature since these models share several common features. Usually models are referred to as stochastic when the time variability is taken into account explicitly, and they are referred to as statistical if based on the use of some stationary statistical approaches (such as the clustering techniques or the Bayesian inference mechanism).

Models taking uncertainty into account, instead of deterministic ones, do not aim to describe the level of pollution as a phenomenology-driven, cause-effect problem; instead, their identification is based on the direct use of air quality measurements. In particular, statistical models are generally useful when the information available from measured concentration trends is more relevant than the one obtained from the deterministic analyses. The use of statistical models may also be encouraged by the presence of a significant amount of data, recorded by several monitoring networks, in the territory of interest.

Some general information about basic statistical and stochastic models for air pollution data is given in Gilbert (1987) and Zannetti (1990). More recent techniques are described by Finzi et al. (2001) and Jorquera et al. (2004).

In the development of such models, measurements of pollutants and related meteorological variables have been considered as time series and analysed by means of a wide variety of methods, including:

- spectral analysis
- component approach (trend + seasonal + residuals)
- regression analysis
- trend analysis
- clustering analysis
- principal component analysis
- hybrid models
- *black-box* ARIMA models
- *grey-box* (non linear and non stationary) models
- Bayesian models

More recently other statistical approaches have been published generically referenced as AI (*Artificial Intelligence*) or *Soft-computing* techniques, including:

- *neural models*
- *fuzzy models*
- *neuro-fuzzy models*

Statistical modelling approaches can be used in a *black box* mode (i.e., pollutant concentration time series analysed without any phenomenological information),

to evaluate their intrinsic statistical variations without attempting any physical explanation. Otherwise, they can be used in a *grey box* mode, in which other information (such as meteorological or emission variables, seasonal cycles) may be taken into account explicitly or implicitly, although always in an uncertain environment.

In the following each model class will be shortly referenced, pointing out its main features with respect to a possible use as air pollutant concentrations predictor.

- **Spectral analysis** (Jenkins and Watts, 1968) allows the identification of cycles in meteorological and air quality time series measurements. Early applications of spectral analysis of SO_2 concentration were carried out by Tilley and McBean (1973) and Trivikrama et al. (1976). They showed the existence of semi-diurnal, diurnal and three-and-half-day period oscillations in SO_2 and wind time series recorded in the study area (Northern USA). Semidiurnal cycles were interpreted in terms of local phenomena, such as sea breezes. Longer periods seem to be induced by synoptic weather variations. More recently, spectral analysis has been applied by Schlink et al. (1997) for analysing SO_2 data recorded at Leipzig in the years 1980-1993.

- **Component models** are based on a spectral decomposition of time series into a trend component, a periodic component, and a residual component. Each component shows a peculiar behaviour; in particular, trend one is very smooth and slowly varying with fixed periods for the cyclic components. Each component is predicted on the basis of its features and the final forecast is a combination of all of them (Schlink et al., 1997). In spectral decomposition, low-pass and band-pass filters are used to quantify the components. Young et al. (1991) gave very convenient and flexible filtering algorithms based on the Kalman filter.

- **Regression analysis** is a particular type of stationary multiple input time-series analysis, in which meteorological measurements are statistically related to air quality concentrations by means of a linear model.

- **Trend analysis** allows evidencing and eliminating any *a priori* significant trend or seasonal variation in time series (see contribution by Buishand et al., 1988).

- **Clustering analysis** allows finding a set of *a priori* unknown data categories (or clusters), based on available set elements and observations. Formally, a clustering process orders observations in clusters (i.e., subsets with high degree of association among members of the same group and low among members of different groups). Clustering analysis for modelling pollution data has been considered for instance by Sanchez et al. (1990) and Huang (1992).

- **Principal component analysis** allows reduction of multivariate data through transformation of the original variables into a new set of uncorrelated ones, progressively accounting for decreasing proportions of explained variance in the data. Aim of this methodology is to reduce the dimensionality of the model. The new variables (principal components), are defined as linear functions of the original ones. If a limited number of components account for a large percentage of explained variance in the observations, they can be used to simplify the subsequent analysis. Principal components have been used by many authors (e.g., Lins, 1987).

- **Hybrid models**, mainly the ones based on Kalman filters, have been frequently used for updating the forecast capabilities of a deterministic predictor based on the availability of real-time measurement of pollutants. For instance, the application proposed by Melli et al. (1981) suggested this approach for real-time control of SO_2 emissions in the industrial area of Porto Marghera, which is located in the Venetian lagoon region (Italy).

- ***Black-box* ARIMA and ARMAX models** (Box and Jenkins, 1970, 1976; Box et al., 1994) have been considered as one of the most cost-effective approaches for time series analysis. Many authors have been inspired to apply this technique in developing pollutant forecast models for SO_2 time series as well as for other pollutants such as ozone, NO_x, etc. (see Finzi et al., 1983).

- ***Grey-box* models** are extended ARMAX (Auto Regressive Moving Average with eXogenous inputs) models, which allow the user to take into account the non-stationarity of the process through parameters depending on time-varying classes in order to treat the complexity of air pollution dynamics (a first example of application to SO_2 urban pollution forecast is illustrated in Finzi and Tebaldi, 1982).

- The **neural approach** consists of using artificial neural networks (ANN) to identify air quality prediction models. Neural networks are virtually parallel computational architecture based on the emulation of the human brain. Applications of similar computational architectures to the prediction of SO_2 concentration have been described by Boznar et al. (1993), Arena et al. (1996), Finzi et al. (1998), and Nunnari et al. (2001).

- The **fuzzy approach** aims to describe the behaviour of dynamic systems by using linguistic representation (i.e., the system model is represented by a set of rules, the rule base, in the "if…then" form). A similar methodology has been proposed to implement NARX (Non-linear Auto Regressive with eXogenous inputs) models of pollutant time series (e.g., Nunnari et al., 1998; Nunnari, 2000), allowing an easily understandable way to model complex air quality phenomena.

- In **neuro-fuzzy** systems, neural networks are used to tune the *membership functions* of the fuzzy system and to automatically extract *fuzzy rules* from numerical data (see Finzi and Volta, 2000).

- **Bayesian models** are recent techniques applied to forecast critical pollution episodes. Some contributions referring to this topic are reported by Maffeis (1999), Cossentino et al. (2003), and Nunnari and Cannavò (2004). The peculiarity of this approach is that the prediction problem is formalised in terms of a Bayesian Network (BN). Deep *a priori* knowledge about the model structure is not required to build a BN; instead, heuristic knowledge that is usually available can easily be taken into account. Bayesian models can work even in the presence of partially missing input information. When some input values are missing, they can simply be neglected or replaced by surrogates such as the corresponding probability distributions. Finally, BNs can operate in the so-called *diagnostic mode* in order to infer the causes of poor air quality.

2 Some Literature Results

Some of the most significant results, which appeared in the literature referred above, are shortly reported in the following: they mainly concern the application of statistical and stochastic models to predict critical pollutant concentrations both in industrial and urban areas.

Regression analysis was employed by Bringfelt (1971) to assess significant variables driving SO_2 concentrations in Central Stockholm. One of the aims of the study was to set up a warning system for SO_2 episodes based on the forecast of atmospheric mixing layer height and wind speed. SO_2 mean daily concentrations, averaged over four monitoring sites in Central Stockholm, were compared with temperature, wind speed and mixing height by means of a multiple regression analyses for the winter periods of 1967-1969. Different methods to extract the best meteorological predictors from routine weather data were compared. As temperature predictor, the difference of mean daily temperature monitored at the local airport was used when below a threshold of 25°C. The mixing height was estimated on the basis of night and day radio sounding measures (01 and 13 hours) and the minimum temperature in the city. The multiple correlation coefficient came out quite high (0.84). The daily SO_2 levels were predicted with a standard error of about 25%.

A **Kalman filtering** approach was proposed by Melli et al. (1981) for modelling SO_2 emissions in the Venetian Lagoon industrial area (Northern Italy) with the aim of implementing a real time system for emission control; the authors considered real-time emission control as an air quality strategy alternative to permanent emission reduction. They proposed an emission control scheme characterised by the following steps:

- Collect current concentration and meteorological measurements from monitoring networks.
- Forecast future values of relevant local meteorological variables, by means of simple stochastic mathematical predictors.
- Predict future concentrations on the basis of information from current concentration values, forecast meteorology and scheduled emissions. The predictor model is based on a complex forecast algorithm (Kalman predictor) derived from the "stochastic version" of the numerical solution of the advection-diffusion partial differential equation.
- If future concentrations exceed some reference level, the scheduled emissions are reduced. The assumed control policy consists of mixing fuel with a cleaner one, under the constraint of maintaining the production scheduled by each polluting plant.

The results of the case study are supplied as cost-effectiveness curves (cost versus effectiveness of the control action). The authors showed that real-time emission control was economically cheaper and technically possible.

Finzi and Tebaldi (1982) applied a non-stationary and non-linear **grey–box** auto regressive model with exogenous inputs to predict daily SO_2 average concentration in Milan urban area (Northern Italy). This city is now using mostly methane or low sulphur fuels for domestic heating, but at the time when the study was performed (late 1970s), there were serious problems with SO_2 pollution due to low quality fuels widely used for domestic heating. Data analysis carried out on historical time series allowed the authors to evidence that the urban pollution level was particularly high during cold season with anti-cyclonic synoptic conditions, low ambient temperature, and low wind. So they implemented a forecast model for daily SO_2 DAP (Dosage Area Product) computed over Milan urban area. In particular, two different meteorological categories were defined respectively, corresponding to cyclonic and anti-cyclonic synoptic regimes over Northern Italy; the local wind velocity and temperature were considered as non-linear exogenous inputs. The model was validated comparing predicted and measured data during the winter season of 1975-1976, with fairly good agreement mainly during severe critical episodes.

A more complex *grey-box* model was further applied in a study concerning SO_2 air pollution forecast in Madrid metropolitan area (Hernandez et al., 1983; Finzi et al., 1983). In this study the role of meteorological variables in statistical pollution forecasting models was highlighted. A comparison was also performed between *black-box* and *grey-box* models, showing how the use of basic physical knowledge of the phenomenology allows a higher cleverness in episodes prediction.

Another statistical model to forecast SO_2 concentrations in the surroundings of a thermal power plant was studied by Brusasca and Finzi (1986) with the purpose of emission real-time control. The plant taken as a case study, with a nominal

power of 1365 MW, is located at Turbigo (in the Po Valley near the Alps Mountains, the Ticino River Natural Park, and not far from the city of Milan) and managed by ENEL (the Italian National Electricity Board). The authors implemented a **Cyclostationary** Auto Regressive model (referred also as ARCX in the following Sections) to forecast half-hourly and daily mean concentrations of SO_2 at some hours in advance during the day. Thanks to a recursive computation scheme, the prediction of half-hourly average concentrations was performed and updated starting from the morning data considered for the study. They recorded at 5 stations around the plant (in the range of 5 Km) during the cold seasons of 1982/83, 1983/84 and 1984/85. The real-time daily SO_2 forecast model was compared with a more trivial one, the so-called persistent model; the results evidenced that the stochastic predictor performed much better in terms of a statistical analysis of the forecast errors, and allowed the power plant managers to prevent critical episodes and meet law standards.

Boznar et al. (1993) carried out one of the first studies that appeared in the literature concerning a comparison between a deterministic and a **neural network** based approach. They modelled SO_2 concentrations due to Slovenian thermal power plant emissions and pointed out the difficulties in using deterministic models when the terrain is not flat. Moreover, they stressed the fact that also the simplest stationary Gaussian model needs several input parameters to work properly. Since it is often difficult to get inputs and a reliable parameterisation in real time, deterministic models may give unrealistic estimates, mostly based on the assumption of a stationary emission and meteorological scenario. Alternatively, they proposed a **neural network** based approach for short-term prediction.

Artificial neural networks have also been taken into consideration by Arena et al. (1996) to set up a model of SO_2 time series monitored in an industrial area very close to Siracusa (Sicily). In particular, a short-term prediction (six hours ahead) of the SO_2 pollutant mean value has been performed. A neural architecture was implemented, based essentially on a *Multilayer Perceptron* devoted to predict alarming situations and to estimate the mean pollutant value. The results showed that **neural network** based strategies for short-term prediction of SO_2 levels are promising.

Statistical methods have also been developed by Schlink et al. (1997) to set up an advanced smog warning system in central Leipzig (Germany) by modelling winter time SO_2 concentrations recorded during 1980-1993. The authors essentially use a **recursive Kalman algorithm**, based on a preliminary spectral analysis of the SO_2 concentration time series; the smog episodes with low frequencies and time-dependent power spectra were well represented by the trend component alone. This component was therefore investigated in the phase space, where it exhibited a typical trajectory feature. One data subset was used to identify the model parameters and another was left for validation; the results were still not satisfactory. So a modified method was proposed to extrapolate the time-

dependent spectrum of the trend component, namely a local harmonic approximation. This method was tested and compared with simple linear extrapolation. It provided a generalization, producing closer correspondence between predicted and observed concentration values.

Another study concerning **neural network** models, applied to SO_2 pollution, has been reported by Reich et al. (1999). They address the problem of the apportionment of a small number of SO_2 sources from a data set of ambient concentrations. A three layer feed-forward artificial **neural network** trained with a back-propagation algorithm was employed. A subset of hourly meteorological conditions and measured concentrations constituted the input patterns to the network, which was mainly designed to identify relevant emission parameters of unknown sources as outputs. The remaining model data were degraded by adding noise to some meteorological parameters and the effectiveness of the method was tested. The model was applied to a realistic case where 24 h SO_2 concentrations were previously measured. Some of the limitations of the artificial **neural network** approach and its capabilities are discussed in this paper.

The role of statistical and stochastic modelling techniques as a practical tool for air quality prediction and forecasting has been recognized by the European Union in the framework of the Fifth Framework Program funded projects devoted to develop and test a variety of advanced statistically based modelling techniques. Among these, it is worth to mention the APPETISE project (*Air pollution episodes: modelling tools for improved smog management*, http://www.uea.ac.uk/env/appetise/). This work has been carried out over a period of two years (Greig et al, 2000) by a consortium of 9 institutions from 5 different European Countries. The project has focused essentially on four key pollutants: nitrogen oxides, particulates, ground level ozone and sulphur dioxide.

A significant number of statistical modelling techniques were applied to model pollutant time series of a rich database representing different meteorological and emission conditions throughout Europe. The performances of the considered techniques were inter-compared rigorously. The variety of inter-compared techniques together with the different locations of the area considered, and different kind of sites (i.e., urban, suburban, rural, industrial) and targets make the results of this inter-comparison exercise more general and interesting. The main results of the project have been published by Schlink et al. (2003) for ground level O_3, by Kukkonen et al. (2003) for NO_2 and PM_{10}, and by Nunnari et al. (2004) for SO_2.

In Section 5, *grey-box* models, *neural*, *fuzzy* and *neuro-fuzzy* networks will be taken into consideration and applied to other recent real case studies, and their performance will be compared as real time predictors for atmospheric urban pollutant concentrations. As introduced above, all the approaches are based on the analysis of time series of pollutant concentration measures recorded by air quality control networks. The first methodology requires a minimum physical

understanding of the phenomenon in order to drive the model to the description of the non-linearity and non-stationarity of the process by means of a limited number of parameters, while the other ones can be applied by a non-expert user by means of largely automatic procedures (*black-box* approach).

3 Time Series Modelling

Techniques for modelling time series can be roughly classified as linear and non-linear.

3.1 Linear Techniques

Linear techniques represent the simplest way to model statistical time series. Despite the fact that the largest part of natural phenomena are non-linear (e.g., Kantz and Schreiber, 1997), several ideas can be generalised from the theory of linear modelling techniques.

Until recently, linear multi-variate methods have been considered to be one of the most cost-effective approaches for time series analysis. Many authors have been inspired to apply these techniques, after some appropriate modifications, in developing pollutant forecasting models. These techniques have been applied to modelling SO_2 time series as well as other pollutants such as O_3, NO_x etc. The original Box-Jenkins approach (Box and Jenkins, 1976; Box et al., 1994) has been adapted by some authors in order to treat the complexity of air pollution data such as non-stationarity (e.g., Finzi et al., 1998).

The basic structure of linear techniques is outlined as follows. Let us denote a discrete time random process by *{y(t)}* and a discrete purely random process with zero mean and variance σ^2 by *{e(t)}*. A process *{y(t)}* is said to be a moving average process of order q, and indicated as MA(q) if

$$y(t) = e(t) + \gamma_1 e(t-1) + ... + \gamma_q e(t-q) = (1 + \gamma_1 B + ... + \gamma_q B^q) e(t) \quad (1)$$

where γ_i are constants and B is the backward shift operator. Similarly a process is said to be an autoregressive process of order p, and indicated as AR(p) if

$$y(t) = \alpha_1 y(t-1) + ... + \alpha_p y(t-p) + e(t) \quad (2)$$

where α_i are constants and $|\alpha_1| < 1$.

A very simple AR model is the so-called persistent model (i.e., *y(t) = y(t-1)*, tomorrow equals today) that is often considered as a reference model during inter-comparison exercises.

By combining the AR and the MA structures, it is possible to obtain an ARMA process of order (p,q) as follows:

$$y(t) = \alpha_1 y(t-1) + ... + \alpha_p y(t-p) + e(t) + \gamma_1 e(t-1) + ... + \gamma_q e(t-q) \qquad (3)$$

Such structure is sometimes useful to model linear stationary univariate time series. To handle linear non-stationary time series, the ARMA model can be appropriately extended in order to obtain the ARIMA model (Box and Jenkins, 1970). An ARIMA(p, d, q) model is a particular ARMA model where the original time series *{y(t)}* is substituted by the d-times differenced series $\{B^d y(t)\}$.

Multivariate time series (i.e., time series in which one variable is related to others), referred to as exogenous variables (or inputs), can be modelled by using the ARMAX models. As an example, if we want to model the ozone time series recorded at a given point, we can try to use as exogenous variables (the solar radiation and the concentration of Nitrogen Dioxide, NO_2) since it is known that these variables play a role in the ozone cycle. The structure of the ARMAX process is the following:

$$\begin{aligned} y(t) = \alpha_1 y(t-1) + ... + \alpha_p y(t-p) + \beta_1 u(t-1) + ... + \beta_r u(t-r) + \\ + e(t) + \gamma_1 e(t-1) + \cdots + \gamma_q e(t-q) \end{aligned} \qquad (4)$$

Equation (4) has been considered in the presence of an individual exogenous variable *u(t)*, also referred to as the input variable, but the generalization to the case of a generic number of exogenous variables is trivial. Furthermore, the above-mentioned structure can be generalised to the case when the stochastic variables *y(t)* and *u(t)* are vectors. ARMAX models lead to a simpler but quite useful representation. The ARX model is given by:

$$y(t) = \alpha_1 y(t-1) + ... + \alpha_p y(t-p) + \beta_1 u(t-1) + ... + \beta_r u(t-r) + e(t) \qquad (5)$$

Identification of parameters of ARX models (i.e., the determination of constants α_i and β_i from experimental data), can be obtained by the standard Least Square (LS) method. Identification of ARMAX models can be performed instead using the Generalized Least Square (GLS) approach (e.g., Soderstrom and Stoica, 1989).

Finzi et al. (1982) proposed a particular kind of ARMAX models referred to as *cyclo-stationary* or *grey-box* ARMAX model that are able to deal with particular (cyclic) non-stationary phenomena that often affect pollution time series. The structure of a *cyclo-stationary* ARMAX model is as follows:

$$y(t) = \sum_{i=1}^{p} \alpha_i[s(t)] \cdot y(t-i) + \sum_{j=1}^{r} \beta_j[s(t)] \cdot u_j(t-k_j) + \sum_{h=1}^{q} \gamma_h[s(t)] \cdot e(t-h) + e(t) \quad (6)$$

where *s(t)* is a properly defined category at time *t* and k_j is the lag time. These kinds of models can be useful when the process to be modelled is affected by some underlying periodic (daily, weekly, seasonal or yearly) components.

3.2 Non-Linear Techniques

Linear methods allow interpreting all the regular structure in a data set such as dominant frequencies. However, the linear paradigm, which can be roughly stated as "small causes lead to small effects", is not always true for natural phenomena. This can be explained by bearing in mind that linear differential equations can only lead to exponentially growing or periodically oscillating solutions. This means that all irregular behaviours of a given system must be attributed to some random external input. However, it is known from the *System Theory* that random inputs are not the only ones responsible for irregular behaviours of the system output; nonlinearities or chaos can produce very irregular data even with purely deterministic equation. So, it is better to try to explain irregularities in a given time series with both the presence of random inputs and nonlinearities (Kantz and Schreiber, 1997). The literature about non-linear modelling techniques is very rich and we will deal with the most widely considered ones here.

3.2.1 NARX Models

A general way to represent nonlinear systems is the NARX (Non-linear Auto Regressive with eXogenous inputs) representation, which can be considered as a generalization of the ARX model:

$$\begin{aligned} y(t) = f(y(t-1), \cdots, y(t-p), u_1(t-1), \cdots, u_1(t-r_1), u_2(t-1), \cdots, u_2(t-r_2), \\ \cdots, u_n(t-1), \cdots, u_n(t-r_n)) + e(t) = f(X) + e(t) \end{aligned} \quad (7)$$

Here *f* is an unknown non-linear function, *y(t)* is the system output, $u_1, u_2, \ldots, u_n$ are related input variables (e.g., meteorological and/or emission variables), and *e(t)* is a random term. In expression (7) the variable *X* is expressed as:

$$\begin{aligned} X^T(t) = [y(t), y(t-1), \cdots y(t-n_y+1), u_1(t), \cdots, \\ u_1(t-n_1+1), \cdots, u_q(t), \cdots u_q(t-n_1+1)] \quad \in R^d \end{aligned} \quad (8)$$

in order to indicate the data vector, also referred to as the model input pattern.

Several of the most powerful time series modern techniques - such as the Multi-layer Perceptron (MLP) artificial **neural networks**, the *Fuzzy* and *Neuro-Fuzzy*

techniques - can be considered for approximating the unknown function *f*, given an appropriate set of measured data, as explained below.

3.2.1.1 The MLP Modelling Technique

Multi-layer Perceptron (MLP) Neural Networks are parallel computational architectures where their structure is based on the emulation of the human brain. If suitably "trained" using a set of examples, they can "learn"; that is, they can extract the link between the input data and the corresponding output data (Lippmann, 1987). MLPs can thus be used to solve a number of problems of classification, and more generally, *black-box* identification, in which *a priori* knowledge of the model is not needed (Chen and Billings, 1992). Moreover, operations are relatively simple and can be performed quite systematically; the learning phase is entrusted to special optimisation algorithms such as the *back-propagation* algorithm (Rumelhart et al., 1986). Properties of MLPs in non-linear system identification are described in Sioberg et al. (1994). From the mathematical point of view, MLPs perform automatic search of models in the class of NARX structure. It has to be stressed that such a modelling approach can be equally applicable for both scalar and vector sequences. The use of this non-linear auto-regression can be justified as follows. For a wide class of deterministic systems, it can be assumed a *diffeomorfism* (i.e., a one-to-one differential mapping) between a finite window of the time series [*y(t-1), ... y(t-p), u(t-1), ... u(t-n)*] and the underlying state of the dynamic system which gives rise to the time series. This implies the existence of the non-linear auto-regression of the form (7). The MLP neural network thus forms an approximation of the ideal function *f(·)*. Furthermore, it has been shown (Cybenko, 1989) that a feed-forward **neural network**, with an arbitrary number of neurons in the hidden layer, can approximate any uniformly continuous function with an arbitrary degree of accuracy.

The internal model representation by using a *Multi-layer Perceptron* structure is based on the expression:

$$N_i(U) = \Gamma(W_i \cdot U + B_i) \tag{9}$$

which is used in a recursive frame. The meaning of the symbols is as follows:

- U = output of the $(i-1)^{th}$ layer
- W_i = weight matrix associated to the i^{th} layer
- B_i = bias vector corresponding to the i^{th} layer
- Γ = an appropriate activation function

Formally, the output of the **neural network** can be expressed as follows:

$$Y = F(U) = \Gamma(W_n\,(\Gamma(W_{n-1}(\ldots\,\Gamma(W_1 \cdot U)))) \tag{10}$$

In the expression above, the biases have not been reported for sake of simplicity. Since a MLP with one hidden layer can solve the same class of problems that can be solved by using MLP with more than one hidden layer, below we will refer to MLP with one hidden layer only. Under this hypothesis, expression (10) assumes the following simpler form:

$$Y = F(U) = \Gamma(W_2\,(\Gamma(W_1 \cdot U + B_1) + B_2) \tag{11}$$

MLPs have been considered in the literature for air pollution time series modelling by several authors such as Boznar et al. (1993), Arena et al. (1996), Nunnari et al. (1998), Gardner and Dorling (1998, 1999), Schlink et al. (2003), Kukkonen et al. (2003), and Nunnari et al. (2004).

A problem with using MLP trained by the traditional back-propagation algorithm is, apart from the often excessively low speed of convergence (e.g., Sarkar, 1995), the possibility of obtaining configurations of the weights corresponding to a local minimum. The literature reports numerous variations of this algorithm aimed at improving the performance (see Gori and Tesi, 1992). Another problem using MLP is the choice of the most appropriate number of neuron in the hidden layers. A small number of hidden neurons yields low accuracy models; on the other hand, a large number produces the problem of over fitting, which means poor generalization capabilities of the model. Since there are no *a priori* formulas to compute the best number of hidden neurons (Barron, 1993), this is usually done by a trial and error approach by searching for a trade-off between accuracy and generalization capabilities of new input data. In order to correctly evaluate the generalisation capabilities of MLP trained by using the traditional back-propagation (BP) algorithm, the so-called *Early Stopping* approach (Sjoberg and Ljung, 1995) can be considered.

Despite the fact that BP algorithm is the most widely considered **neural networks** for training MLP, others algorithms are available such as the conjugate gradient optimisation approach. Moreover, the traditional Sum of Squares Error (SSE) that is the standard cost function minimised by the BP algorithm is sometimes replaced by other kinds of cost function. As an example, Dorling et al. (2003) proposed to use the *maximum likelihood* cost functions to model air quality data.

3.2.1.2 The Fuzzy Modelling Technique

The fuzzy set theory represents a different approach to dealing with uncertainty than the traditional probabilistic and statistical methods. The essential feature of the fuzzy logic is the concept of membership function, which ranges between 0 and 1 and represents the degree of membership of an individual element to a given set, referred to as a fuzzy set. From the seminal paper by Zadeh (1965), a lot of work has been carried out in the field of fuzzy logic, and it is beyond the purpose of this chapter to deal with the huge amount of theoretical and practical aspects of this theory. Here, we only deal with the problem of approximating a NARX model of the form (7) by using a fuzzy rule base of the form:

$$
\begin{aligned}
R_i : \; & if \; y(t) \; is \; A_{i,1} \; \; and \; y(t-1) \; is \; A_{i,2} \; and, \ldots, y(t-n_y+1) \; is \; A_{i,ny} \quad and \\
& u_1(t) \; is \; A_{i,ny+1} \; \; and \; u_1(t) \; is \; A_{i,ny+2} \; and, \ldots, u_1(t) \; is \; A_{i,ny+n1} \qquad and \\
& \ldots \\
& u_q(t) \; is \; A_{i,ny+n1+\ldots} \; \; and \; u_q(t) \; is \; A_{i,ny+2} \; and, \ldots, u_q(t) \; is \; A_{i,p} \\
& \qquad\qquad\qquad\qquad\qquad\qquad y(t+1) \; is \; B_i \qquad (i = 1, \ldots k)
\end{aligned}
\tag{12}
$$

where $A_{i,j}$ *(j= 1,...p)* and B_i *(i=1,...K)* are fuzzy sets. Particularly in the case considered below, the consequent fuzzy sets B_i are assumed to be *singletons* (i.e., real numbers).

The fuzzy modelling approach consists of the following steps:

1. Positioning of the membership functions $A_{i,j}$ in their respective universe of discourse. This step is based on the determination of the matrix centres of the input data clusters by using one of the clustering algorithms described in the subsequent Section 3.2.6, or it can be simply performed on a trial and error basis. The shape of the membership function must also be selected among a large variety (e.g., trapezoidal, Gaussian, etc.)
2. Generation of all possible rules according with the input patterns available
3. Pruning of the unnecessary rules; this step is based on approximating the input patterns with the closest cluster centre
4. Determination of the consequent part of each rule; this is done by using an appropriate optimisation approach. It can be demonstrated that for rules of the form (12), the consequent part of each rule can be also obtained by using the least square algorithm (e.g., Nunnari, 2000)
5. Further pruning phase (last step is optional) according to a statistical criterion which takes into account the number of activation of each rule

More details about this algorithm for modelling air pollution time series can be found in (Nunnari, 2000). A quite similar approach, based on the use of the *Fuzzy* C-means clustering algorithm, is described in Section 3.2.6.

3.2.1.3 The *Neuro-Fuzzy* Technique

In *neuro-fuzzy* systems, **neural networks** are used to tune the *membership functions* of the fuzzy system and to automatically extract *fuzzy rules* from numerical data (Shing and Jang, 1993). The internal structure of a *neuro-fuzzy* network is illustrated in Figure 1. The nodes of the first layer represent the *crisp* inputs. The activation functions of the second layer nodes are Gaussian and act as membership functions. Each neuron of the third layer acts as a *rule node* so that this layer provides the fuzzy rule base. The output of this layer determines the activation level at the output memberships. As ordinary neural nets, the *neuro-fuzzy* one learns on a training data set, tuning membership functions and rules by means of a *back-propagation* algorithm.

When x_i is the i^{th} node in layer A, O_j^L is the j^{th} output of generic layer L, and W_{ij}^L is the weight of the link between j^{th} neuron at layer $L+1$ and i^{th} neuron at layer L, each layer output can be described as follows:

Layer B: $$O_j^B = \left(1 + \exp(-\frac{x_i - w_{ij}^{Ap}}{W_{ij}^{Al}}) \right)^{-1}$$

Layer C: $$O_j^C = \min_i (w_{ij}^B \cdot O_j^B)$$

Layer D: $$O_j^D = \frac{\sum_i w_{ij}^C \cdot O_i^C}{\sum_i O_i^C}$$

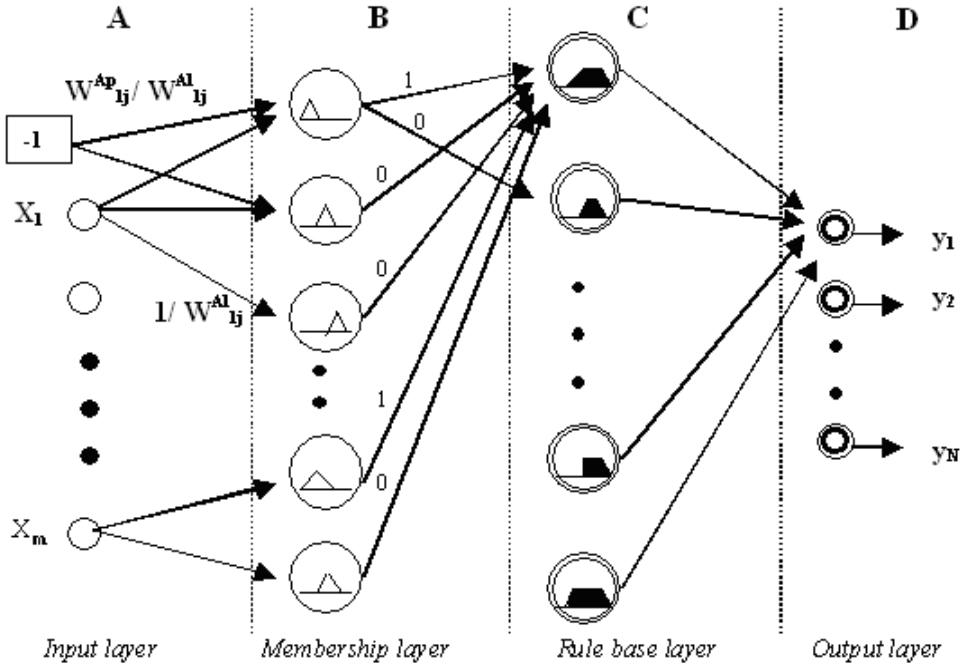

Figure 1. ***Neuro-fuzzy*** **network architecture.**

3.2.1.4 The Wavelet based Modelling Technique

Wavelet functions have been reported in the literature due to their capability of modelling transient phenomena occurring in particular geophysical time series (e.g., earthquakes). Some insights on the use of wavelets for system identification improvements can be found in Zhang (1997). A strategy, based on wavelets for modelling air pollution data, was proposed by Nunnari (2003). The mentioned approach can be summarised as follows. Let $\psi(t)$ be a basic wavelet function and let $s\ (s \neq 0)$ and u be real numbers; the family of wavelets corresponding to $\psi(t)$ is

$$\Psi_{s,u}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-u}{s}\right) \tag{13}$$

Here s represents the dilatation and u gives the translation. With reference to expressions (7) and (8), let us introduce the scalar quantities $t_j\ (j=1,\ldots,\ M)$ as a map between the vectorial argument of expression (7) and the scalar argument t of the generic *j-th* wavelet function (13):

$$t_j = \frac{\left[A_j \otimes (X(t) - U_j)\right] \cdot \left[A_j \otimes (X(t) - U_j)\right]^T}{s_j} \tag{14}$$

where A_j and U_j are appropriate vectors of unknown parameters. The WAG approach consists of approximating *y(t+1)* in expression (7) as:

$$y(t+1) = \sum_{j=1}^{M} wj \cdot \Psi_j(t_j) \tag{15}$$

In the WAG approach, the number of approximating wavelet functions M is obtained by a *trial and error* iterative procedure, a trade-off between accuracy and generalisation capabilities, while the remaining model parameters, namely A_j, U_j, s_j, and w_j *(j=1, M)* are searched by using a genetic algorithm (GA) optimisation approach (Holland, 1975). The reason for using GAs is that they are capable of finding the global minimum of a function with many variables, overcoming the limitation of typical gradient-based optimisation techniques (Goldberg, 1989). Even though by using a search algorithm, such as GAs, the modeller can define very complicated cost functions, good results can still be obtained by the traditional minimisation of the sum of error squares.

3.2.2 The Generalized Additive Modelling Technique

The Generalised Additive Model (GAM) uses smoothing techniques, such as locally weighted regression, to identify and represent possible non-linear relationships between the output and the model inputs (i.e., the explanatory variables).

This approach represents an alternative to considering polynomial terms or searching for the appropriate transformations of both output and input variables. By using these models, the link function of the expected output value variable is modelled as a sum of a number of smooth functions of the explanatory variables as expressed by (16) rather than in terms of explanatory variables themselves:

$$y(t+1) = f_1(u_1(t)) + f_2(u_2(t)) + \ldots + f_k(u_k(t)) + e(t) \tag{16}$$

GAM is discussed in Breiman and Friedman (1985), Cleveland (1979), Davis et al. (1998), Davis and Speckman (1999), Hastie and Tibshirani (1986, 1987), and Wood (2000). Applying the GAM technique, the non-linear functions f_1 in equation (16) are specified in a nonparametric fashion by means of scatterplot smoothers (i.e., weighted average of neighbouring observations). Cubic or fourth order splines functions were applied for instance by Schlink et al. (2003) to model tropospheric ozone and SO_2 time series (Nunnari et al., 2004)

3.2.3 Local Prediction in Phase Space

Local Prediction in Phase Space is a method for non-linear time series analysis, which is based on the paradigm of deterministic chaos (Kaplan and Glass, 1995;

Abarbanel, 1996; Kantz and Schreiber, 1997). The chaos theory offers completely new concepts and algorithms to model irregular behaviour and anomalies in systems, which do not seem to be inherently stochastic. An original reference on phase space embedding is Takens (1981).

The first step of the Local Prediction in Phase Space (LPH) technique is phase space embedding of the observed pollution data *y(t)*. This is done by forming delay vectors

$$\vec{y}(t) = (y(t-(m-1)d),...,y(t-d),y(t))^T \tag{17}$$

with *m* representing the embedding dimension and *d* representing the delay time. Methods for estimating the optimal embedding parameters are discussed by Grassberger and Procaccia (1983), Kennel et al. (1992), and Sugihara and May (1990).

The second step is the local non-parametric extrapolation in the phase space, beginning at a starting point $\vec{y}(t)$. In a neighbourhood $U_z(\vec{y}(t))$ of this point, all points and their tracks are considered. According to Equation (18), the forecast is calculated as the average of these tracks

$$y(t+1) = \frac{1}{\left|U_\varepsilon(\vec{y}(t))\right|} \sum_{\vec{y}(s) \in U_\varepsilon(\vec{y}(t))} y(s+1), \qquad \text{with} \quad s < t \tag{18}$$

A peculiarity of the LPH technique is that it requires time series without missing values. Application of this technique for modelling O_3 and SO_2 time series were considered by Schlink et al. (2003).

3.2.4 The Kalman Filtering Approach

The Kalman filtering approach is based on the assumption that the linearization and discretization - of the differential equation that describes the concentration of a given pollutant - are described by a system of equations of the form (19):

$$\begin{aligned} X(t+1) &= A \cdot X(t) + B \cdot U(t) + C \cdot V(t) \\ Y(t) &= D \cdot X(t) + E \cdot Z(t) \end{aligned} \tag{19}$$

where *A, B, C*, *D* and *E* are matrices of parameters of appropriate dimensions, *X(t)* and *U(t)* are the state and input vectors respectively, and *V(t)* and *Z(t)* are purely random signals with zero mean and known covariance. The first part of Equation (19) is referred to as the state equation since it allows computation of the system state at time *t+1* on the basis of information available up to time *t*. The second equation is referred to as the output (or the observation equation), and reflects the

fact that the state of the system is considered as an internal variable (i.e., it may not be directly observed, but it can be computed based on an appropriate measurement process) involving the variables *Y(t)* and *U(t)*. The presence of the random variables, *V(t)* and *Z(t)*, is the way to represent the incomplete knowledge of the process. Interested readers can find details about methods available to estimate the unknown parameters of the model (19) in several books, including Brown and Hwang (1996). A comprehensive framework for analysing time series of environmental data, based on the use of recursive Kalman filters, was proposed by Young et al. (1991) and further improved by Young et al. (1997) and Young (1998). Application of such a methodology has been proposed by Schlink et al. (1997) and Ng and Yan (1998).

3.2.5 Clustering Approaches

A statistical model to predict episodes of poor air quality can be formalized as a classification problem. To understand this, it is necessary to bear in mind that often we are not interested to know the *exact* value that will assume the concentration of a given pollutant, but rather if the value will belong to a certain class. For instance, we might classify air quality in classes such as (excellent, good, acceptable, worse, or bad) and we may be interested to forecast the class of air quality for tomorrow. Or we might classify the concentration of a given pollutant in classes according to threshold levels suggested by actual legislation (e.g., the attention level of the alarm level) and we want to know if the concentration of a given pollutant will exceed the attention level tomorrow. Classification models can be obtained using a large variety of approaches. For instance, a traditional MLP **neural network** can be *trained* as a classifier to forecast if the concentration of a given pollutant will exceed the attention level or not, rather than try to predict the *exact* value. In this case, it would be possible to use well-known algorithms such as the back-propagation to find the model parameters. However, more appropriately, the classification can be achieved by using one of the numerous algorithms proposed in literature, such as the K-means developed by MacQueen in 1967 and the Hard C-Means. For a description of these algorithms, see Anderberg (1973) and Hartigan (1975), respectively, or the good synthesis made by Jorquera et al. (2004). A K-means clustering approach was considered by Sanchez et al. (1990), who presented results of a synoptic meteorological classification oriented at forecasting particulate matter in the city of Valladolid (Spain). Clustering was also used by Huang (1992) to predict air quality in the city of Xiamen (China).

The advent of Fuzzy Logic has stimulated the development of other clustering algorithms such as the Fuzzy C-Means developed by Dunn (1973), further improved by Bezdek (1981). This method is based on minimizing the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2, \quad 1 \le m < \infty \tag{20}$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the i^{th} of d-dimensional measured data, c_j is the d-dimension centre of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the centre. The real number m is referred to as the fuzzyfication parameter. If m is zero, the clusters are conventional. However, the larger the parameter value, the fuzzier the cluster will be. The recommended value for m is 2, which also assures convergence of the algorithm.

Fuzzy partitioning is carried out through an iterative optimisation of the objective function shown above, with the update of membership u_{ij}:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\left\| x_i - c_j \right\|^2}{\left\| x_i - c_k \right\|^2} \right)^{\frac{2}{m-1}}} \tag{21}$$

and the cluster centres c_j.

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{22}$$

This iteration will stop when:

$$\max_{ij} \left\{ \left| u_{ij}^{k+1} - u_{ij}^{k} \right| \right\} < \varepsilon \tag{23}$$

where ε is a termination criterion (between 0 and 1) and k is the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

3.2.6 Identification of Air Quality Models by Using Fuzzy C-Means

The Fuzzy C-means algorithm can be used to identify non-linear air quality models using the approach proposed by Sugeno and Yasukawa (1993). This approach allows minimization of the number of rules in the fuzzy rule base thus avoiding the drawback of traditional fuzzy model identification, where the number of fuzzy rules increases exponentially with the number of inputs. This is obtained by partitioning the input universe of discourse based on the Fuzzy C-means.

The method can be stated as follows. Let B represent a fuzzy cluster defined in the output variable universe and A the projection of B in the input space. The projection of A in the axes of the input variables x_j, (i =1,..,p) yields the fuzzy sets A_j. The projection must satisfy the following relation:

$$A_1(x_{1k}) = A_2(x_{2k}) =, \cdots, A_p(x_{pk}) = B_i(y_{1k}) \tag{24}$$

where $A_j(x_{jk})$ is the degree of membership of the k^{th} sample of the input variable x_j to the fuzzy set and A_j and $B_i(y_k)$ are the degree of membership of the k^{th} sample to the cluster B_i. Expression (24) gives the following fuzzy rule:

$$if\ x_1\ is\ A_1\ and\ x_2\ is\ A_2\ and \cdots x_p\ is\ A_p\ then\ y\ is\ B_1 \tag{25}$$

Note that in order to avoid generating two or more fuzzy rule for each input variable, it is necessary to assume that the fuzzy clusters cannot be convex. After generating the fuzzy set A, the fuzzy sets A_j are approximated by trapezoidal-type set. Hence, the consequent part of the rule generated from equation (24) is changed by consequents in the typical Takagi-Sugeno form:

$$y^i = c_0^i + \sum_{j=1}^{p} c_j^i \cdot x_j \tag{26}$$

The unknown parameters c_j^i are identified by using the traditional least square approach.

The original Takagi-Sugeno approach has a number of drawbacks. It performs the clustering process using only the information from the output space, thus ignoring the input space. Furthermore, non-convex clusters in the input space must be split into two or more sets. Some of these limitations were overcome by other authors such as Briseño and Cipriano (1996).

3.2.7 Bayesian Air Quality Models

Bayesian models, also known as Directed Acyclic Graphs (DAGs), are of increasing interest to the scientific community since they provide a natural tool for dealing with uncertainty and complexity (Jordan, 1999). They are essentially graphical models obtained as a combination of graph theory and probability theory. Bayesian models have been considered, in particular, in the machine learning and statistics communities. More recently, they have been applied to modelling dynamic systems because they can encode the time variable (Dynamic Bayesian Networks, DBNs). A very simple DAG is represented in Figure 2. It consists of two nodes labelled as X and Y respectively, which represent two

random variables. The arrows from X to Y can be informally interpreted as indicating that X causes Y. This model can be represented as:

$$P(X,Y) = P(X) \cdot P(Y \mid X) = P(Y) \cdot P(X \mid Y) \tag{27}$$

where $P(X,Y)$ represents the joint probability, $P(X)$ and $P(Y)$ are *a priori* probabilities, and $P(Y|X)$, $P(X|Y)$ are conditional probabilities. If we consider Y as the observed variable and X as the hidden variable, one goal of the model could be to infer X given Y (i.e., to invert the causal arrow).

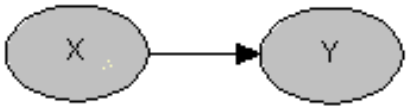

Figure 2. A very simple DAG.

This ability is referred to as inference. In order to make inferences, it is necessary to estimate the model parameters (learning) that can be represented as a conditional probability table (CPT). There are two main kinds of inference: exact and approximate. Exact inference, in the sense of having a closed form solution, is only possible in a very limited set of cases, most notably when all hidden nodes are discrete or when all nodes (hidden and observed) have linear Gaussian distributions. One of the most relevant exact inference algorithms for DAG models is the "chain-rule" decomposition that is illustrated by the following expression:

$$\begin{aligned} p(X) = p(X_1, X_2, ..., X_n) = p(X_1) \cdot \prod_{i=2}^{n} p(X_i \mid X_1, ..., X_{i-1}) = \\ = p(X_1) \cdot p(X_2 \mid X_1) \cdot ... \cdot p(X_n \mid X_1, ..., X_{n-1}) \end{aligned} \tag{28}$$

This algorithm essentially pushes sums inside products to marginalize the irrelevant hidden nodes efficiently; this is the so-called variable elimination algorithm (Pearl, 2000). The result of the computation is a single marginal $P(X_i \mid X_j)$. However, even in cases where exact inference is possible, it might not be computationally feasible; the cost of inference depends on the width of the inference tree. In this case, or when a closed form does not exist for the inference, one can use approximate inference. Several algorithms for approximate inference have been proposed, including the Expectation Maximization (EM) algorithm. The reader is referred to the recent book by Neapolitan (2004) for deeper insight into learning Bayesian Networks.

An example of Dynamic Bayesian Network is shown in Figure 3. This network was considered by Nunnari and Cannavò (2004) to model SO_2 daily mean time series at Melilli (Siracusa, Italy).

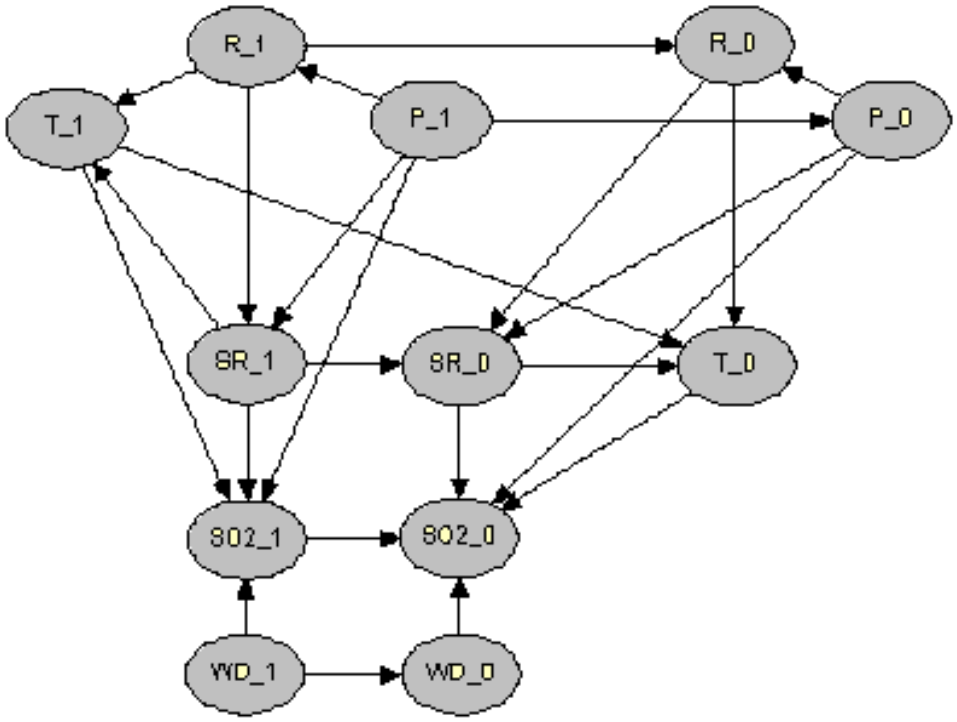

Figure 3. Example of Dynamic Bayesian Network. The meaning of the variables at the nodes is as follows: (T) air temperature, (P) atmospheric pressure, (SR) Solar Radiation, (WD) wind direction, (R) level of rain, (SO_2) sulphur dioxide. The symbol _0 or _1 close to the variable indicates that it is evaluated at time t (the day of prediction) or at time t-1 (the day before).

In this figure, the output node is $SO_2(t)$ (represented as SO2_0 in Figure 3), which allows us to compute the probability that the output belongs to one of the four classes defined for this variable. Hence, the model output is represented by the most probable class. Software packages such as Matlab® or Netica® can be used to learn the CPT for each node in the DAG.

4 Building a Model for Air Quality Forecast

Since statistical models for air quality forecast are based on extracting semi-empirical relationships from pollution time series, they are strictly dependent on the point where information is recorded. Furthermore, they depend on several factors such as the type of pollutant (e.g., primary or secondary pollutant), the type of target (e.g., prediction of the daily mean values, daily maximum value, or hourly mean values), the horizon of the forecast (e.g., 1 day, 12 hours, etc.), the type of area (e.g., urban, suburban, rural, industrial) and so on. We assume that the designer has already chosen these elements and is aware about the current normative about the limit values for the considered pollutant. An example of

normative is the EC–Council directive 1999/30/EC of 22 April 1999 that provides the limit values for sulphur dioxide, nitrogen dioxide, oxide of nitrogen, particulate matter, and lead in ambient air.

Furthermore, we assume that an adequate set of data is available in order to build and test the model under development. It is not easy to give rules for judging if a given available data set is adequate or not, since this depends on numerous factors enumerated above. Roughly speaking, we assume that a data set spanning over at least two years is available. The data set should concern both pollution and meteorological data. Emission data are in general unavailable, but in some cases one can try to use information related with emissions. As an example, in an urban area one can assume, as a surrogate of emissions data, the information from traffic flow data (if available).

4.1 Structure of Prediction Models

When building a short-term air quality forecast model, a crucial step is finding the most appropriate set of arguments of the unknown function in expression (7). In other terms, we have to establish the exogenous inputs (sometimes referred to as the "explanatory variables") and the number of regressions for each considered variable. It is necessary to stress here that the solution of this problem is perhaps one of the major problems for the modelers. First, all the candidate variables are often numerous and not necessarily known *a priori*. Moreover, the link between the pollutant concentration and the exogenous inputs is nonlinear and depends on the investigated geographical area. In addition, the selected variables depend on the particular target (daily maximum, daily mean, hourly mean, etc.). Finally, we must stress that the observed data are affected by various kinds of noises. Although several authors have addressed the problem of input variable selection, it is still resolved in an unsatisfactory way. This problem has been studied by Zickus et al. (2001) who evaluated which of 20 input variables are relevant for predicting exceedances of the European PM_{10} daily average limit value in the Helsinki Metropolitan Area. These authors showed good agreement with some selected predictors, but also variability among different methods. Aware of the high level of complexity in the variable selection problem, the designer can try to have some rough indication about candidate exogenous variables by using correlation analysis and calculation of typical days. Correlation and typical days can be computed as explained below.

4.1.1 Correlation Analysis

The internal correlation $\rho(\tau)$ of the observation in a time series $y(t)$ is expressed as a function of the time lag τ between observations and is defined mathematically as:

$$\rho_y(\tau) = \frac{E(y_t - \mu_y)(y_{t+\tau} - \mu_y)}{E(y_t - \mu_y)^2} \tag{29}$$

where y_t, $t = 0, \pm1, \pm2, \ldots$ represent the values of the series $y(t)$ and μ_y is the mean of the series. The symbol E denotes the expected value. Expression (29) can be calculated as:

$$\rho_{yc}(\tau) = \frac{\sum_{i=1}^{n-\tau}(y_i - \mu_{yc})(y_{i+\tau} - \mu_{yc})}{\sum_{i=1}^{n}(y_i - \mu_{yc})^2} \tag{30}$$

where μ_{yc} is the mean of the observed values $y_1, y_2, \ldots, y_n$. A plot of the values of the autocorrelation against the lag is known as the autocorrelation function. Similarly, the correlation between two time series $y(t)$ and $u(t)$ is defined as:

$$\rho_{yu}(\tau) = \frac{E(y_t - \mu_y)(u_{t+\tau} - \mu_u)}{E(y_t - \mu_y)(u_t - \mu_u)} \tag{31}$$

and can be computed as:

$$\rho_{yc}(\tau) = \frac{\sum_{i=1}^{n-\tau}(y_i - \mu_{yc})(u_{i+\tau} - \mu_{uc})}{\sum_{i=1}^{n}(y_t - \mu_{yc})(u_t - \mu_{uc})} \tag{32}$$

A serious limitation using correlation analysis is that with this technique, it is possible to identify only linear associations between the considered pollutant and other explanatory variables (e.g., meteorological variables) when we expect that the associations are non-linear.

4.1.2 Typical Day Analysis

A typical day is calculated using series of average hourly values as follows: 24 averages are calculated, one for each hour of the day being considered, for each day of the year. Each average is therefore calculated on 365 values (366 in a leap year) recorded at the same time of the day; this is expressed in a formula as:

$$\bar{y}(k) = \frac{1}{364}\sum_{h=0}^{364} y(k + h \cdot 24), \qquad k = 1,\ldots,24 \tag{33}$$

As an example of a typical day, SO_2 concentrations recorded at the station referred to as Melilli in the industrial area of Siracusa (Italy) from 1995 to 1999, are shown in Figure 4a. Typical days of SO_2 plotted with typical days of wind direction (WD) are shown in Figure 4b.

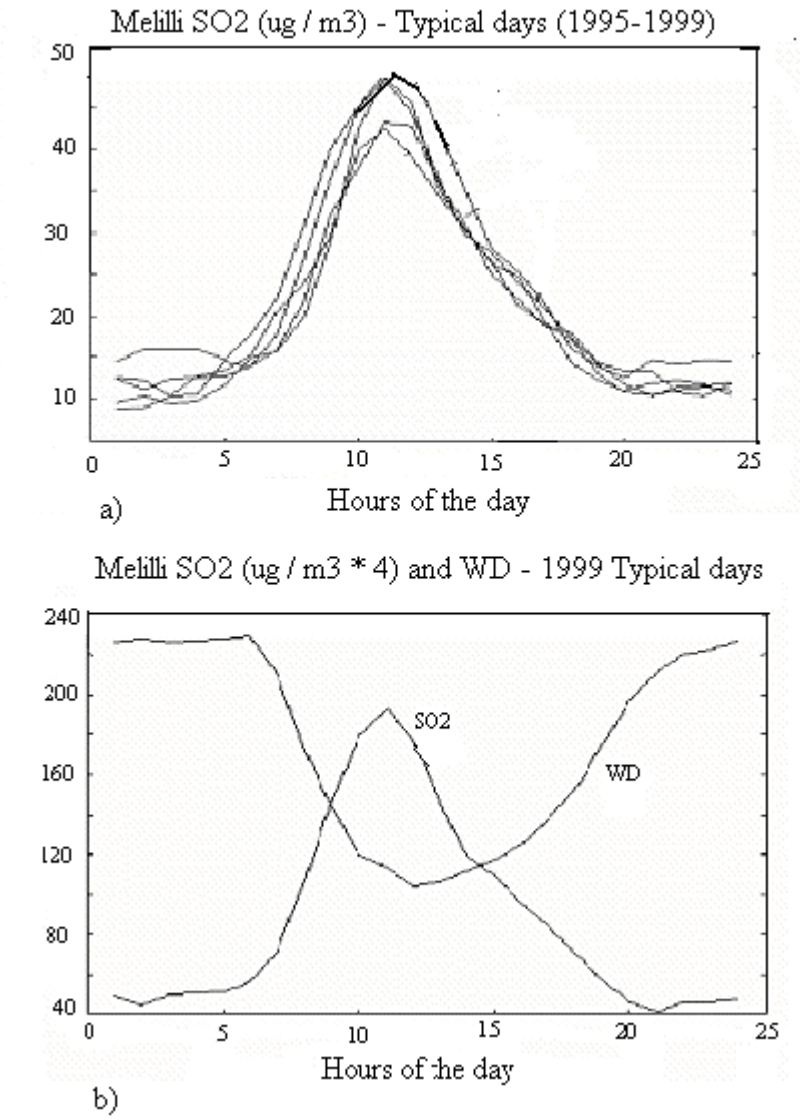

Figure 4. Example of typical day a) SO_2 b) SO_2 and wind direction.

The typical day SO_2 concentration at Melilli (Figure 4a) shows peaks between 10 a.m. and 12 p.m. local time that are somewhat surprising. In fact, pollution in this area is mainly due to oil refinery industries that are characterized by almost constant emission rates during 24 h. Hence, it seems reasonable to attribute the peaks to local atmospheric condition and, in particular, to the wind regime as suggested by Figure 4b. From Figure 4b, one can easily conclude that wind direction could be a candidate to explain the behaviour of SO_2 in this area. A

possible explanation of what has been observed was studied by Nunnari et al. (2004).

4.2 An Iterative Procedure for Building Models

Let us assume that, depending on the considered pollutant and target, a set of candidate explaining variables has been obtained by using correlation analysis and/or day type analysis, or any other *a priori* knowledge. Furthermore, let us suppose that a representative data set is available.

Independently of the modelling technique considered, the identification of a model for short term forecast of air quality involves an iterative process consisting of the following steps:

1. Divide the data set into (at least) two subsets: 1) a calibration data set which will be used to identify the model parameters; and 2) a testing data set that will be used to test the model performance (i.e., test its generalization capabilities). If possible, depending on the extension of the data set available, it is more suitable dividing that data set into three subsets. The first will be referred to as the calibration (or the learning) data set, the second as the validation set and the third as the test set that will be used to finally evaluate the model performance. The presence of the third data set is justified by the fact that in an iterative process, the second data set (i.e., the test) is in some sense still used to find an optimal set of model parameters, as it will be clear after reading the following steps.
2. Identify the model parameters by using one of the available methods (e.g., those enumerated in Section 3).
3. Validate the model against the validation data set.
4. Modify the model structure when it does not perform satisfactorily. In this case, go back to step 2 and repeat the whole process.

4.3 Evaluating the Model Performances

From the procedure outlined in the previous section, it is evident that the problem of evaluating the model performance plays a crucial role in the success of the model. It is important to consider indices that are able to evaluate performances in an objective manner. Indices usually considered for evaluating statistical air quality models can be grouped into two separate sets: 1) global fit indices, for example, those indices that give measures of the fit of the overall time series (i.e., RMSE error); and 2) those that give a measure of the capability of a given model to predict critical episodes, referred to as exceedance indices. A list of the considered performance indices is reported below. Interested readers can find detailed information about performance indices in Willmott (1982) and Willmott et al. (1985). Further insights can be found in Van Aalst and De Leeuw (1997).

4.3.1 Global Fit Indices

Let us indicate μ_o and μ_p as the mean of the observed time series (O) and predicted time series (P), and σ_o and σ_p as the corresponding standard deviations. The most widely considered global fit indices are the following:

Bias

$$\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i) \tag{34}$$

The mean bias error is the degree of correspondence between the mean forecast and the mean observation. Lower numbers are best. Values < 0 indicate under-forecasting.

MAE

$$\frac{1}{N}\sum_{i=1}^{N}\left|P_i - O_i\right| \tag{35}$$

The Mean Absolute Error is the mean of the absolute value of the residuals from a fitted statistical model. Lower numbers are best.

RMSE (the Root Mean Square Error)

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \tag{36}$$

σ_e^2 (the variance of the error)

$$\sigma_e^2 = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2 \tag{37}$$

σ^2_{un} (the unexplained variance in percent)

$$\sigma^2_{un} = \frac{\sigma_e^2}{\sigma_O^2}\cdot 100 \tag{38}$$

where σ_O^2 represents the variance of the observed time series.

d (the index of agreement)

$$d = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(|P_i - \mu_O| + |O_i - \mu_O|)^2} \quad (39)$$

It gives the measure of the degree of which predictions are error-free. With respect to a good model, the index of agreement should approach one.

ρ (the correlation coefficient observed-predicted)

$$\rho = \frac{\sum_{i=1}^{N}(O_i - \mu_O)(P_i - \mu_P)}{\sigma_O \sigma_P} \quad (40)$$

4.3.2 Exceedance Indices

This kind of performance indices were adopted by the European Environment Agency (Van Aalst and de Leeuw, 1997) to test the capabilities of a short term forecast model to predict exceedances of photochemical smog episodes, with particular attention to tropospheric ozone concentrations. These indices are defined according to the following standard *contingency table* (Table 1):

Table 1. The EEA contingency table.

Alarms	**Observed**		
Forecasted	**Yes**	**No**	**Total**
Yes	*A*	*f-a*	*f*
No	*m-a*	*N+a-m-f*	*N-f*
Total	*M*	*N-m*	*N*

where:

N = total number of data points
f = total number of forecast exceedances
m = total number of observed exceedances
a = number of correctly forecast exceedances

Using these definitions, the following indices can be defined:

SP (the probability of detection)

$$SP = \frac{a}{m}100 \tag{41}$$

SP is the *fraction of correct forecast* of critical events. Its values range from 0 to 100 (100 being the best value).

SR (the percentage of predicted exceedances actually occurred)

$$SR = \frac{a}{f}100 \tag{42}$$

SR is the *fraction of realised forecast* critical events (range from 0 to 100 with a best value of 100).

FA (the false alarm rate)

$$FA = (100 - SR) \tag{43}$$

FA is the percentage of instances when predicted exceedances do not occur. With respect to a good model, *FA* should approach zero.

SI (the success index)

$$SI = (\frac{a}{m} + \frac{N + a - m - f}{(N - m)} - 1) \cdot 100 \tag{44}$$

SI indicates how well the exceedances were predicted. *N* is the total samples in the time series. Since *SI* is not affected by a large number of correctly forecasted non-exceedances, it is useful for evaluating rare events. *SI* ranges from –100 to 100 (100 being the best value).

5 Identification of Statistical Air Quality Models

In this Section we report some case studies concerning the application of statistical modelling techniques to different areas. Results refer to different time-horizon (from hours to 1 day), targets (e.g., 1 hour average, 1 day average, 1 day maximum, etc.), and modelling techniques.

5.1 Ozone *Grey-Box* and MLP at Brescia and Catania (Italy)

The application of *grey-box* and *artificial neural networks* models has been performed in two Italian urban areas: 1) Brescia in the Northern part of Italy and 2) Catania in the Southern part (Finzi et al., 1998). See Figure 5.

Figure 5. Location of Brescia and Catania.

For each city, the examined data records consist of 1 h average O_3, CO, NO and NO_2 concentrations measured by the urban air quality monitoring network. In particular, tropospheric *ozone* is a photochemical oxidant that may cause serious health problems to people and damage to materials and crops. The European Community directive 92/72/EEC, following the WHO guidelines, prescribes air quality standards for ozone in terms of threshold values for health protection, population information and warning (Sluyter and Van Zantvoort, 1996). The critical anthropogenic emissions (mainly traffic and combustion processes), the frequent stagnating meteorological conditions and the high solar radiation in Mediterranean regions cause ozone peaks, especially during the summer months. In order to take short-term abatement actions to prevent critical episodes, a proper real time concentration exceedances alarm system was set up for population information and warning; different forecast modelling methodologies have been used and compared.

Due to the particular relevance of photochemical pollution during summer season, two time series of hourly data measured during June-August 1996 and 1997 respectively have been taken into account for both cities. The first one has been used as the training set in identifying the model stage, while the validation of the predictors has been performed on the second one.

A pre-processing phase was required to remove the patterns containing incomplete data due to non-working or re-calibration of the measuring instruments.

5.1.1 *Grey-Box* Model Identification

Different classes of *grey-box* models have been considered and identified for O_3 concentration among stationary and cyclo-stationary autoregressive models having as inputs other chemical compounds taking part in the photochemical reactions. The most significant phenomenon in explaining O_3 dynamics is the 24 h period of solar radiation, which is directly connected to the photochemical atmosphere reactivity and indirectly to the regular variation of vehicular urban traffic emissions throughout the day.

The particular *grey box* model examined is cyclo-stationary with period of 24 h and assigned ranges of internal stationarity of the parameters during sub-periods of the day (night, sunrise, morning, afternoon, sunset). Model performances were evaluated in terms of Bias, σ_e, σ^2_{un}, and ρ. The results obtained are reported in Table 2 and 3 for Brescia and Catania respectively.

Table 2. Performance indexes for *grey-box* predictors at Brescia.

Performance Indexes	Identification (1996)	Test (1997)		
		1 hour	3 hours	6 hours
Bias	0	0.32	0.79	1.68
σ_e	7.25	8.78	11.53	12.56
σ_{un}^2	0.13	0.32	0.54	0.63
ρ	0.93	0.83	0.69	0.62

Table 3. Performance indexes for *grey-box* predictors at Catania.

Performance Indexes	Identification (1996)	Test (1997)		
		1 hour	3 hours	6 hours
Bias	0	0.65	0.25	3.61
σ_e	4.57	5.64	9.73	11.31
σ_{un}^2	0.06	0.08	0.23	0.31
ρ	0.97	0.96	0.86	0.83

5.1.2 MLP Neural Network Model Identification

It is well known that there are no practical criteria for the definition of the topology of the MLP solving a given problem. Hence, the best network topology has been searched by a trial and error procedure. As a first attempt, the input pattern has been defined as:

$$y(t),\ y(t-1),\ \ldots, u_1(t), u_1(t-1), \ldots,\ u_1(t-n_1),\ \ldots, u_p(t), u_p(t-1), \ldots,\ u_p(t-n_p)$$

where y(t) is the ozone concentration recorded at time t and $u_1(t),\ ..., u_p(t)$ are exogenous inputs representing other pollutants of the nitrogen cycle recorded with the ozone (e.g., NO, NO_2, NMHC). The structure of the output pattern has been defined as $[y(t+k_a)]$, k_a being the number of steps ahead of the prediction model. Several attempts have been performed varying the parameters $p,\ n_1,\ ...,\ n_p$, and the number of units in the hidden layer n_h . The parameter k_a has been set to 1, 3 and 6. For each attempt, the back-propagation algorithm has been used to train the network. At the end of the trial-and-error procedure, the network topology has been obtained as the one giving the best set of performance indexes. In this application, Auto Regressive non-linear models have been investigated. The best network topology for the considered problem was a MLP of the type 3-12-1 (i.e., 3 neurons in the input layer, 12 neurons in the hidden layer and 1 neuron in the output layer). This result is reliable, independently from the value of k_a. In Table 4 and 5, the performance indexes are reported.

Table 4. MLP neural network predictors at Brescia.

Performance Indexes	Learn (1996)	Test (1997)		
		1 hour	3 hours	6 hours
Bias	-0.49	-0.53	-0.18	2.86
σ_e	4.84	5.92	6.71	8.50
σ_{un}^2		0.15	0.06	0.30
ρ		0.92	0.97	0.90

Table 5. MLP neural network predictors at Catania.

Performance Indexes	Learn (1996)	Test (1997)		
		1 hour	3 hours	6 hours
Bias	0.01	-0.45	-1.19	-3.68
σ_e	4.68	5.55	9.06	11.73
σ_{un}^2	0.07	0.08	0.21	0.34
ρ	0.97	0.96	0.89	0.86

5.1.3 Discussion of Results

The performance indexes reported in Tables 2 to 5 show that the mean square error, the ratio of the unexplained variance and the correlation coefficient in the considered case are generally more satisfying for the neural model. This is evident from the validation concerning the Brescia data set. In particular, the last index reveals a greater efficiency of neural models to *capture* the deterministic and persistent part of the historical time series.

In order to test the capabilities of the predictors to predict episodes of poor air quality threshold, the SP and FA indices were computed for different time horizon and thresholds. The results are shown in Table 6 and 7 for Brescia and Catania respectively.

Table 6. Performance indexes corresponding to different O_3 concentration threshold value at Brescia.

Forecast Step	Threshold value	Outl. N.	Grey box		Neural network	
			SP	FA	SP	FA
3 hours	50 ppb	443	77.9	27.2	74.7	12.7
6 hours			75.4	30.4	81.0	34.1
3 hours	70 ppb	34	55.9	7.0	52.9	2.3
6 hours			64.7	9.0	38.2	1.8

Table 7. Performance indexes corresponding to different O_3 concentration threshold value at Catania.

Forecast Step	Threshold value	Outl. N.	Grey box		Neural network	
			SP	FA	SP	FA
3 hours	50 ppb	240	88.7	12.9	88.7	11.1
6 hours			86.2	13.2	80.4	9.1
3 hours	70 ppb	30	23.3	1.6	6.7	0.0
6 hours			0.0	0.0	0.0	0.0

For the cleverness of the two classes of models in foreseeing the O_3 peaks correctly, the results show that the *grey-box* models tend to have a higher performance in forecasting critical episodes, although they give a larger number of false alarms. This fact can be related to the different features of the two model classes; in particular, the considered *grey-box* models are time-variant while the identified neural models have a stationary structure. So it seems that neural models give more *conservative* predictions than *grey-box* models. The comparison between temporal O_3 patterns (measured and forecast values), reported in Figures 6 and 7 for Brescia and Catania respectively, looks quite

satisfying. The two periods (10-15 August 1997 for Brescia and 3-7 July for Catania) have been chosen as significant both for their criticality with respect to ozone pollution over Europe and for their typical photochemical feature.

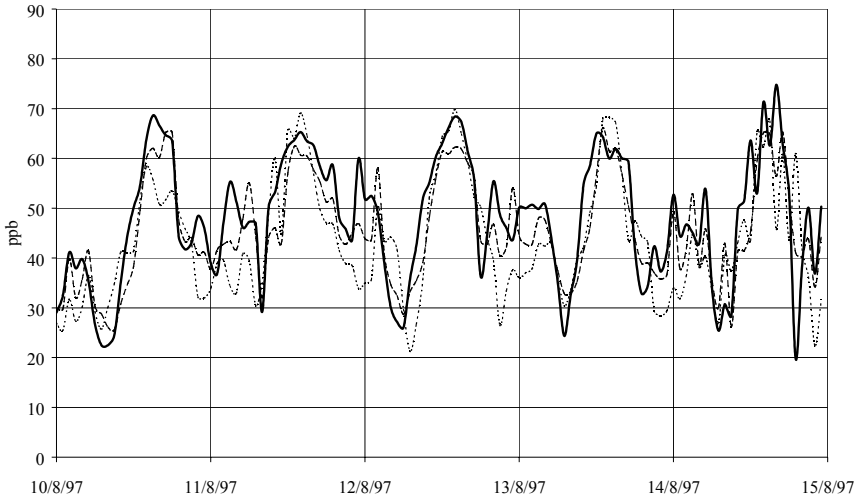

Figure 6. Brescia O_3 patterns: measured concentrations (__), 3 hours ahead forecast by means of *grey-box* model (. .) and neural network (- -).

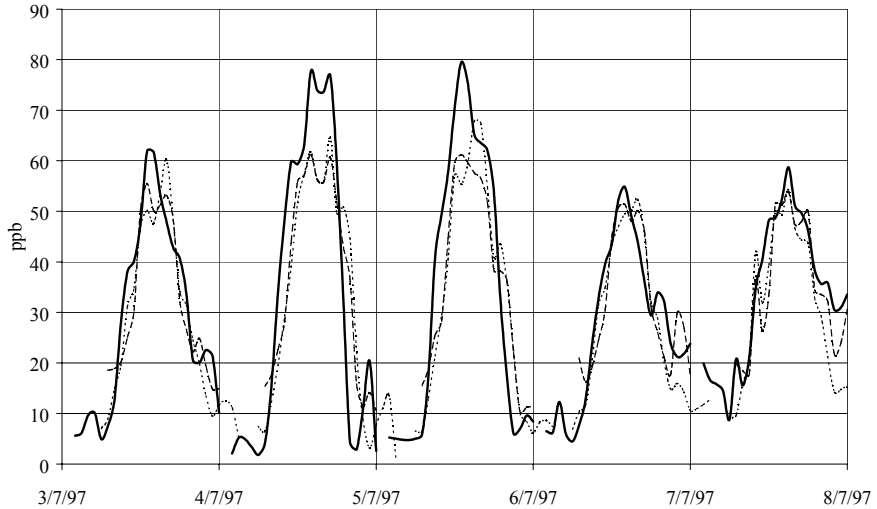

Figure 7. Catania O_3 patterns: measured concentrations (__), 3 hours ahead forecast by means of *grey-box* model (. .) and neural network (- -).

5.2 Ozone Fuzzy and Neuro-Fuzzy Model at Brescia and Siracusa (Italy)

In this Section, *fuzzy* and *neuro-fuzzy* models for ozone at Brescia and Siracusa (industrial area) are considered. The industrial area of Siracusa is located in the eastern cost of Sicily, about 50 Km south of Catania (Figure 5). In the post-WWII

period, one of the largest concentrations of petrochemical industries in Europe developed here and it is considered to be an area of high environmental risk.

5.2.1 Brescia Metropolitan Area

The examined data records consist of O_3, CO, NO and NO_2 hourly concentrations measured by the urban air quality monitoring station in the city of Brescia. Local temperature monitored and forecast data were available from the meteorological office. The models were identified on 1994-1998 and validated on 1999 summer season data (May to September).

The *neuro-fuzzy* network forecast is performed on the maximum expected hourly concentration value during the afternoon. The model has been identified assuming triangular membership functions and *sum-prod* inference mechanism. The *crisp* model inputs are O_3 concentrations and the most relevant meteorological parameter (temperature) taking part in the photochemical reactions during the day (Finzi and Volta, 2000).

Table 8 shows the inputs and their respective fuzzy set number for the best model as a trade-off between a satisfying forecast performance and a possible operational implementation.

Table 8. The *neuro-fuzzy* model inputs.

Inputs	Value	Fuzzy sets
O_3 conc.	10 a.m.-12a.m. average	3
O_3 gradient	12 a.m.-6 a.m. difference	4
Temperature	10 a.m.-12a.m. average	3
Temperature	12am- 6a.m. difference	2

The rule base came out to be a set of 30 rules. The *persistent model* skill parameters have been also computed as lower bound performance indexes. The forecast evaluation has been related to an O_3 threshold value of 140 $\mu g/m^3$. Figure 8 compares the European skill parameters computed for the persistent and the *neuro-fuzzy* predictor. The *neuro-fuzzy* model seems worthy to be used mainly for its cleverness in avoiding false alarms, while the *SP* index claims for a forecast improvement in enhancing some episodes.

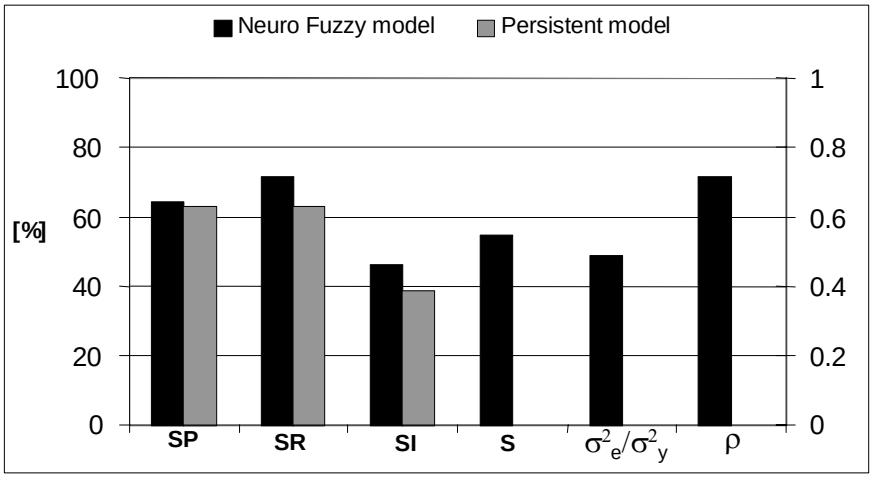

Figure 8. Performance indexes referred to the persistent model.

5.2.2 Siracusa Industrial Area

Fuzzy models have been obtained by using the approach described in Section 3.2.1.2. Data for model identification were recorded during 1995-1998, while the test was done using data recorded in 1999. The *fuzzy* model forecasts are performed at 8 p.m., giving the maximum expected hourly concentration value during the day after. The inputs of the *fuzzy* prediction model studied for the Siracusa industrial area are shown in Table 9.

Table 9. The *fuzzy* model inputs.

Inputs	Value	Fuzzy sets
O_3 conc.	1 a.m. – 8 p.m. average	3
NO_2	1 a.m. – 8 p.m. average	3
NO_X	1 a.m. – 8 p.m. average	3
Temperature	10 a.m. – 6 p.m. average	3
Solar Radiation	10 a.m. – 6 p.m. average	3
Pressure	10 a.m. – 6 p.m. average	3
Wind Direction	10 a.m. – 6 p.m. average	3

The proposed model was compared with a persistent model. The *fuzzy* model identified consist of 42 rules (3 fuzzy sets of trapezoidal type for each considered). The results of the comparison carried out with the persistent model for a threshold of 140 $\mu g/m^3$ are reported in Figure 9.

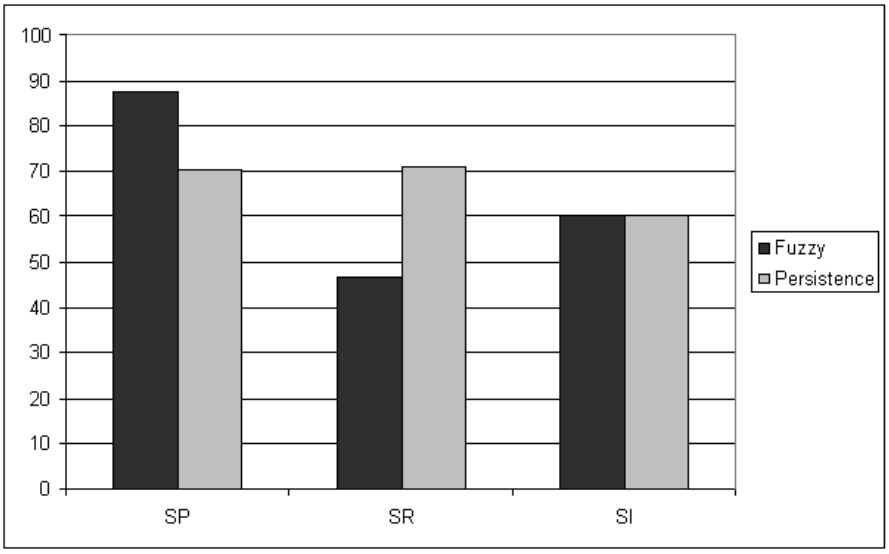

Figure 9. Performance indices for the Fuzzy model compared with the persistent model for the Siracusa industrial area.

5.2.3 Comments on Results

The *neuro-fuzzy* model identified for the Brescia metropolitan area, which was designed specifically to predict critical episodes, shows a satisfying performance both in forecasting exceedances of the threshold level and in avoiding *false alarms*.

The *SP* index obtained for the *fuzzy* model identified for the Siracusa industrial area is considerably better than the one exhibited by the persistent model. On the contrary, the *fuzzy* model shows worse performances in terms of *SR*, which results in a larger number of false alarms. This is due to the fact that the performances of the *fuzzy* model have been optimised with respect to the *SP* parameters.

Finally, it must be observed that:

- *fuzzy* and *neuro-fuzzy* predictors perform better than persistent model (i.e. linear ones)
- *fuzzy* and *neuro-fuzzy* model complexity and flexibility allow the optimization of model performances, stressing the capability to forecast exceeding values or avoiding *false alarms*
- *neuro-fuzzy* models, although they are non-linear, are more *readable* (due to the typical "if … then" form) than the MLP neural network models, and can suggest physical explanation of pollutant processes.

5.3 Inter-Comparison Among ARCX, MLP, NFU and FU Forecast Models

In the preceding Section 5.2 *grey-box* and **neural networks** predictors were compared to each other, while in Section 4.2 the pair of *neuro-fuzzy* and *fuzzy*

models were considered. In this Section, a wider inter-comparison exercise involving *grey-box* (ARCX), **neural networks** (NN), *neuro-fuzzy* (NFU) and *fuzzy* (FU) models will be reported. The selected areas are still Brescia and Siracusa, and the target is the forecast of ozone daily maximum concentration. For all the considered models, exogenous inputs were:

- the average ozone concentration computed between 4 p.m. and 8 p.m. of the day before
- the maximum temperature between 1 a.m. and 8 p.m.
- the NO_2 average concentration between 4 p.m. and 8 p.m. of the day before

The data set for both areas was represented by meteo–chemical measures recorded in the years 1995 to 2001. Instead of splitting the data into two different sets (the so called *learning* set and *validation* set), as it is usual among the neural network practitioners, in order to take the generalisation capabilities of the prediction models into account, the data set was divided into three subsets: the *learnin*g set, the *validation* set and the *testing* set. It has been pointed out indeed (e.g., Sjoberg and Ljung, 1995) that the validation error rate, periodically computed during the *learning* phase, is not a good estimate of the *generalisation* error. One way to evaluate an unbiased estimate of this last error is to run the prediction model on a third set of data, the *test* set, not used at all before for the training process. In the inter-comparison exercise, the *learning* set, the *validation* set and the *testing* set were composed respectively by data recorded in 1995 to 1998, 1999, 2000 and 2001. The results, referring to a threshold of 140 $\mu g/m^3$ in terms of forecasting performances, are shown in Figures 10 and 11 for Brescia and Siracusa respectively.

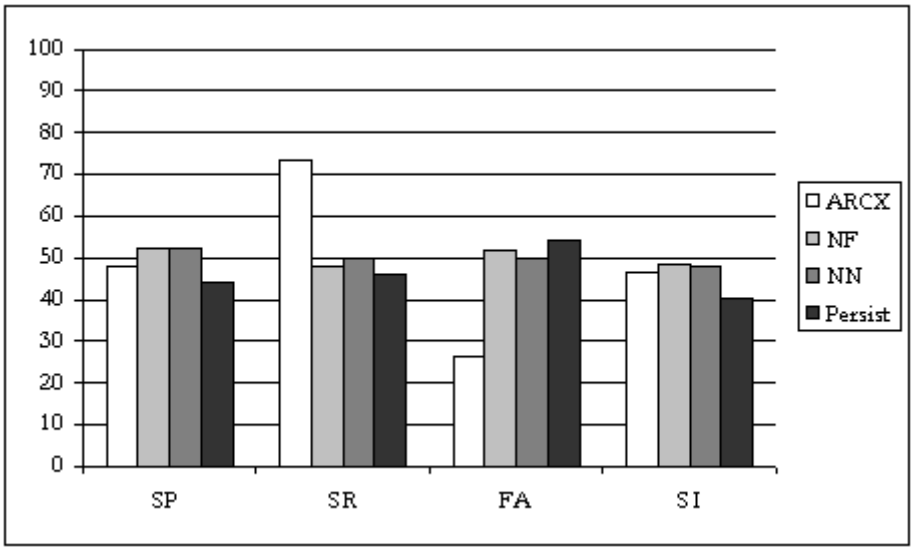

Figure 10. Performance indices for Brescia.

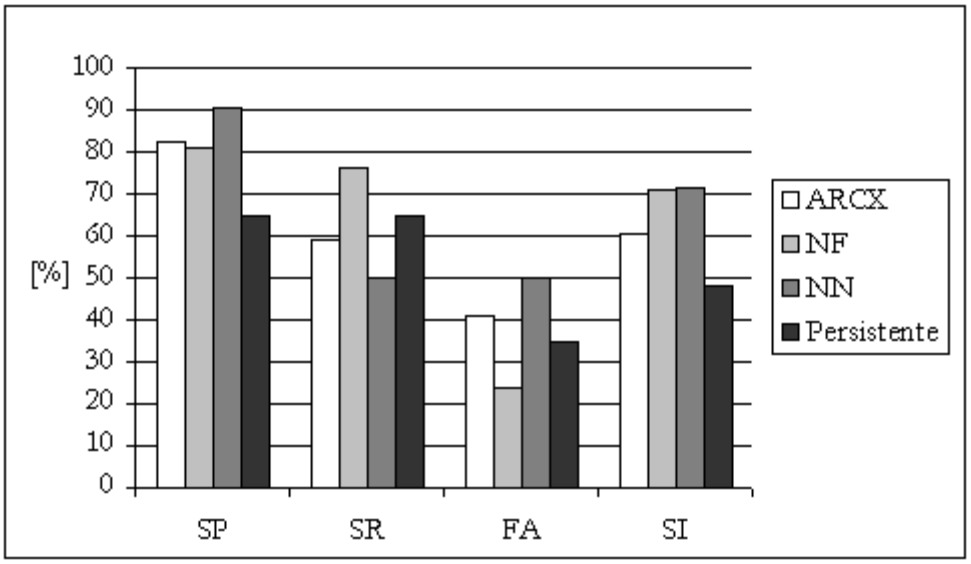

Figure 11. Performance indices for Siracusa.

Based on the reported results, the following considerations can be made:

- The forecasting performances are generally better for Siracusa than for Brescia. This means that the performance levels are point dependent no matter what modelling technique is considered. This fact can be easily explained by taking the different geographic meteo-climatic conditions of the two areas into account. Probably there were ozone accumulations in Siracusa that depend on a quite regular sea breeze regime. The concentration daily peaks can be predicted more easily with respect to Brescia.
- ARCX, NF and NN predictors work better than the persistent model in terms of SP and SI indices.
- While NN and NF usually perform slightly better than ARCX in terms of SP and SI, ARCX performs better in terms of FA in some cases.

5.4 Conclusive Remarks

At the end of this Section some further remarks are made referring to the results reported in the preceding Sections 5.1 to 5.3. The first consideration is that there is no single modelling approach exhibiting all the performance indices at the best level. The structure of a model, as well as its performance, is strictly dependent on the particular monitoring site. The forecast performance is usually higher if reliable meteorological information is provided as inputs. This means that, when information about pollutant emissions is not available, meteorological conditions play a key role in improving the reliability of predictions. The results reported show that not all observed critical episodes can be explained on the basis of historical concentration time series and meteo-climatic data measured at ground level. Neural based approaches, even the *neuro-fuzzy* version, seem more promising (though moderately) to critical events forecast.

6 An Operational Decision Support System

As example of possible operational use of forecast models, a prototype of Decision Support System (DSS) for short-term emission reduction measures, is described in Figure 12 (Finzi, 2001). It implements two feedback loops, which may be based on different methodologies taken into consideration. Air quality status forecast, given by the daily model, is supplied to the Control Authority in order to support the decisions relevant to the emission abatement strategies (vehicle traffic reduction or restriction in different metropolitan areas, temporary adoption of closer industrial emission limits, health prevention policies, etc.). These measures can prevent smog episodes if they are planned ahead of time. It is also possible to inform the population by means of media, in order to limit the unhealthy exposures; in this way, the feedback can prevent and reduce both pollution and sanitary risks.

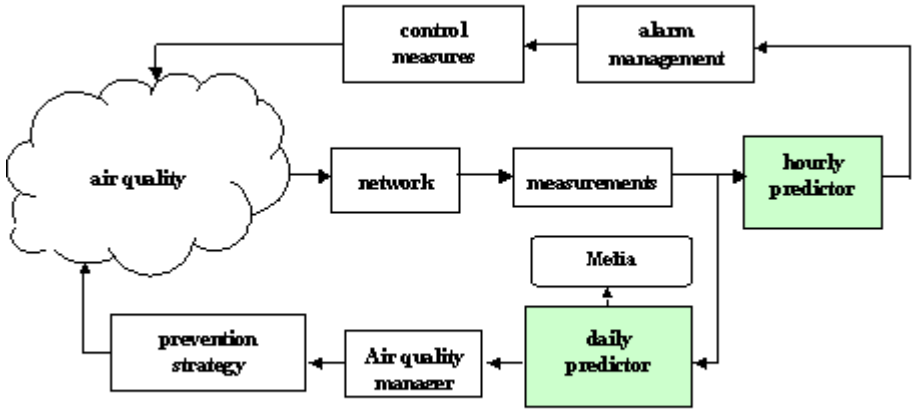

Figure 12. A Decision Support System for Air Quality Alert.

The designed system has a second internal feedback improving operational effectiveness in the short term (hours). The information, provided both by the hourly predictor model and the on-line meteo-chemical networks, allows the air quality managers to monitor the current pollutant evolution with a high confidence.

Moreover, any exceedance of the threshold, if not correctly forecast a day in advance, can be quickly recognised in the morning by the alarm system in order to apply short-term pollution control measures (traffic information through road panels, mobilisation of the metropolitan police, traffic control by means of a computerised system of traffic lights, etc.).

The metropolitan areas of Brescia and Milan have been considered as case studies and the performance of the designed DSS in both cities is examined in the following paragraphs.

6.1 Brescia Metropolitan Area Case Study

The examined data records consist of O_3, CO, NO and NO_2 hourly concentrations measured by the urban air quality monitoring station in the centre of Brescia (see Figure 13). The city is located in the Po Valley in Northern Italy and is characterised by high industrial, urban, and traffic emissions, and continental climate.

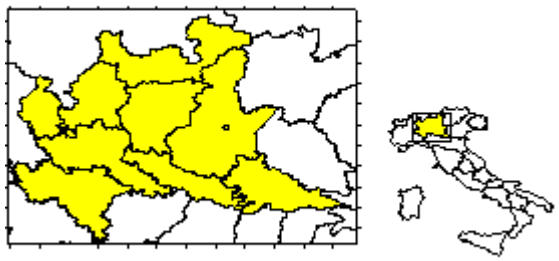

Figure 13. The Brescia area (in yellow).

Local temperature monitored and forecast data are available from the meteorological office. Both classes of models were identified on the period 1994-1998, and validated on 1999 summer season data (May to September). A pre-processing phase was required to remove the patterns containing incomplete data due to non-working or re-calibration of the measurement instruments.
Two alternative alarm DSS (see Figure 12) were set up:

- **a)** *neuro-fuzzy* network forecasts were performed on the maximum expected hourly concentration value one day in advance, while *grey-box* predictors provided 4 hour ahead forecast of O_3 concentrations from sunrise to noon during the same day
- **b)** *grey-box* forecast was performed on the maximum expected hourly concentration value one day in advance, while *neuro-fuzzy* network predictor provided the maximum expected hourly concentration value at noon for the afternoon during the same day.

Case a)

The *neuro-fuzzy* model has been identified for O_3 maximum daily concentration, assuming Gaussian membership functions and max-min inference mechanism. The *crisp* model inputs were the chemical compounds (O_3 and NO_2 concentrations) and the most relevant meteorological parameter (temperature) taking part in the photochemical reactions during the day. Table 10 shows the inputs and their respective fuzzy set number for the best model as a trade-off between a satisfying forecast performance and a possible operational implementation. The rule base came out to be composed by 135 rules.

Table 10. The neuro-fuzzy model inputs.

Inputs	Value	Fuzzy sets
O_3 conc.	Max concentration	6
O_3 conc.	4p.m. - 8p.m. average	4
NO_2 conc.	4p.m. - 8p.m. average	4
Temperature	Max forecast for the following day	3
Temperature	Max	5

Different *grey-box* models have been considered and identified for O_3 hourly concentration in the classes of stationary and cyclo-stationary autoregressive models (Finzi and Volta, 2000).

The most significant phenomenon in explaining O_3 hourly value dynamics appears to be the 24 h period of solar radiation, which is directly connected to the photochemical atmosphere reactivity and indirectly to the regular variation of vehicular urban traffic emissions throughout the day. So, the particular *grey-box* model considered is a *cyclo-stationary* one of 24 h period, with assigned sub-period internal stationarity ranges for the parameters (night, sunrise, morning, afternoon, and sunset).

The *persistent model* (tomorrow equals today) skill parameters have also been computed as lower bound performance indices. The indices, related to an O_3 threshold value of 140 $\mu g/m^3$, have been estimated for both daily *neuro-fuzzy* and hourly *grey-box* models. On the basis of the skill parameters computed for the persistent, the *neuro-fuzzy* and *grey-box* predictor, the second model seems worthy to be used mainly for its cleverness in avoiding false alarms, while *SP* index claims for a forecast improvement in enhancing some episodes. The hourly *grey-box* model provides a second internal system feedback, improving operational effectiveness in the short term (hours). The results (Figure 14)

underline the improvement mainly in forecasting ozone threshold exceedances (*SP*) and in performing *success index* (SI).

The comparison among temporal O_3 patterns (measured and forecast values), reported in Figure 15, points out the improvement provided in detecting alarms by the hourly model during a particular critical episode between 28 June and 7 July 1999.

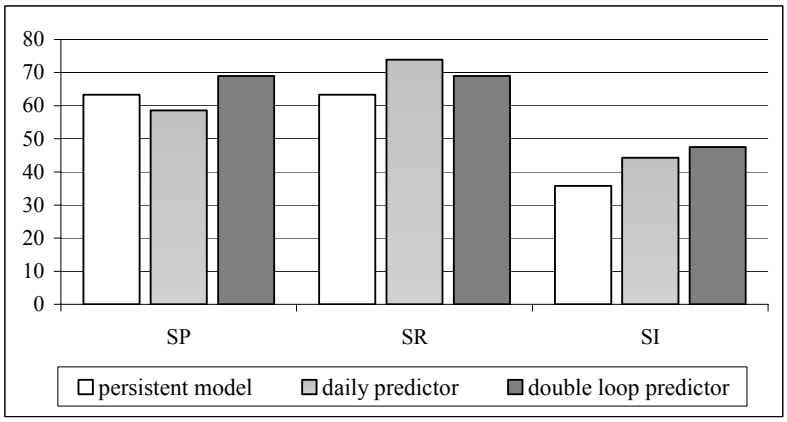

Figure 14. The estimated forecast skill parameters.

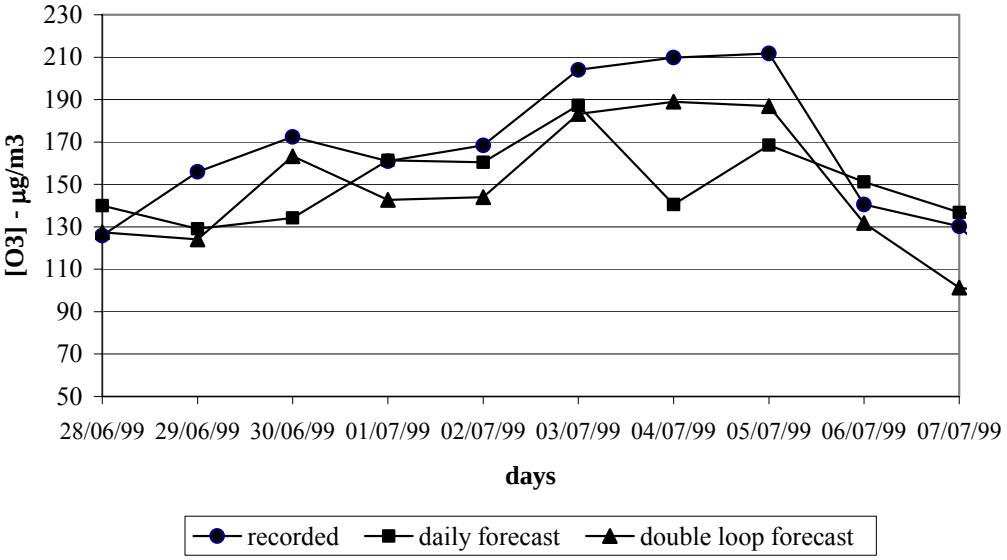

Figure 15. Ozone patterns: measured concentrations, daily forecast by means of single loop *neuro-fuzzy* model and double loop 4 hour ahead *grey-box* predictor.

Case b)

Different *grey* box models have been considered and identified for O_3 maximum daily concentration forecast. The most significant phenomena in explaining O_3 maximum value dynamics appear to be the solar radiation and temperature, directly connected to the photochemical atmosphere reactivity, and the regular variation of vehicular urban traffic emissions during the day. So, the particular *grey box* model considered is a *non-linear* ARX(1), with assigned temperature categories (Table 11).

Table 11. The *grey-box* model inputs.

Inputs	Value
O_3 concentration	4 p.m. to 8 p.m. average
NO_2 concentration	4 p.m. to 8 p.m. average
Temperature	Daily maximum value squared

The *neuro-fuzzy* approach has been used to forecast the maximum expected hourly concentration value during the afternoon. The model has been identified assuming *triangular* membership functions and *sum-prod* inference mechanism. The *crisp* model inputs were O_3 concentrations and the most relevant meteorological parameter (temperature) taking part in the photochemical reactions during the day.

Table 12 shows the inputs and their respective fuzzy set number for the best model as a trade-off between a satisfying forecast performance and a possible operational implementation. In this case, the rule base came out to be composed by 30 rules.

Table 12. The neuro-fuzzy model inputs.

Inputs	Value	Fuzzy sets
O_3 conc.	10 a.m. to noon average	3
O_3 conc.	8 a.m. to noon gradient	4
Temperature	10 a.m. to noon average	3
Temperature	6 a.m. to noon gradient	2

The performance indices, related to the O_3 threshold value of 140 $\mu g/m^3$, have been estimated for both *neuro-fuzzy* and *grey-box* models. The *persistent model* (tomorrow equals today) skill parameters have also been computed as lower bound performance indices.

Figure 16 compares the skill parameters computed for the four models: the daily (*grey-box*) and hourly (*neuro-fuzzy*) predictors, their combination in the air quality system and the persistent model. The first two predictors have good performance in avoiding false alarms, while the *SP* index for daily model provides a forecast improvement in enhancing some episodes. The *neuro-fuzzy* predictor matches the persistent model in correctly forecasting smog events.

The air quality system of Figure 12, implementing the second internal feedback, improves operational effectiveness in the short term (hours), taking into account all recent available meteo-chemical measurements. The results underline the system synergy, mainly in forecasting ozone threshold exceedances (SP) and in global performance (SI).

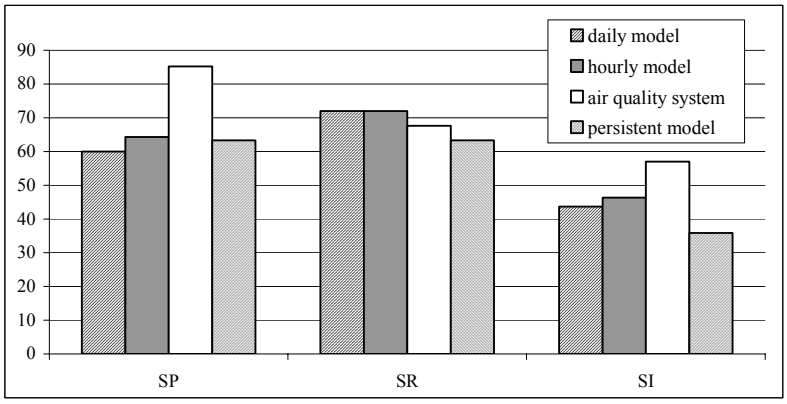

Figure 16. The estimated forecast skill parameters.

6.2 Milan Metropolitan Area Case Study

The examined data records consist of photochemical pollutants hourly concentrations measured by the urban air quality monitoring network in Milan during 1994-1999. Local and synoptic meteorological data are also available from the meteorological office. The assessment of the results, in terms of forecast performance indices and statistical indicators, according to European Environment Agency guidelines, is presented in the following.

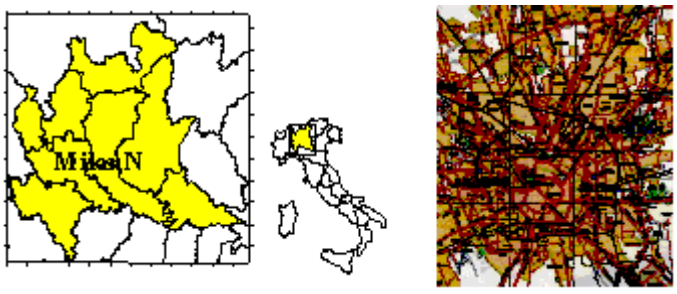

Figure 17. The Milan metropolitan area (in yellow).

The Milan metropolitan area, the city and its suburbs (see Figure 17), is a very industrialised and populated region in the Po Valley. The data records examined in this study consist of O_3, CO, NO and NO_2 hourly concentrations measured by three urban and suburban air quality monitoring stations (Parco Lambro, Via Juvara, Piazza Zavattari). Local temperature monitored and forecast data are available from the meteorological office. The models have been identified during the period 1994-1998 and validated during 1998-1999. A pre-processing phase was required to remove the patterns containing incomplete data due to non-working or re-calibration of the measuring instruments.

Different predictors have been considered and identified for O_3 and NO_2 maximum daily concentrations in the classes of stationary and cyclo-stationary auto-regressive models (Finzi, 2001).

The most significant phenomenon in explaining O_3 hourly value dynamics appears to be the 24 h period of solar radiation, which is directly connected to the photochemical atmosphere reactivity and indirectly connected to the regular variation of vehicular urban traffic emissions along the day. So, the particular *grey box* model considered is an auto-regressive model with exogenous inputs and categories. The inputs are O_3 average concentrations measured from 4 p.m. to 8 p.m. and the maximum temperature recorded during the day. The categories are defined for the future trend of the maximum temperature. The model has been identified for use in the summer seasons.

The model, referring to the maximum NO_2, is similar to the preceding one. The particular *grey box m*odel considered is an auto-regressive model with exogenous inputs and categories; the inputs are NO_2 average concentrations measured mostly in the morning (from 7 a.m. to 2 p.m.) and the maximum temperature recorded during the day. The categories are built for the future trend of the maximum temperature. The model has been identified for use in the winter seasons.

The *persistent model* (tomorrow equals today) skill parameters have been also computed as lower bound performance indices. The assessment of the results is presented in the following Figures 18 (*a, b, c, d*), both in terms of statistical indicators and real-forecast series comparison examples. All the indices have been computed with reference to a O_3 threshold value of 150 $\mu g/m^3$ and to a NO_2 threshold value of 135 $\mu g/m^3$. As it can be seen, *grey-box* models, although structurally simple, give more reliable alarm forecasts with respect to persistent models for all the examined pollutant measurement locations.

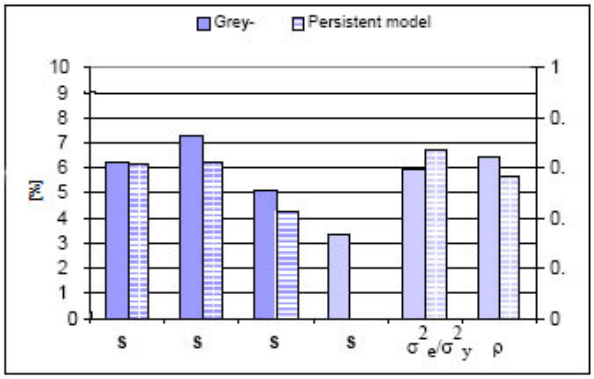

Figure 18a. Forecast System Validation Juvara station. NO_2 skill parameters (winter 1998).

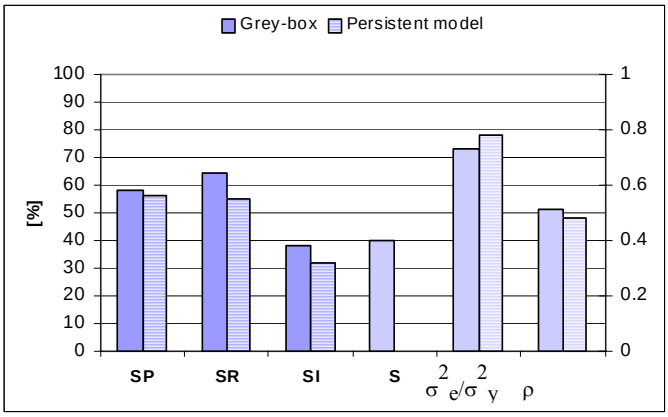

Figure 18b. Zavattari station. NO_2 skill parameters (winter 1998).

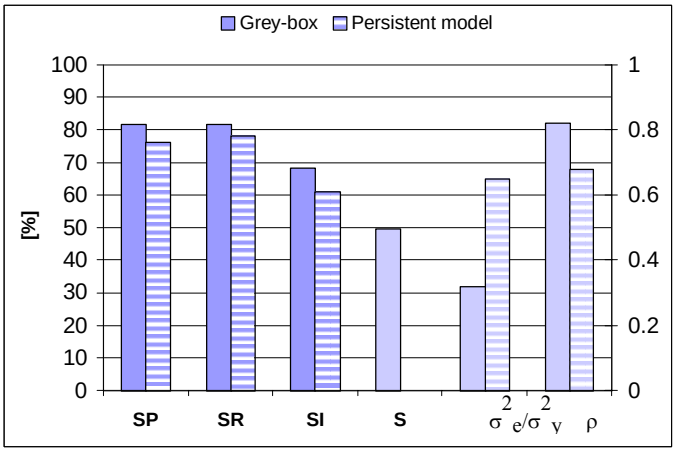

Figure 18c. P. Lambro station. O_3 skill parameters (summer 1999).

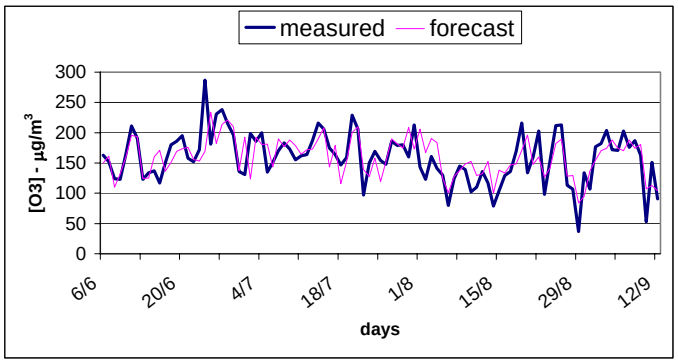

Figure 18d. P. Lambro station. O_3 real and forecast. Time series along the validation season (summer 1999).

7 Conclusions

Statistical modelling approaches have been the object of a growing interest among researchers involved in air quality modelling. This is proved by the large amount of studies published in literature, partially outlined in the first part of this chapter. This is also a consequence of the gradually increasing number of pollution data available from the monitoring networks. In addition, unlike deterministic models requiring a lot of input information (in most cases not available from monitoring networks in real time), statistical approaches represent a straightforward way to fit available pollution and meteorological time series.

As shown in the preceding paragraphs, the statistical models' performance in real cases is strictly point dependent and may require some time and trials to be properly tuned. However, a statistical approach provides an excellent and easy way to capture the global dynamics involved with the complex phenomena of air pollution, overcoming the drawbacks of using deterministic models when predictors have to work in real time (hours, few days).

Also, statistical models are good ways to approach the problem of pollution peaks prediction. The results show that the number of critical episodes correctly forecast ranges from 60% to 90%, depending on the particular target, area and statistical modelling technique considered, while the rate of false alarms ranges from 30% to 50%. This apparently large number of false alarms can be ascribed to the lack of emission data in real time and correctly reflect the fact that not all critical episodes can be interpreted using the ground-recorded meteo-chemical information. However, a false alarm rate around 40% appears to be acceptable if the total number of critical episodes along the year is limited. Moreover, there is a reasonable evidence that these results can be further improved taking into account, when available, vertical profiles of significant meteorological variables (e.g., wind, temperature, etc.). The analysis carried out also highlights the role of meteorological forecasts as inputs to reliable predictors of pollutant critical episodes.

In the last section, the implementation of statistical models in decision support systems (DSS) is suggested as one of the actual possible use in real time. The two case studies reported show how the contribution of peculiar statistical approaches may be integrated in a feedback loop system, increasing the operational effectiveness of the DSS as a whole.

Finally, it must be observed that all the examined models perform better globally than the persistent model, as indicated by the skill score indices computed as lower bound performance thresholds.

Acknowledgements

This research has been partially supported by Italian MURST and Provincia di Brescia, in the frame of EUROTRAC-2/SATURN European scientific project.

The Siracusa industrial area case study has been carried out in the framework of the European Union Project APPETISE (Contract N. IST-99-11764).

References

Abarbanel, H.D.I., *Analysis of observed chaotic data*, Springer, New York, 1996.

Anderberg, M.R., *Cluster Analysis for Applications*, Academic Press, New York, 1973.

Arena, P., Baglio, S., Castorina, C., Fortuna, L., Nunnari, G., A Neural Architecture to Predict Pollution in Industrial Areas, *Proc. of the International Conference on Neural Network - ICNN*, Vol. 4, pp. 2107-2112, 1996, Washington, USA.

Barron, A.R.,. Universal Approximation Bounds for Superposition of a Sigmoidal Function, *IEEE Trans. on Information Theory*, vol. 39, pp. 930-945, 1993.

Bezdek, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

Box, G.E.P., Jenkins, G.M., Time Series Analysis, Forecasting and Control, Holden Day Inc., S. Francisco, 1970, new edition 1976.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Time Series Analysis, Forecasting and Control, Prentice Hall, 1994.

Boznar, M., Lesjak, and P. Mlakar, A Neural Network-Based Method for Short-Term Predictions of Ambient SO_2 Concentrations in Highly Polluted Industrial Areas of Complex Terrain, *Atmospheric Environment*, Vol. 27B, No. 2, pp. 221-230, 1993.

Breiman, L. and Friedman, J.H., Estimating Optimal Transformations for Multiple Regression and Correlation, *American Statistical Association*, Vol. 80, No. 391, pp. 580-598, 1985.

Bringfelt, B., Important Factors for the Sulphur Dioxide Concentration in Central Stockholm, *Atmospheric Environment*, Vol. 5, pp. 949-972, 1971.

Briseño, H., Cipriano, A., Clustering methods applied to fuzzy model identification, *Proc. of the 7th Latin American Congress of Automatic Control*, September 9-13, Buenos Aires, pp. 760-765, 1996.

Brown, R.G., Hwang, P.Y.C., Introduction to Random Signals and Applied Kalman Filtering with Matlab exercises and solutions, Wiley, 3rd Edition, 1996.

Brusasca, G., Finzi, G., Stochastic Models for Real-Time Sulphur Dioxide Pollution Forecast around Thermal Power Plants, *in Proc. of ENVIROSOFT86. Comp Mech. Publications*, pp. 177-191, 1986.

Buishand, T.A., Kempen, G.T., Frantzen, A.J., Reijnders, H.F. and Van den Eshof, A.J. Trend and Seasonal variation of precipitation chemistry data in the Netherlands, *Atmospheric Environment*, Vol. 22(2), pp. 339-348, 1988.

Chen, S., Billings, S.A. "Neural networks for nonlinear dynamic system modelling and identification", *Int. Journal Control*, **56,** No. 2, pp. 319-346, 1992.

Cleveland, W. S., Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association*,. **74**, No. 368, pp. 829-836, 1979.

Comrie, A.C. Comparing Neural Networks and Regression Models for Ozone Forecasting, *Air & Waste Management Association*, Vol**. 47**, pp. 653-663, 1997.

Cossentino, M., Damiani, A., Gaglio, S., Raimondi, F.M., Vitale, M.C., Exploring the use of Soft-computing and artificial intelligence techniques in atmospheric pollution modelling, *Air Pollution XI*, WIT press (Brebbia, C.A. and Patania, F., Editors), 2003.

Cybenko, G. Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals and Systems*, Vol. 2, pp. 303-314, 1989.

Dorling, S.R., Foxall, R.J., Mandic, P.D., Cawley, G.C. Maximum likelihood cost functions for neural networks models of air quality data, *Atmospheric Environment,* Vol. 37, pp. 3435-3443, 2003.

Davis, J.M., Eder, B.K., Nychka, D. and Yang, Q., Modelling the effects of meteorology on ozone in Houston using cluster analysis and generalised models, *Atmospheric Environment,* 32(14/15), pp. 2505-2520, 1998.

Davis, J.M. and Speckman, P., A model for predicting maximum and 8 h average ozone in Houston, *Atmospheric Environment*, 33, pp. 2487-2500, 1999.

Dunn, J.C., A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, *Journal of Cybernetics*, 3, pp. 32-57, 1973.

Finzi G., Tebaldi G., A mathematical model for air pollution forecast and alarm in an urban area", *Atmospheric Environment*, Vol.16, 2055- 2059, 1982.

Finzi, G., Hernandez, E., Garcia, R., "The SO_2 pollution in Madrid. Part II: A comparison between two stochastic models for real-time forecast purposes", *Il Nuovo Cimento*, Vol. 6C, No.6, pp. 605-617, 1983.

Finzi, G., Bonelli, P., Bacci, G., "A stochastic model of surface wind speed for air quality control purposes", *Journal of Climate and Applied Meteorology*, Vol. 23, pp. 1354-1361, 1984.

Finzi, G., Volta, M., Nucifora, A., Nunnari, G., Real-Time Ozone Episode Forecast: a Comparison between Neural Network and Grey-Box Models, *Proc. International ICSC/IFAC Symposium on Neural Computation - NC'98*, Vienna, 1998.

Finzi, G., Volta, M., Real time urban ozone alarm system including neuro-fuzzy networks and grey box models, *Proc. Second International ICSC/IFAC Symposium on NEURAL COMPUTATION-NC2000*, ICSC Academic Press, pp. 659-663, 2000.

Finzi, G., An air quality alarm system for Milan Metropolitan Area, *proc. EUROTRAC Symposium 2000*, P.M. Midgley, M. Reuther, M. Williams (Eds.), Springer Verlag Berlin, Heidelberg, 2000.

Finzi, G., Pirovano, G., Volta, M., Gestione della Qualità dell'Aria - Modelli di Simulazione e Previsione, McGraw-Hill Libri Italia Srl, Milano, (in Italian), 2001.

Gardner, M.W., S.R. Dorling, Artificial neural networks (the multilayer perceptron), A review of applications in the atmospheric sciences, *Atmospheric Environment*, Vol. 32, No. 14, pp. 2627-2636, 1998.

Gardner, M.W., S.R. Dorling, Neural network modelling and prediction of hourly NO_x and NO_2 concentration in urban air in London, *Atmospheric Environment*, Vol. 33, pp. 709-719, 1999.

Gilbert, R.O., Statistical Methods for Environmental Pollution Monitoring, Van Nostrand Reinhold, New York, 1987.

Gori, M., Tesi, A. On the problem of local minima in backpropagatioin, *IEEE trans. Patt. Anal. Machine Intell*. Vol. PAMI-14, No. 1, pp 76-86, 1992.

Grassberger, P. and Procaccia, I., *Characterisation of strange attractors*, Physical Review Letters, Vol. 50, pp. 346, 1983.

Greig, A.J., Cawley, Dorling S., Eben, K., Fiala, A. J., Karppinen, A., Keder, J., Kolehmainen, M., Kukkonen, J., Bertucco, L., Macoun, J., Niranjan, M., Nucifora, A., Nunnari, G., Palus, M., Pelikan, E., Ruuskanen, J., Schlink, U., Air pollution episodes: modelling tools for improved smog management (APPETISE), *Proc. of the Eighth International Conference AIR POLLUTION 2000*, New Hall, Cambridge University, UK, 24 – 26, pp. 89-98, 2000.

Hartigan, J., *Clustering Algorithms*, John Wiley & Sons, New York, 1975.

Hastie, T.J. and Tibshirani, R.J., Generalized Additive Models, Chapman & Hall, London, 1986.

Hastie, T.J. and Tibshirani, R.J., Generalized Additive Models: Some Applications, *Journal of the American Statistical Association*, Vol. 82, No. 398, pp. 371-386, 1987.

Hernandez, E., Garcia, R., Finzi, G., The Sulphur Dioxide Pollution in Madrid. Part 1: A Study of the Meteorological and Statistical Aspects, *Il Nuovo Cimento*, 6C(6), pp. 595-604, 1983,.

Holland, J.H. Adaptation in Natural and Artificial System, University of Michigan Press, 1975.

Huang, G., A stepwise cluster analysis method for predicting air quality in a urban environment, *Atmospheric Environment*, Vol. 26B, pp. 349-357.

Jenkins, G.M, Watts, D.G., Spectral Analysis and Its Applications, Holden-Day, San Francisco, 1968.

Jordan, M.I., Editor, Learning in graphical Models, Mit Press, 1999.

Jorquera, H., Perez-Correa, R., Cipriano, A., Acuna, G., Short Term Forecasting of Air Pollution Episodes, Chapter 3 of Environmental Sciences and Environmental Computing, Vol. II (P. Zannetti, Editor), The EnviroComp Institute, 2004.

Kantz, H., Schreiber, T., Nonlinear Time Series Analysis, Cambridge Nonlinear Science Series 7, Cambridge University Press, 1997 (reprinted 2000, 2002).

Kaplan, D. and Glass, L., Understanding Nonlinear Dynamics, New York: Springer, 1995.

Kennel, M.B., Brown, R. and Abarbanel, H.D.I., Determining embedding dimension for phase space reconstruction using a geometrical construction, *Phys. Rev*., Vol. A 45, pp. 3403, 1992.

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G., Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentration, compared with deterministic modelling system and measurements in central Helsinki, *Atmospheric Environments,* Vol. 37, pp. 4539-4550, 2003.

Lins, H.F., Trend analysis of monthly sulfur dioxide emissions in the conterminous United States, *Atmospheric Environment*, 21, 11, pp. 2297-2309, 1987.

Lippmann, R.P., An introduction to computing with neural nets, *IEEE ASSP Magazine*, Vol. 4, No. 2, pp. 4-22, 1987.

Ljung L., System Identification Toolbox for use with MATLAB, The Math Work Inc., Natick, Massachusetts, 1991.

Maffeis, G., Prediction of carbon monoxide acute air pollution episodes. Model formulation and first application in Lombardy, *Atmospheric Environment,* Vol. 33, pp. 3859-3872, 1999.

Melli, P., Bolzern, P., Fronza, G., Spirito, A., Real-Time Control of Sulphur Dioxide Emissions from an Industrial Area, *Atmospheric Environment*, 15, 5, pp. 653-666, 1981.

Ng, C.N. and Yan, T.L., Recursive modelling and adaptive forecasting of air quality data in C.A. Brebbia, C.F., Ratto, and H. Power (Eds.), *Air Pollution VI,* Boston, Computational Mechanics Publications, pp 869-877, 1998.

Neapolitan, R.E., Learning Bayesian Networks, Prentice Hall series in artificial intelligence, 2004.

Nunnari, G., Nucifora, A., Randieri, C., The application of neural techniques to the modelling of time-series of atmospheric pollution data", *Ecological Modelling,* Vol. 111, pp. 187-205, 1998.

Nunnari G., (2000) Fuzzy models of pollutant time-series with reduced number of rules, *Neural Network World*, No. 6, pp. 983-1000.

Nunnari, G., Bertucco, L, Milio, D., Predicting Daily Average SO_2 Concentrations in the Industrial Area of Syracuse (Italy), *Proceedings of ICANNGA 2001*, 5th International Conference on Artificial Neural Networks and Genetic Algorithm, Prague, Czech Republic, April 22-25, 2001.

Nunnari, G., Modelling air pollution time-series by using wavelet functions and genetic algorithms, *Soft Computing*, Sprinter Verlag, Vol. **8,** No. 3, pp. 173-178, 2004.

Nunnari, G. Dorling, S., Schlink, U., Cawley, G., Foxall, R., Chatterton, T., Modelling SO_2 Concentration at a Point with Statistical Approaches, *Environmental Modelling and Software*, Vol. **19/10,** pp. 887-905, 2004.

Nunnari, G., Cannavò, F., A Bayesian approach to model pollution time series, Proc. o f the ANDIS 2004 International Symposium (CR-ROM), Taormina (Italy), June 2004, pp. 1-6, 2004.

Pearl, J., (2000), Causality: Models, Reasoning and Inference, Cambridge Univ. Press.

Reich, S.L., Gomez, D.R. and Dawidowski, L.E., Artificial neural network for the identification of unknown air pollution sources, *Atmospheric Environment*. Vol. 33, pp. 3045-3052, 1999.

Rumelhart, D.E., Hinton, G.E., McLelland, L., Parallel Distributed Processing Exploration in the Microstructure of Cognition. MIT Press, Cambridge, Mass. Vol. 1, pp. 45-76, 1986.

Sanchez, M.L., Pasqual, D., Ramos, C., Perez, I., Forecasting particulate pollutant concentrations in a city from meteorological variables and regional weather patterns, *Atmospheric Environment*, Vol. 24A, pp. 1509-1519, 1990.

Sarkar, D., Methods to speed-up error backpropagation learning algorithm, *ACM Computing Surveyes*, Vol. 27, No. 4, pp. 520-542, 1995.

Schlink, U., Herbarth, O. and Tetzlaff, G., A component time-series model for SO_2 data: forecasting, interpretation and modification, *Atmospheric Environment*, Vol. 31, No. 9, pp. 1285-1295, 1997.

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondrácek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M. and Doyle, M., A rigorous inter-comparison of ground-level ozone predictions, *Atmospheric Environment,* Vol. **37**, pp. 3237-3253, 2003

Shing, J., R. Jang, "ANFIS: Adaptive-Network-based Fuzzy Inference System", IEEE *trans. on System, Man and Cybernetics*, Vol. 23, pp. 665-683, 1993.

Sjoberg J., Hjalmarsson H., Ljung L., Neural Networks in System Identification, *Proceedings of the Sysid '94*, Copenhagen, Denmark, 1994.

Sjoberg, J., Liung, L., Overtraining, regularisation and searching for a minimum, with application to neural networks, *International Journal of Control*, 62(6), pp. 1391-1407, 1995.

Sluyter, R., E. van Zantvoort, Information Document concerning Air Pollution by Ozone. Overview of the situation in the European Union during the 1997 summer season, Report to the commission by European Environment Agency, European Topic Centre on Air Quality, 1997.

Sugihara, G. and May, R., Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, Vol. 344, pp. 734, 1990.

Sugeno, M. and Yasukawa, T., A fuzzy-logic based approach to qualitative modeling, *IEEE Transactions on Fuzzy Systems*, **1**, pp. 7-31.

Takens, F., Detecting strange attractors in turbulence. Lecture notes in Math. Springer, New York, Vol. 898, 1981.

Tilley, T., McBean, T.A., An application of spectrum analysis to synoptic-pollution data, *Atmos. Environment*, No. 7, pp. 793-801, 1973.

Trivikrama, S.R., Samson, P.I., Pedadda, A.R., Spectral analysis approach to the dynamics of air pollutants, *Atmos. Environment*, No. 10, pp. 375-379.

TRW, Inc. Air Quality Display Model, Prepared for NAPCA, PHS, U.S. DHEW, Washington DC, NTIS PB-189-194, 1976.

Van Aalst, R.M., F.A.A.M. de Leeuw (Eds.), National Ozone Forecasting System And International Data Exchange In Northwest Europe, *European Topic Centre on Air Quality*, 1997.

Volta, A., Nucifora, G., G. Nunnari. "Real time ozone episode forecast: a comparison between neural network and grey box models", *Proc. International ICSC/IFAC Symposium on NEURAL COMPUTATION-NC '98*, ICSC Academic Press, pp. 854-860, 1998.

Willmott, C.J., Some comments on the evaluation of model performance. *Bulletin American Meteorological Society,* 63 (11), pp. 1309–1313, 1982.

Willmott, C.J., et al., Statistics for the evaluation and comparison of models. Journal of *Geophysical Research,* 90 (C5), pp. 8995–9005, 1985.

Wood, S.N., Modelling and smoothing parameter estimation with multiple quadratic penalties, *J.R. Statistic*. Soc. B, Vol. 62, Part 2, pp. 413-428, 2000.

Young, P.C., Ng, C.N., Lane, K., Parker, D. Recursive forecasting, smoothing and seasonal adjustment of non-stationary environmental data, *Journal of Forecasting*, Vol. 10, pp. 57, 1991.

Young, P.C., Jakeman, A.J. and Post, D.A., Recent advances in data-based modelling and analysis of hydrological systems, *Water Science & Tech.* **36,** pp. 99-116, 1997.

Young, P.C., Data-based mechanistic modelling of environmental, ecological, economic and engineering systems, *Env. Modelling and Software*, **13,** pp. 105-122, 1998.

Van Aalst, R.M. and De Leeuw, F.A.A.M., National Ozone Forecasting System and International Data Exchange in Northwest Europe, *European Topic Centre on Air Quality*, 1997.

Zadeh, L.A. Fuzzy sets, *Information and Control*, **8**, pp. 338-353, 1965.

Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning, Part I, *Inf. Sci.,* **8**, pp. 199-249, 1975.

Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning, Part II, *Inf. Sci*., **8,** pp. 301-357, 1975.

Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning, Part III, *Inf. Sci*., **9**, pp. 43-80, 1976.

Zannetti, P., Air Pollution Modeling, Theories, Computational Methods and Available Software, Van Nostrand Reinhold (Ed.), New York, 1990.

Zhang, Q., Using wavelet network in non parametric estimation, *IEEE trans on Neural Networks*, **8**, pp. 227-236, 1997.

Zickus, M., Greig, A.J. and M. Niranjan, Evaluating the Variable Selection and Prediction Performance of Several Machine Learning Algorithms Applied to a PM_{10} Data Set, *Proceedings of 3rd International Conference on Urban Air Quality*, Loutraki, Greece, March 2001.

Appendix

A Survey of Software Packages for Developing Forecast Models

In this appendix, we give some information concerning the software packages for building forecasting models by using the techniques described in Section 3.

ARX, ARIMA, and ARMAX models can be identified and simulated by using the Matlab® developed by Ljung (1991).

Cyclo-stationary or *grey-box* ARMAX models can easily be implemented by using the Matlab® programming features. However, a specific tool for *cyclo-stationary* models, referred to as Winast, was developed by the book of Finzi et al. (2001) and it is in the CD as part of the referenced book.

MLP **neural networks** devoted to implement NARX model for air quality forecast can be applied using the general purpose Matlab® Neural Network Toolbox, which gives the user the possibility of using the high level graphical and training features. However, a large number of other tools are available through the Internet such as JANN (a Java Artificial Neural Network) developed at the University of Catania in the framework of the APPETISE project (IST - 99-11746). This tool is available for interactive use at the following URL: http://www.dees.unict.it/users/gnunnari/appetise/jann/index.html.

Fuzzy and *Neuro-Fuzzy* models can be implemented by using the Matlab® *Fuzzy* Toolbox

Wavelet based models can be implemented by using the Matlab® Wavelet Toolbox. A specific tool has been coded as a Matlab script by G. Nunnari.

Generalised Additive Models (GAM) and Local Prediction in Phase Space (LPH) models can be implemented by using a tool called TISEAN, which is free and available at the following URL: www.mpipks-dresden.mpg.de/~tisean/TISEAN_2.1/index.html.

Such software tool allows the analysis of time series with methods based on the theory of nonlinear deterministic dynamical systems, or chaos theory. The software has grown, with contributions from various groups, during the last few years and was put into distributable form for http://www.mpipks-dresden.mpg.de/~tisean98, held in Dresden, 11-21 Feb 1998. Some of the routines built around the programs are given in the book by Kantz and Schreiber (1997).

Kalman Filtering models can be implemented by using Mathematica®: http://www.wolfram.com/products/mathematica/index.html.

Cluster analysis and modelling can be performed with routines available in the Matlab® Fuzzy tool box, or by using one of the available statistical software packages such as Mathematica® http://www.wolfram.com/products/mathematica/index.html or XLstat http://www.xlstat.com/.

Bayesian Modelling can be performed by using a number of software tools such as the Bayes Net toolbox for Matlab® developed by K. Murphy (see the following URL: http://www.ai.mit.edu/~murphyk/Software/BNT/bnt.html or Netica http://www.norsys.com/), which is one of the world's most widely used Bayesian network development software.