Estimating Pollutant Emission into the Atmosphere: a Neural Network Approach

Crescenzio Gallo[*1], Alessandro Rinaldi, Vito Capozzi

Department of Clinical and Experimental Medicine, University of Foggia
Viale Pinto c/o OO.RR. – 71122 Foggia, ITALY
[*1]crescenzio.gallo@unifg.it

***Abstract*- A very significant issue today concerns the problem of air pollution caused mainly by human activity. The statistics show that most of the pollutants in the atmosphere are due to emissions caused by anthropogenic factors (e.g. power and industrial plants, traffic and combustion phenomena in general). In this paper we evaluate the implementation of a model using artificial neural networks to forecast short-term rate of air pollution for supporting environmental policy decisions.**

Keywords- Air Pollution; Forecasting; Artificial Neural Network

I. INTRODUCTION

The close correlation between environmental conditions and population health emerged from studies carried out jointly by the European Environmental Agency (EEA) and the European Centre for Environment and Health/World Health Organization (Eceh/Who) which identify air pollution due to emissions into the atmosphere as one of the more responsibles for the consequences on human health.

Whereas the quality of air emissions is closely related to combustion technology, within the EU five areas have been identified of major sources of pollution, listed below in order of increasing incidence:

- Energy production;
- Transport vehicle;
- Industrial production;
- Agriculture;
- Other.

The first three sectors, i.e. energy, industry and transport, originate almost total emissions into the atmosphere, as shown in Table I:

TABLE I BREAKDOWN OF POLLUTANTS EMISSION BY SECTOR (SOURCE: EEA)

Emissions	Energy	Industry	Transports
CO_2	33%	24%	24%
CO	n.a.	n.a.	69%
NO_x	<20%	13%	63%
NMVOC	n.a.	37%	47%
SO_2	60%	25%	n.a.
Particulates	40-55%	15-30%	10-25%

where NMVOC stands for “non-methane volatile organic compounds”. The origin of the emission sources of air pollution is thus essentially of two types:

1. Natural (forest fires, volcanic activity, erosion or decomposition of organic matter);
2. Anthropogenic (traffic, industrial plants, electrical or heating systems).

In this paper we want to analyze the concentrations of pollutants of both natural and anthropogenic origin in order to detect any exceedances of Air Quality Regulatory limits. The operational phases of our work start from the calculation of emission factors and the entries related to obtaining maps of concentrations for different meteorological scenarios, with the subsequent processing through neural networks.

To do so, we therefore first recall some theoretical models for the diffusion of air pollutants in Section III. Then, we examine the importance of neural networks for the analysis of air pollution (Section IV), presenting the chosen model in Section V. Finally (Section VI), we present the comparison between the obtained maps and the concentration limits imposed by law on the subject: from this comparison it was possible to identify existing criticalities, thus providing a tool to test different hypotheses of the project prior to a simulation of the same.

II. BACKGROUND

A. The Existing European Institutional Framework

In Europe, air quality regulatory limits are mainly stated in the Directive 96/62/EC, which defines:

- The objectives for the protection of human health and the environment,
- The methods for evaluation in member States,
- The acquisition of information to be made known to the population,
- The improvement and maintenance of air quality in areas of high risk of pollution.

Subsequent Legislative Decrees establish for the pollutants NO_2, NO_x, PM10, lead, benzene and CO:

- Limit values and its terms;
- Alarm thresholds;
- Tolerance margins;
- Criteria for data collection;
- Upper and lower evaluation thresholds;
- Information to the public;
- Data communication.

B. Air Quality Monitoring Networks

The air quality control system was designed as a learning tool that can provide information on the general state of air quality of the entire region and not intended solely for the verification of compliance with regulatory limits in the most critical areas. It includes fixed measurement stations located in different sites representative of typical situations of the various regions in terms of orography, meteorological conditions and presence of sources of pollutant emissions into the atmosphere.

C. Classification of Monitoring Stations

According to Directive 2001/752/EC, the monitoring stations are classified based on type of area and major sources of emission.

Type of area:

- Urban (continuously built-up area);
- Device (largely built-up area);
- Rural (which do not meet the criteria for urban).

Major sources of emissions:

- Traffic (pollution is influenced by vehicular traffic);
- Industry (pollution is influenced by industrial areas);
- Fund (pollution is influenced by all emission sources in the area in question).

III. THE IMPORTANCE OF MODELS

To study the emission of pollutants into the atmosphere by different sources (such as industrial plants, vehicles, heating systems, etc.) and air quality of a given area, you can resort to the use of diffusion models of air pollutants [1, 2, 3]. Models for the estimation of concentrations of air pollutants [4, 5] can be grouped into (see Fig. 1):

- White Box models (deterministic);
- Grey Box Linear models (stochastic);
- "Black Box" Non-linear models (of neural network type).

A. White Box Models (deterministic)

The model is well known and it is possible to build it entirely from the description of the constituent parts of the system by means of mathematical-physical laws that govern its behavior.

1) Eulerian Model

According to the Eulerian approach the behavior of pollutant concentrations is described by a differential equation (that expresses the instantaneous mass balance) relative to a coordinate system fixed in space, written for the various pollutants we are interested in simulating.

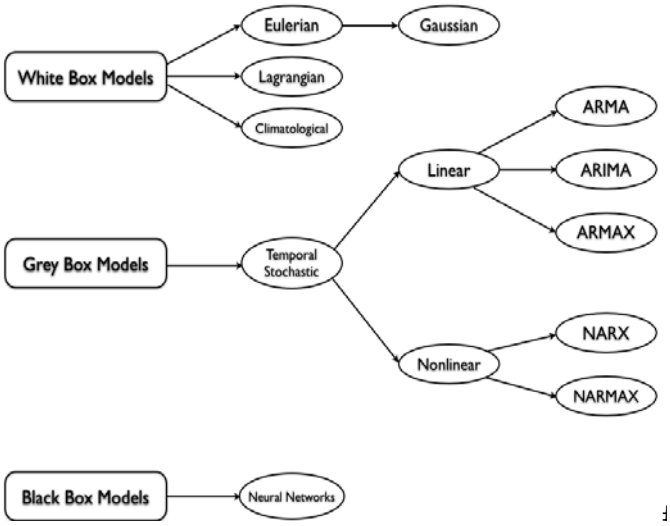

Fig. 1 Models for estimating concentrations of air pollution

If we consider N polluting species, for a generic i-th substance such equation turns out to be (once the molecular diffusion has been neglected) the following:

$$\frac{\partial c_i}{\partial t} + \frac{\partial (uc_i)}{\partial x} + \frac{\partial (vc_i)}{\partial y} + \frac{\partial (wc_i)}{\partial z} = R_i(c_1, c_2, \ldots, c_N) + E_i - S_i \tag{1}$$

In this equation the term R_i, which a priori depends on the instantaneous concentration of all substances, synthesizes all the chemical kinetics considered, while the term E_i represents the sources of pollutants and the term S_i takes into account globally for all removal processes taking place in the atmosphere. It should be noted that the generic concentration c_i is a molar concentration (expressed as moles $\cdot$ m^{-3}). Obviously there are many equations as there are polluting species, and the system is closed if the wind field is known.

Since it is virtually impossible to use this equation directly, we prefer the Reynolds hypothesis [6] that each variable (also the concentration of various pollutants, therefore) is equal to the superposition of a slowly time-varying mean value and a turbulent fluctuation with zero mean. Applying this hypothesis and by averaging, after some simplifications [3] we obtain the following relationship:

$$\frac{\partial \overline{c}_i}{\partial t} + \left(\overline{u}\frac{\partial \overline{c}_i}{\partial x} + \overline{v}\frac{\partial \overline{c}_i}{\partial y} + \overline{w}\frac{\partial \overline{c}_i}{\partial z} \right) - \left(\overline{u}\frac{\partial \overline{u'c_i'}}{\partial x} + \overline{v}\frac{\partial \overline{v'c_i'}}{\partial y} + \overline{w}\frac{\partial \overline{w'c_i'}}{\partial z} \right) = R(\overline{c}_1, \overline{c}_2, \ldots, \overline{c}_N) + E_i - S_i \tag{2}$$

where

- The first term on the left represents the time evolution of the average concentration,
- The second term on the left represents the transport of the pollutant caused by the mean motion of air masses (advection),
- The third term represents the interaction with the atmospheric turbulence and is indicated by the term turbulent diffusion,
- The first term on the right (which is globally the chemical reaction) depends, with the limitations outlined by Seinfeld and Pandis [3], on the mean concentration of pollutant species.

Again there are many equations as the pollutants considered, but - in this case - even if we assume to know the average wind field, the system is not closed for the presence of covariance between the components of wind speed and concentration, that is of turbulent flows.

Eulerian Models: Gaussian

This is a model that refers to a Gaussian scheme [6] for solving the diffusion equation. In essence, the Gaussian method defines the temporal evolution of physical quantities, and in the case of pollutant concentrations, measured at a fixed point of the flow field. The equation is:

$$C = \frac{Q}{2\pi u \sigma_y \sigma_z} \cdot w \cdot (v_1 + v_2 + v_3) \tag{3}$$

where

- $w = e^{-\frac{y^2}{2\sigma_y^2}}$ (crosswind dispersion parameter)
- $v_1 + v_2 + v_3$ (vertical dispersion parameter)
- $v_1 = e^{-\frac{(z-H)^2}{2\sigma_z^2}}$ (vertical dispersion with no reflection)
- $v_2 = e^{-\frac{(z+H)^2}{2\sigma_z^2}}$ (vertical dispersion for reflection from the ground)
- $v_3 = \sum_{k=1}^{\infty}\left[e^{-\frac{(z-H-2kL)^2}{2\sigma_z^2}} + e^{-\frac{(z+H+2kL)^2}{2\sigma_z^2}} + e^{-\frac{(z+H-2kL)^2}{2\sigma_z^2}} + e^{-\frac{(z-H+2kL)^2}{2\sigma_z^2}} \right]$ (vertical dispersion for reflection from an inversion aloft)
- C = concentration of emissions, in g/m^3, at any receptor located x meters downwind from the emission source point, y meters crosswind from the emission plume centerline, z meters above ground level
- Q = source pollutant emission rate, in g/s
- u = horizontal wind velocity along the plume centerline, m/s
- H = height of emission plume centerline above ground level, in m
- σ_z = vertical standard deviation of the emission distribution, in m
- σ_y = horizontal standard deviation of the emission distribution, in m
- L = height from ground level to bottom of the inversion aloft, in m

Figure 2 shows a three-dimensional concentration profile of a pollutant emitted from a point source in a coordinate system oriented along the average direction of the wind. When in the atmosphere there are very unstable conditions the air flow inside the boundary layer is turbulent and ascending and descending movements are present. This causes the pollutants emitted from a source to be carried quickly to the upper layers of the boundary layer or close to the surface. The ascending movements have higher speed than the descending transport and cover smaller area.

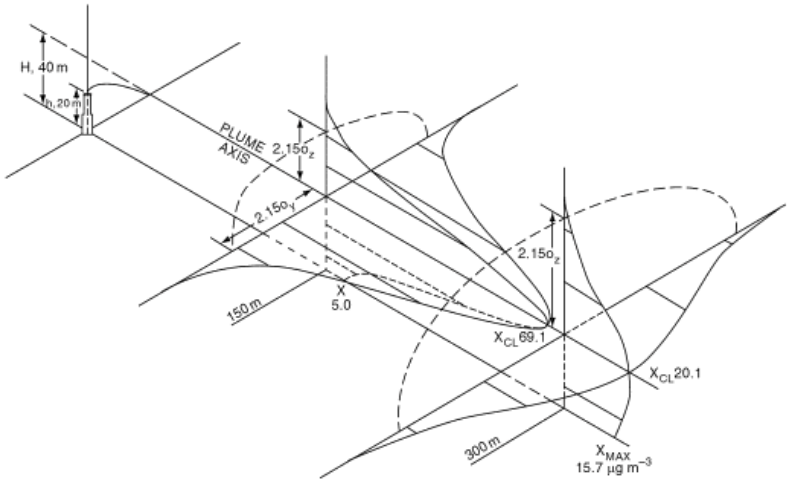

Fig. 2 Gaussian description of the emission of pollutants from a point source

Therefore pollutants have higher probability to be located in a descending air movement, with the main axis of the plume moving to the surface as a final result. In total the plume moves to the surface or to the base of the boundary layer. This fact is not considered in the Gaussian model and requires caution for its application in very unstable conditions.

2) Lagrangian Model

In the Lagrangian approach, instead, the particle motion is expressed in a mobile coordinate system that follows the flow of dispersed concentrations in the atmosphere. The study of pollutant particle motion is at the base of the fundamental Lagrangian equation:

$$\overline{C}(x,t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(x,t \mid x',t')\overline{C}(x_0,t_0)\mathrm{d}x_0\mathrm{d}t_0 + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(x,t \mid x',t')S(x',t')\mathrm{d}x'\mathrm{d}t' \quad (4)$$

Whereas the Eulerian model describes the concentration of a pollutant dispersed into the atmosphere through a differential equation (called *balance equation* or *mass conservation*) relative to a coordinate system fixed in space, the Lagrangian model describes the statistical properties of the concentration in terms of the motion properties of a particle along a trajectory [7].

3) Climatological (or long-term*) Model*

This model takes into account the variations of meteorological variables and emissions which in the previous models are assumed stationary. These variables include:

wind speed class denoted by v_n (n=1, …, 5);

wind direction class indicated by Θ=1,…, 16;

stability class denoted by S = [A, B, C, D, E, F].

B. Grey Box Models (stochastic)

The model is partly built on the basis of mathematical laws, but some parameters are unknown and must be determined from the observed data. These models can predict the temporal evolution of the concentration of a particular kind of pollutant considered a random variable.

1) Linear Models

The temporal linear stochastic models are distinguished in:

- ARMA (AutoRegressive Moving Average);
- ARIMA (AutoRegressive Integrated Moving Average);
- ARMAX (AutoRegressive Moving Average with eXogenous inputs).

The ARMA, ARIMA and ARMAX models are linear and can be briefly described as the composition of a transfer function $G(z)$, which sets out the link between the exogenous variable $u(.)$ and the output $y(.)$ in the absence of noise, and a transfer function $W(z)$ powered by a remote white noise $\xi(.)$ whose output $v(.)$ quantifies the error in describing the system with the function $B(z)$. The resulting relation becomes (see also Fig. 3):

$$y(t) = G(z)u(t-1) + W(z)\xi(t) \tag{5}$$

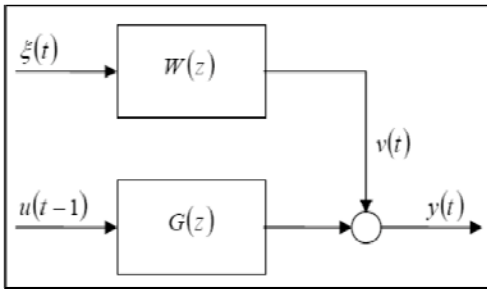

Fig. 3 Block diagram of a linear model

ARMA models

The ARMA models are widely used stochastic models for predicting concentrations of pollutants in the atmosphere, providing excellent results when used for long-term forecasts, while not satisfactory when used for short-term forecasts and in the presence of non-stationary situations. Formally, an ARMA(p, q) model is represented as follows:

$$\underbrace{X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p}}_{\text{autoregressive component}} = \underbrace{a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}}_{\text{moving average component}} \tag{6}$$

ARIMA models

The ARIMA models (also defined on d-th differences of a stochastic process X_t) are used in the analysis of air quality data, as a feature often found in time series data of air pollution is the non-stationarity. The statistical properties of the series so vary if you translate the origin of the time axis. The expanded form of an ARIMA(p, d, q) is given by the following equation:

$$\nabla^d X_t - \phi_1 \nabla^d X_{t-1} - \phi_2 \nabla^d X_{t-2} - \ldots - \phi_p \nabla^d X_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \tag{7}$$

where $\nabla^d = (1-B)^d$ denotes the d-th difference of a process. If d=1 we talk about *first order* differences; for d=2 we talk about *second order* differences.

ARMAX models

ARMAX models are obtained from ARMA models by adding appropriate exogenous variables, which can provide useful information to the prediction of a phenomenon. An ARMAX model can be seen as an improvement of an ARMA model, as it reduces the variance with the introduction of an exogenous variable in the model, which lowers the erratic component. An ARMAX model is represented as follows:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} + \psi_1 u_{T_1} + \ldots + \psi_m u_{T_m} \quad (8)$$

where $\psi(\cdot)$ represents a polynomial of order m, u_{T_m} represents the m-th exogenous input and T_m is the time delay between the output and the m-th input.

2) Non-Linear Models

A non-linear stochastic system with a discrete-time input and output can be described, under some weak assumptions, from the general NARMAX model (see Fig. 4):

$$y(t) = f\left(y(t-1),\ldots,y(t-n_y),u(t-1),\ldots,u(t-n_u),\xi(t-1),\ldots,\xi(t-n_\xi)\right) + \xi(t) \quad (9)$$

where

- y(t) represents the output at time t;
- n_y, n_u and n_ξ are the maximum output, input and noise delays (called *orders of the model*);
- $\xi(\cdot)$ is a white noise;
- $f(\cdot)$ is a nonlinear function.

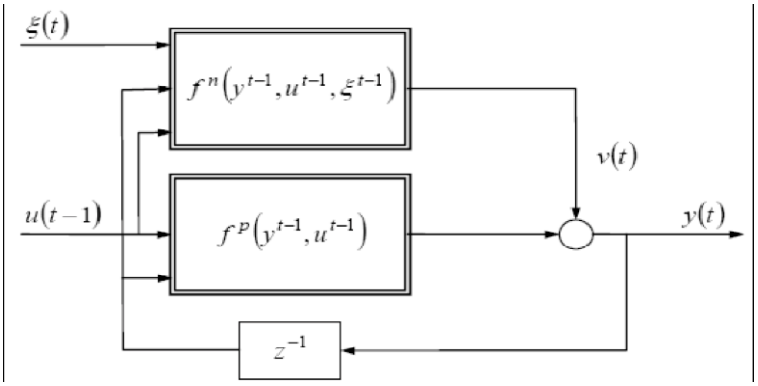

Fig. 4 Block diagram of a NARMAX nonlinear model

C. "Black Box" Non-linear Models (of neural network type)

The model is constructed without any prior information on the physical environment, using "default" facilities characterized by good flexibility and a good functioning in practice. The non-linear models of neural network type that make predictions in the time domain can be of type:

- NARX (Non linear AutoRegressive with eXogenous inputs);
- NARMAX (Non linear AutoRegressive Moving Average with eXogenous inputs).

1) NARX Models

The non-linear autoregressive models with exogenous inputs are obtained by considering only the regressors $y(t\text{-}k)$ and $u(t\text{-}k)$ of (9) where

- $y(t\text{-}k)$ represents the output at time $(t\text{-}k)$;
- $u(t\text{-}k)$ represents the exogenous variable introduced in the model at time $t\text{-}k$.

The output is given by a nonlinear function that refers to the previous output and exogenous variable values, all powered by the white noise $\xi(t)$. In other words, in NARX models only the process model appears, while the noise model is null:

$$y(t) = f^p\left(y(t-1),\ldots,y(t-n_y),u(t-1),\ldots,u(t-n_u),\xi(t-1),\ldots,\xi(t-n_\xi)\right) + \xi(t) \quad (10)$$

2) NARMAX Models

The non-linear autoregressive moving average with exogenous inputs models (NARMAX) are obtained by considering the

regressors $y(t\text{-}k)$, $u(t\text{-}k)$ and $\xi(t\text{-}k)$ where $\xi(t\text{-}k)$ represents the remote white noise at time $t\text{-}k$. Therefore, the output at time t is given by the following relationship:

$$y(t) = f\left(y(t-1),\ldots,y(t-n_y),u(t-1),\ldots,u(t-n_u),\xi(t-1),\ldots,\xi(t-n_\xi)\right)+\xi(t) \tag{11}$$

NARMAX models are a generalization of the NARX model, which allows a better characterization of the noise. In these models both the process and the noise model appear:

$$\begin{aligned} y(t) = {} & f^p\left(y(t-1),\ldots,y(t-n_y),u(t-1),\ldots,u(t-n_u)\right)+ \\ & f^n\left(y(t-1),\ldots,y(t-n_y),u(t-1),\ldots,u(t-n_u),\xi(t-1),\ldots,\xi(t-n_\xi)\right)+\xi(t) \end{aligned} \tag{12}$$

IV. THE IMPORTANCE OF NEURAL NETWORKS FOR THE ANALYSIS OF AIR POLLUTION

Air pollution is an extremely complex phenomenon to be treated. The classical methods are not considered capable of efficiently modeling complex phenomena such as meteorology and air pollution because, usually, they make approximations or too rigid schematizations.

Instead of the classical methods, you can then try to model these complex phenomena by adopting architectures that simulate the operation of biological neurons in our nervous system such as artificial neural networks.

A. Concept of Artificial Neural Network

A neural network is defined as "a set of interconnected artificial neurons" that try to simulate the operation of biological neurons in the human brain (see Fig. 5). The *advantages* of neural networks are:

Robustness

A neural network is resistant to noise, i.e. it is able to continue to give a correct answer even if some of its connections are deleted (damaged) or if noise is added to the input signal, the transmission channels or function activation of nodes.

Flexibility

A neural model can be used for a large number of different purposes: it does not need to know the properties of the specific application domain, because it learns by experience.

Generalization

A neural network, trained on a limited number of examples, is able to produce an adequate response to input patterns never seen before, but which nevertheless have some resemblance to the examples presented during the training.

Content-based retrieval

Artificial neural networks are able to recover their memories based on content from data incomplete, similar or corrupted by noise. As in biological nervous systems, we see that many faculties have parallel pathways which can compensate for damage to one of the routes through the brain.

The *disadvantages* of neural networks are:

- Learning times longer than the linear models;
- The design of the model is empirical (the number of hidden neurons is found out by trial and error as well as the activation function);
- Need for a very large training dataset;
- The neural network acts as a *black box* as it does not reveal, in readable terms, the relations between the input and the equivalent output.

B. Types of Neural Networks

The principal types of neural networks [8] are:

- Associative Memories (or Hopfield Networks);
- Kohonen Maps (SOM);
- Radial Basis Function (RBF) Networks;
- Feed-forward (or Multi Layer Perceptron, MLP) Networks.

1) Associative Memories

This type of network (see Fig. 6) is able to learn associations between patterns (complex set of data as the set of pixels of

an image) so that by submitting an input pattern to the network you get the associated output pattern, even if the input is contaminated by noise or only partial. The typical transfer function of an associative memory is illustrated in Fig. 7.

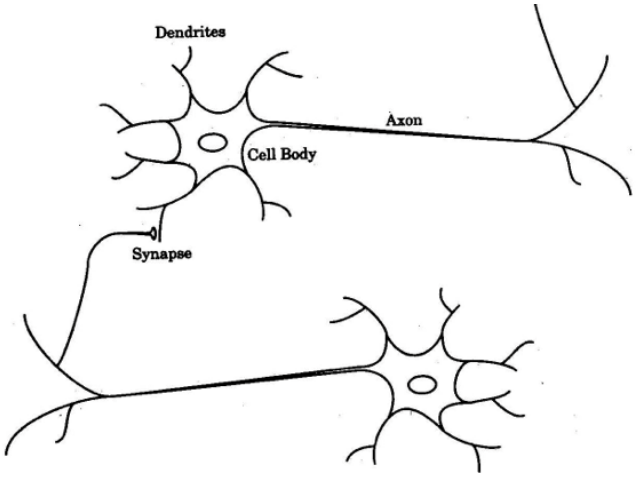

Fig. 5 The biological neuron

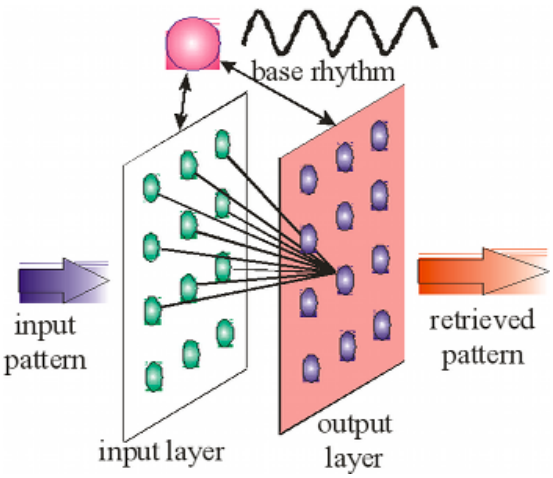

Fig. 6 Associative neural network schema

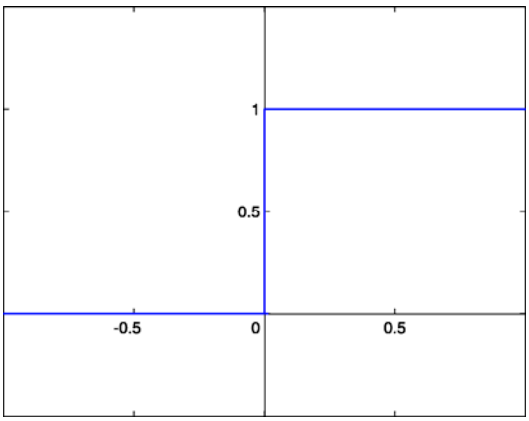

Fig. 7 Step function in an associative network

2) Kohonen Self-Organizing Maps

These are network models (see Fig. 10) with unsupervised competitive learning that perform data clustering, i.e. group similar data in the same category.

3) Radial Basis Function Networks

A RBFN (see Fig. 9) consists of neurons with local field, which activate when the input belongs to a specific region. The learning, of hybrid type (typically sigmoid for the output layer and Gaussian for hidden neurons), allows for interpolation and classification.

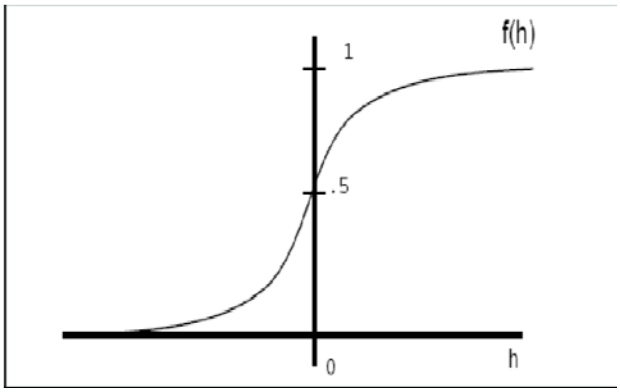

Fig. 8 The sigmoid (or Fermi) transfer function

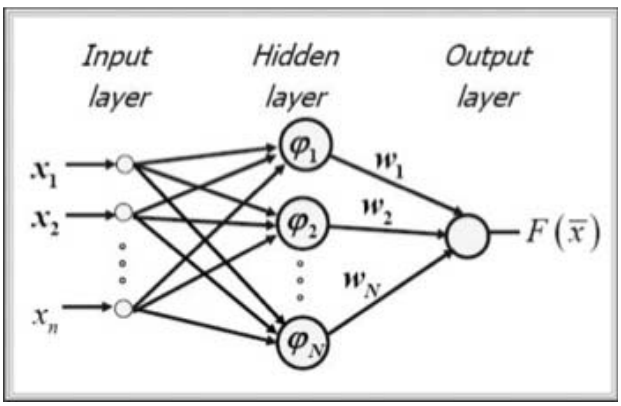

Fig. 9 A Radial Basis Function network

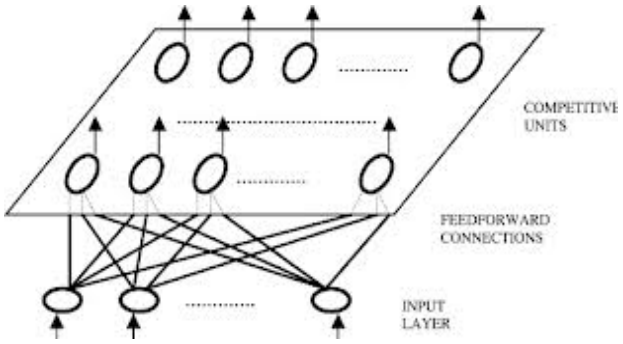

Fig. 10 A bidimensional Kohonen map

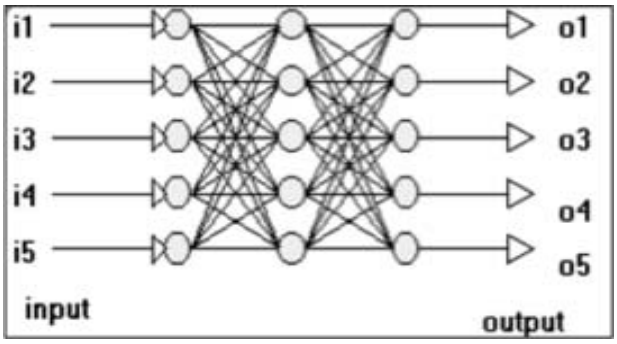

Fig. 11 A Multi-Layer Perceptron network

4) Multi Layer Perceptron

MLPs are networks able to learn the input-output function based on the examples provided in the learning phase. The transfer function of a MLP is a "sigmoid" (also known as Fermi function), as illustrated in Fig. 8. After the learning phase, the network is able to provide an output in response to an input even different from those used in the training examples. The neurons are arranged in layers and the information flow is unidirectional (see Fig. 11). A MLP network can be defined as *static*, because the outputs produced by the network depend solely on the inputs, then the network has no memory of previous computations.

5) Feed-Forward Networks

Feed-forward networks [9] receive, for each node, only signals from neurons in the previous layer and thus, the flow of information goes in one direction, from input nodes to output, and the network graph contains no cycles (see Fig. 12).

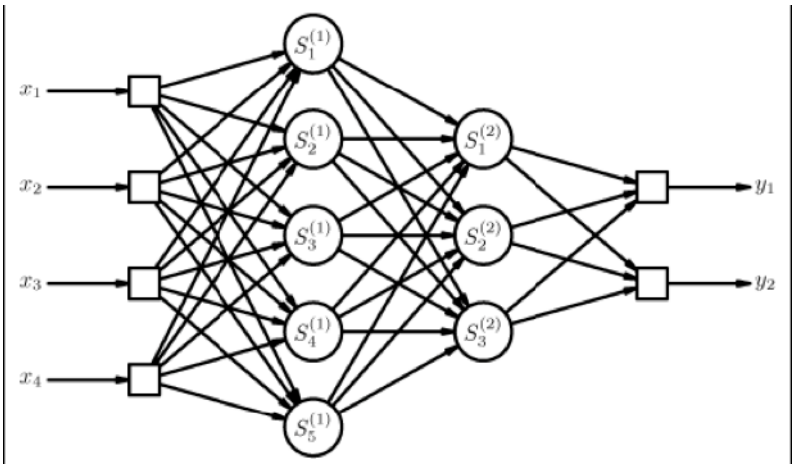

Fig. 12 A Multi-Layer Feed-Forward network

6) Feedback (or recurrent*) Networks*

In Feedback networks, as opposed to Feed-forward, neurons receive their input signals from neurons of the same and previous layers. These networks are defined *dynamical* as the output provided by a network at a given time does not only depend on the input stimuli provided at the considered time, but also on the past history (memory of previous computations).

V. CHOOSING THE ARTIFICIAL NEURAL NETWORK MODEL

In order to estimate pollutant emission into the atmosphere, a neural network approach has been chosen with:

- Feed-forward (static) structure;
- Multi Layer Perceptron architecture;
- Levemberg-Marquardt learning algorithm;
- Sigmoid activation function.

The schematic representation of the chosen artificial neural network of Feed-forward/MLP type is illustrated in Fig. 13.

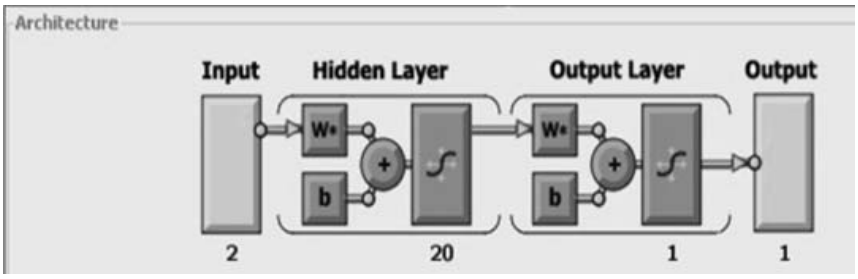

Fig. 13 The chosen artificial neural network

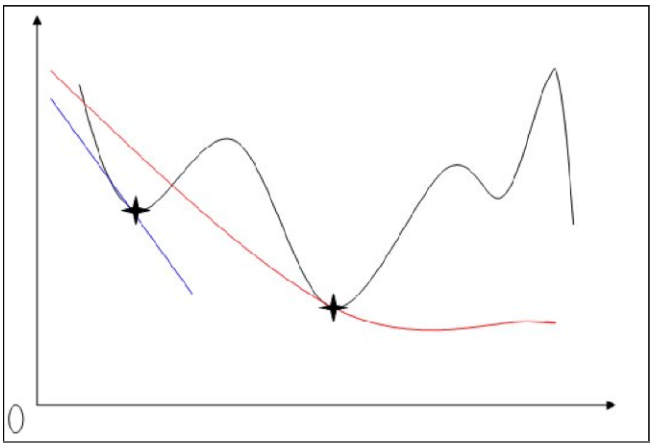

Fig. 14 The Levemberg-Marquardt learning algorithm

A. Feed-Forward Structure

The choice of using a Feed-forward network is essentially linked to a number of advantages over feedback networks:

- Advantages in terms of *efficiency*, i.e. a faster convergence towards local minima than feedback (recurring) networks;
- Less parameters than in feedback networks;
- Feed-forward networks are the most studied in the literature with respect to feedback networks and are used in many application domains;
- Compared to RBF (Radial Basis Function) networks, that require a more complex design before the training phase, they require an easier design in less time.

B. Multi-Layer Perceptron Architecture

The benefits of a Multi Layer Perceptron are:

- It can exceed the limits imposed by the linear separability (such as those in the simple perceptron);
- Its data generalization ability (doing approximations for inputs not previously observed);
- Its high learning ability (good fitting response to the data submitted to the network);
- Its ability to act as a universal approximator, in the sense that any continuous function can be approximated by a neural network with an arbitrarily high level of precision.

C. Levemberg-Marquardt Learning Algorithm

For the training phase the choice fell on the Levenberg-Marquardt (LM) algorithm of iterative down-error type, because the classic Error Back Propagation approach has a rather slow convergence to an absolute minimum as it makes use of the gradient descent optimization method (see Fig. 14). The LM algorithm also uses the information on the error function's hessian without calculating it explicitly: it is therefore particularly fast when the number of inputs is not high.

D. Sigmoid Activation Function

The choice of the sigmoid activation function (see Fig. 8) is due to three important factors:

1. It is a continuous function and therefore is everywhere differentiable;
2. It is a nonlinear function;
3. It is limited to [0, 1] and this translates into benefits in terms of learning.

Its analytical form is:

$$y_k = F(\text{Net}_k) = \frac{1}{1+e^{-\text{Net}_k}} \tag{13}$$

where

- y_k (output)
- Net_k (net input) = u_k - θ_k
- u_k (activation) $= \sum_{i=1}^{n_k} x_i w_{ki}$
- θ_k (threshold)
- x_i (input)
- w_{ki} (weights)
- n_k (number of connected neurons)

E. Characteristics of the Chosen Neural Network

In the Feed-forward/MLP chosen neural network several features have been defined, such as:

Number of hidden layers

set to one, as it has been shown that a neural network with one hidden layer is able to better approximate any continuous function;

Number of hidden nodes

set to 20 for empirical reasons, to avoid the phenomenon of *overfitting* (principle of structural risk minimization);

Number of input data used for training

set to half the input dataset, in order for the network to gain a good ability to "generalize" during the recognition operations.

In our work, these parameters have been defined experimentally (as shown in next section), by trial and error after various tests. The data evaluation indicators used are

Mean Square Error: [1]

$$\text{MSE}^{(p)} = \frac{1}{m}\sum_{i=1}^{m}\left[e_i^{(p)}\right]^2 = \frac{1}{m}\sum_{i=1}^{m}\left[t_i^{(p)} - y_i^{(p)}\right]^2 \tag{14}$$

Correlation coefficient:

$$\rho_{ty} = \frac{\sum_{i=1}^{m}(t_i - \bar{t})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{m}(t_i - \bar{t})^2 (y_i - \bar{y})^2}} = \frac{\text{Cov}(T,Y)}{\sqrt{\text{Var}(T)\text{Var}(Y)}} \tag{15}$$

where

- t_i represents the desired targets;
- y_i is the output obtained from the network;
- (p) represents the input pattern given to the network for training.

VI. APPLICATION OF THE FORECASTING MODEL

In this work we have chosen a neural network model of NARX (Nonlinear AutoRegressive with eXogenous inputs moving average, see Fig. 15) type in order to make predictions [10] in the time domain of certain types of pollutants detected by a control unit for monitoring air quality.

[1] Its square root provides an additional statistical index, the so-called Root Mean Square Error (RMSE) which corresponds to the internal variance given by the ratio between the inner deviance (or deviance within groups) and the total population.

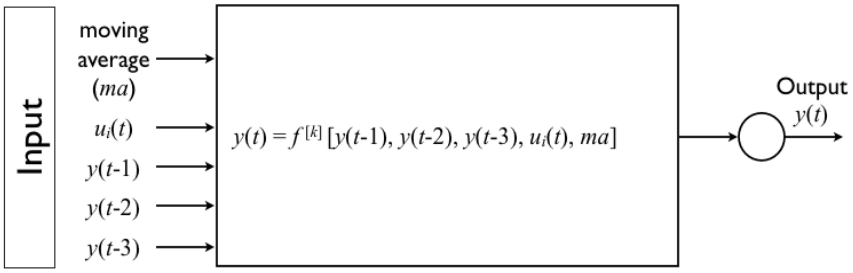

Fig. 15 Block diagram of the NARX adopted model

In order to test our forecasting model, we started from a series of pollutants detected by the control units: sulphur dioxide (SO_2), nitrogen oxide (NO_x), nitrogen monoxide (NO), nitrogen dioxide (NO_2), carbon monoxide (CO), particulate matter (PM10). From these, pollutants analyzed in the model are sulphur dioxide (SO_2) and nitrogen oxide (NO_x), with the following weathering detected: wind direction, wind speed, temperature, rain. The results of the most significant experiments are detailed in the next subsections.

A. Experiment #1 – NARX 1

The followed mathematical model is:

$$y(t) = f\left(t, u(t), y(t-1), \text{mov.avg}\left[y(t-1), y(t), y(t+1)\right]\right) \tag{16}$$

The experimental results are detailed in Table II. The training has been made in MATLAB using the function `trainlm`, a network training function that updates weight and bias values according to Levenberg-Marquardt optimization. Its results are shown in Fig. 16. The regression analysis and the correlation between achieved outputs and desired targets are illustrated in Fig. 17.

TABLE II EXPERIMENT #1 – NARX 1

Experiment	SO_2 – NARX 1
Input	Norm.Time, Weathering, $SO_2(t\text{-}1)$, mov.avg[$SO_2(t\text{-}1)$, $SO_2(t)$, $SO_2(t\text{+}1)$]
Target	$SO_2(t)$
Hidden neurons	20

	%	No. of samples	MSE	ρ
Training	60%	10349	0.198499	0.936143
Validation	20%	3449	0.516760	0.856899
Testing	20%	3449	0.300849	0.914537

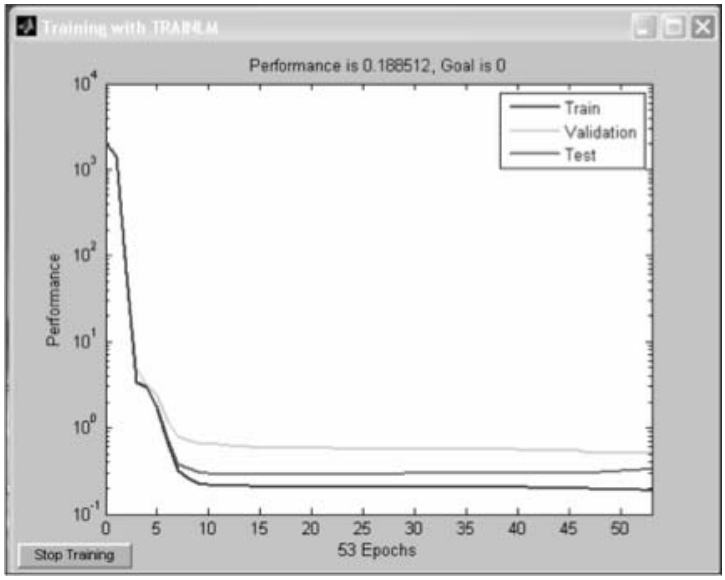

Fig. 16 Experiment #1 – Training results

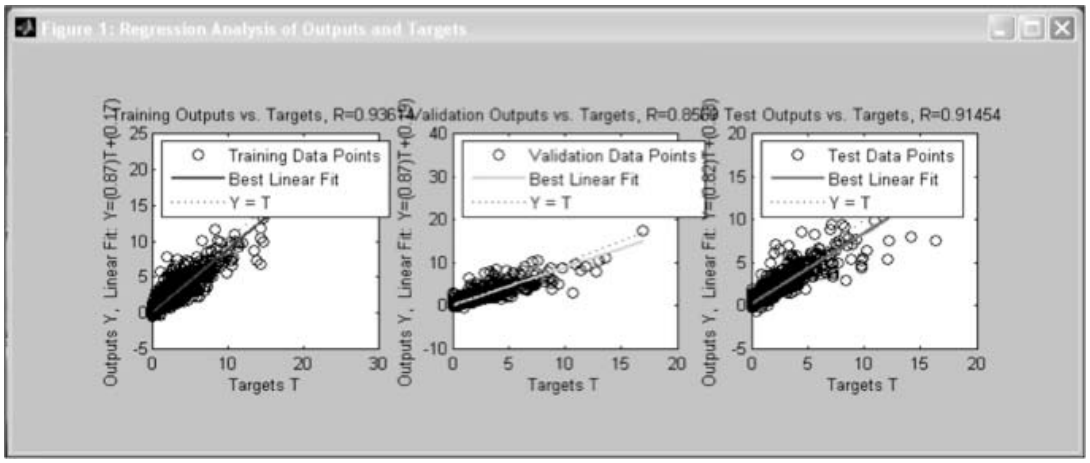

Fig. 17 Experiment #1 – Regression analysis and output-target correlation

B. Experiment #2 – NARX 2

The adopted mathematical model is:

$$y(t) = f\left(t, u(t), y(t-1), y(t-2), \text{mov.avg}\left[y(t-2), y(t-1), y(t)\right]\right) \tag{17}$$

The experimental results are detailed in Table III. The training in MATLAB with the function `trainlm` produced the results shown in Fig. 18. The regression analysis and the correlation between achieved outputs and desired targets are illustrated in Fig. 19.

TABLE III EXPERIMENT #2 – NARX 2

Experiment	SO_2 – NARX 2
Input	Norm.Time, Weathering, $SO_2(t\text{-}1)$, $SO_2(t\text{-}2)$, mov.avg[$SO_2(t\text{-}2)$, $SO_2(t\text{-}1)$, $SO_2(t)$]
Target	$SO_2(t)$
Hidden neurons	20

	%	No. of samples	MSE	ρ
Training	60%	10348	0.476395	0.837379
Validation	20%	3449	0.597655	0.804140
Testing	20%	3449	0.570344	0.839577

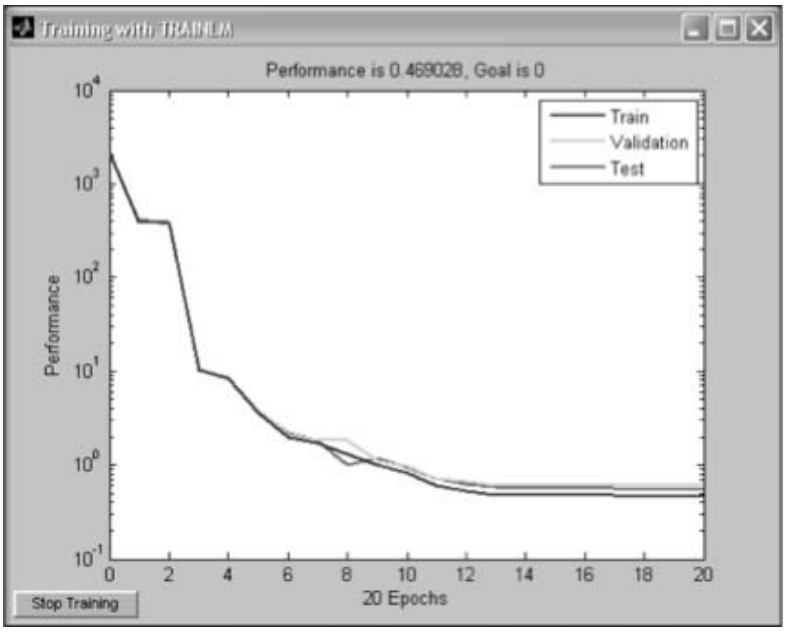

Fig. 18 Experiment #2 – Training results

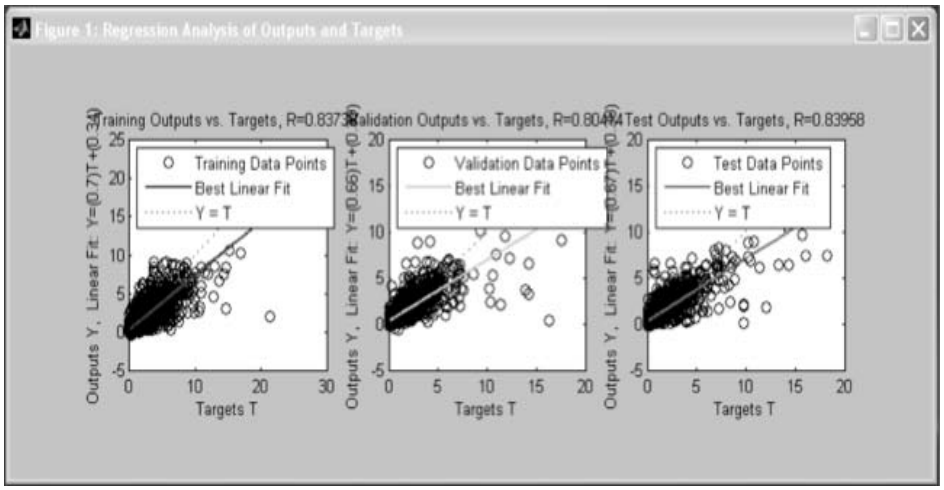

Fig. 19 Experiment #2 – Regression analysis and output-target correlation

C. Experiment #3 – NARX 3

The adopted mathematical model is:

$$y(t) = f\left(t, u(t), y(t-1), y(t-2), y(t-3), \text{mov.avg}\left[y(t-3), y(t-2), y(t-1)\right]\right) \tag{18}$$

The experimental results are detailed in Table IV. The training results shown in Fig. 20. The regression analysis and the correlation between achieved outputs and desired targets are illustrated in Fig. 21.

TABLE IV EXPERIMENT #3 – NARX 3

Experiment	SO_2 – NARX 3
Input	Norm.Time, Weathering, $SO_2(t\text{-}1)$, $SO_2(t\text{-}2)$, $SO_2(t\text{-}3)$ mov.avg[$SO_2(t\text{-}3)$, $SO_2(t\text{-}2)$, $SO_2(t\text{-}1)$]
Target	$SO_2(t)$
Hidden neurons	20

	%	No. of samples	MSE	ρ
Training	60%	10348	0.494369	0.844331
Validation	20%	3449	0.533465	0.810828
Testing	20%	3449	0.573024	0.813115

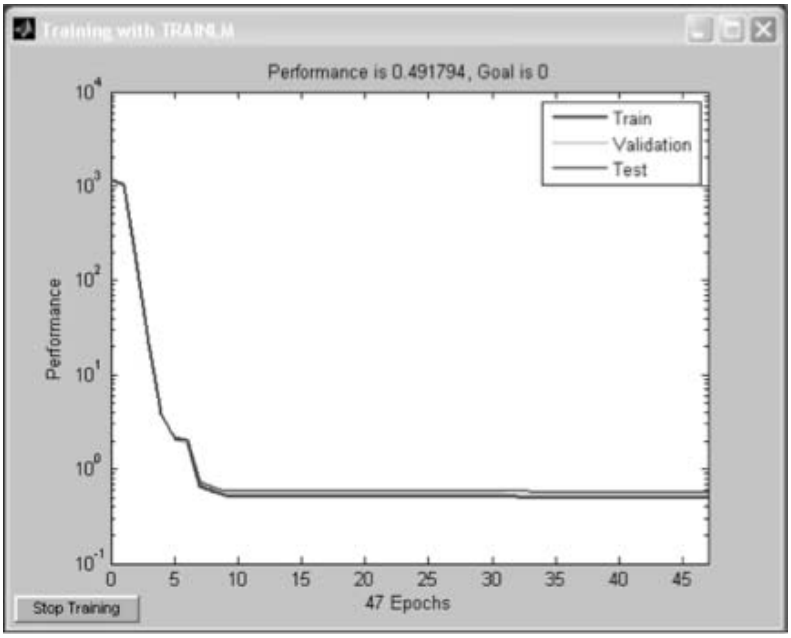

Fig. 20 Experiment #3 – Training results

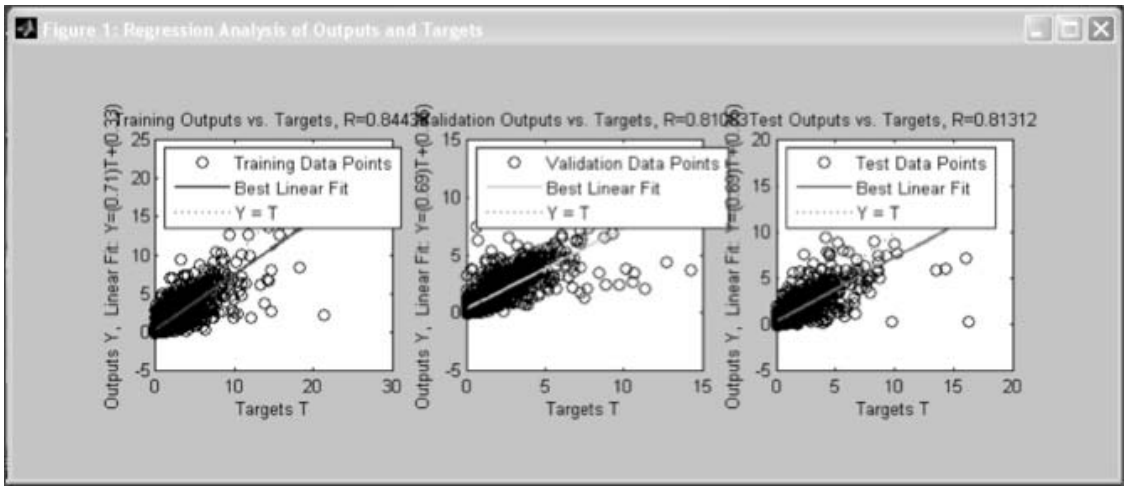

Fig. 21 Experiment #3 – Regression analysis and output-target correlation

D. Experiment #4 – NARX 1

The adopted mathematical model is:

$$y(t) = f\left(t, u(t), y(t-1), \text{mov.avg}\left[y(t-1), y(t), y(t+1)\right]\right) \tag{19}$$

The experimental results are detailed in Table V. The training results are shown in Fig. 22. The regression analysis and the correlation between achieved outputs and desired targets are illustrated in Fig. 23.

TABLE V EXPERIMENT #4 – NARX 1

Experiment	NO_x – NARX 1
Input	Norm.Time, Weathering, $NO_x(t\text{-}1)$, mov.avg[$NO_x(t\text{-}1)$, $NO_x(t)$, $NO_x(t\text{+}1)$]
Target	$NO_x(t)$
Hidden neurons	20

	%	No. of samples	MSE	ρ
Training	60%	10212	143.7393	0.899700
Validation	20%	3404	208.24843	0.900660
Testing	20%	3404	155.09356	0.910463

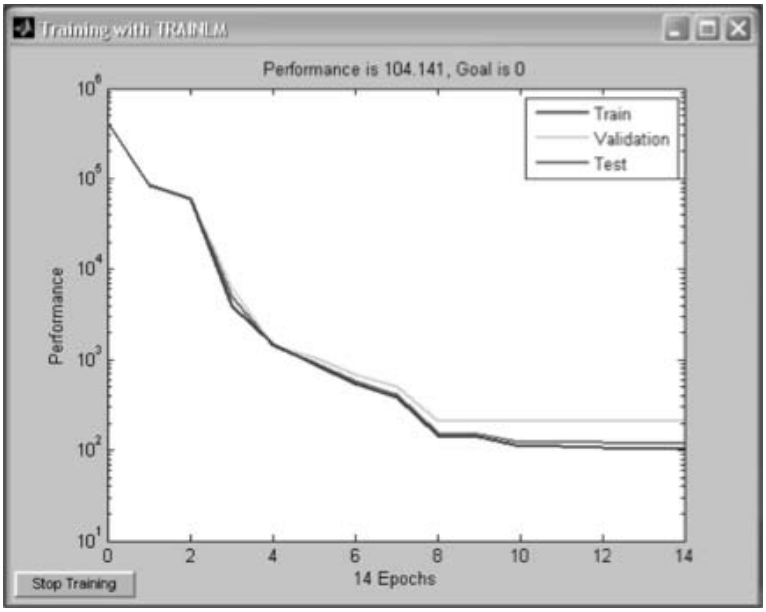

Fig. 22 Experiment #4 – Training results

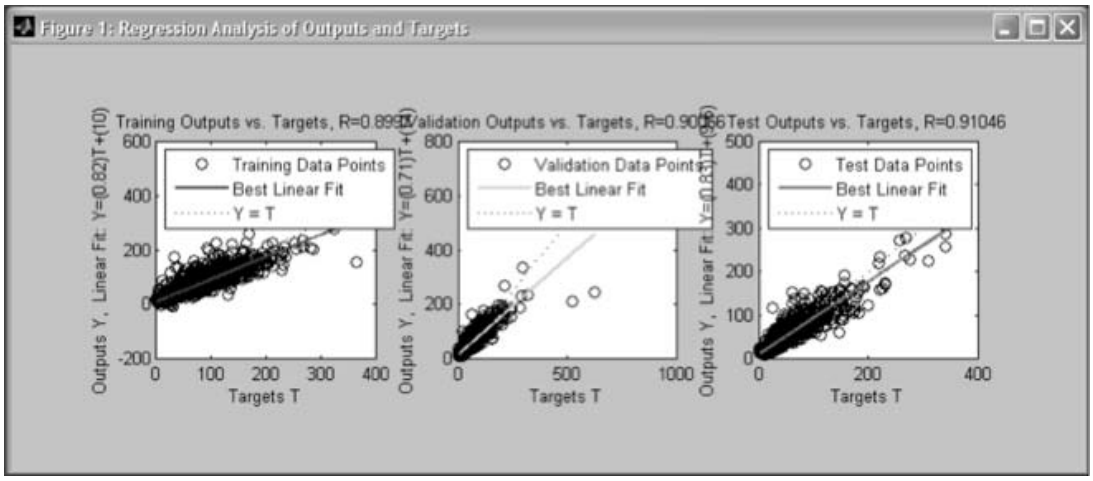

Fig. 23 Experiment #4 – Regression analysis and output-target correlation

E. Experiment #5 – NARX 2

The followed mathematical model is:

$$y(t) = f\left(t, u(t), y(t-1), y(t-2), \text{mov.avg}\left[y(t-2), y(t-1), y(t)\right]\right) \tag{20}$$

The experimental results are detailed in Table VI. The training results are shown in Fig. 24. The regression analysis and the correlation between achieved outputs and desired targets are illustrated in Fig. 25.

TABLE VI EXPERIMENT #5 – NARX 2

Experiment	NO_x – NARX 2
Input	Norm.Time, Weathering, $NO_x(t\text{-}1)$, $NO_x(t\text{-}2)$, mov.avg[$NO_x(t\text{-}2)$, $NO_x(t\text{-}1)$, $NO_x(t)$]
Target	$NO_x(t)$
Hidden neurons	20

	%	No. of samples	MSE	ρ
Training	60%	10212	129.45675	0.907814
Validation	20%	3404	174.98410	0.885201
Testing	20%	3404	164.88923	0.905966

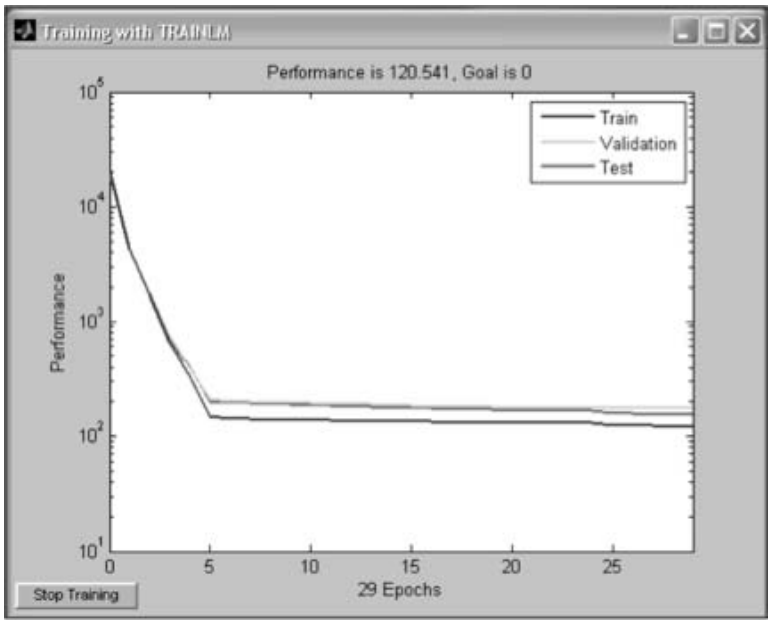

Fig. 24 Experiment #5 – Training results

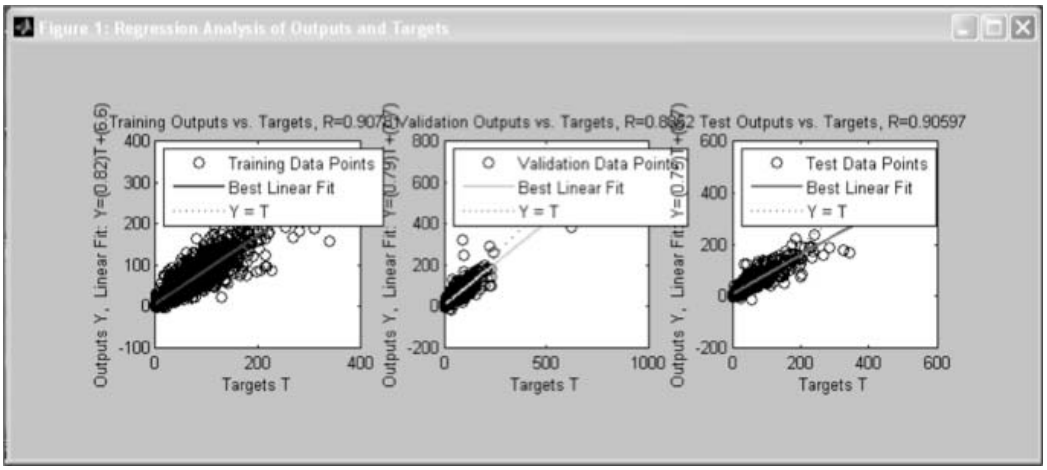

Fig. 25 Experiment #5 – Regression analysis and output-target correlation

F. Experiment #6 – NARX 3

The adopted mathematical model is:

$$y(t) = f\left(t, u(t), y(t-1), y(t-2), y(t-3), \text{mov.avg}\left[y(t-3), y(t-2), y(t-1)\right]\right) \tag{21}$$

The experimental results are detailed in Table VII. The training results are shown in Fig. 26. The regression analysis and the correlation between achieved outputs and desired targets are illustrated in Fig. 27.

TABLE VII EXPERIMENT #6 – NARX 3

Experiment	NO_x – NARX 3
Input	Norm.Time, Weathering, $NO_x(t\text{-}1)$, $NO_x(t\text{-}2)$, $NO_x(t\text{-}3)$, mov.avg[$NO_x(t\text{-}3)$, $NO_x(t\text{-}2)$, $NO_x(t\text{-}1)$]
Target	$NO_x(t)$
Hidden neurons	20

	%	No. of samples	MSE	ρ
Training	60%	10212	128.31340	0.911783
Validation	20%	3404	244.33651	0.851712
Testing	20%	3404	146.58925	0.895261

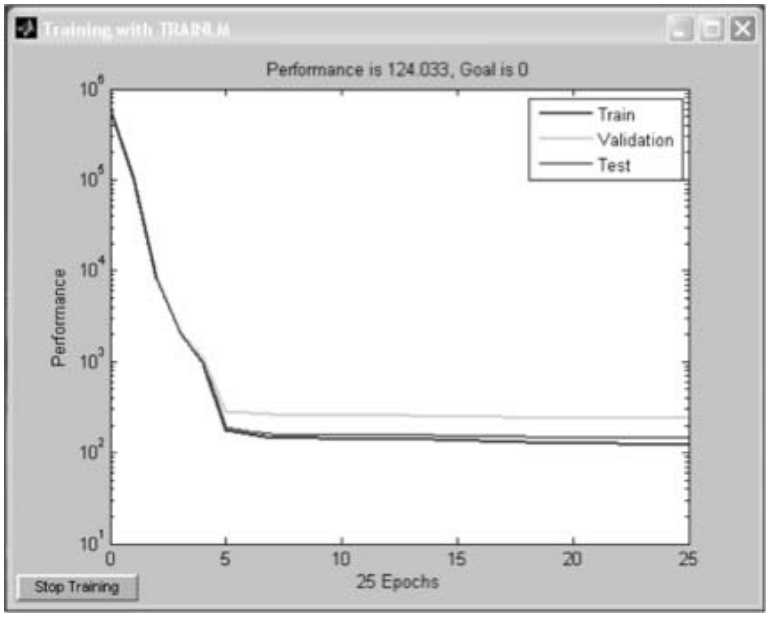

Fig. 26 Experiment #6 – Training results

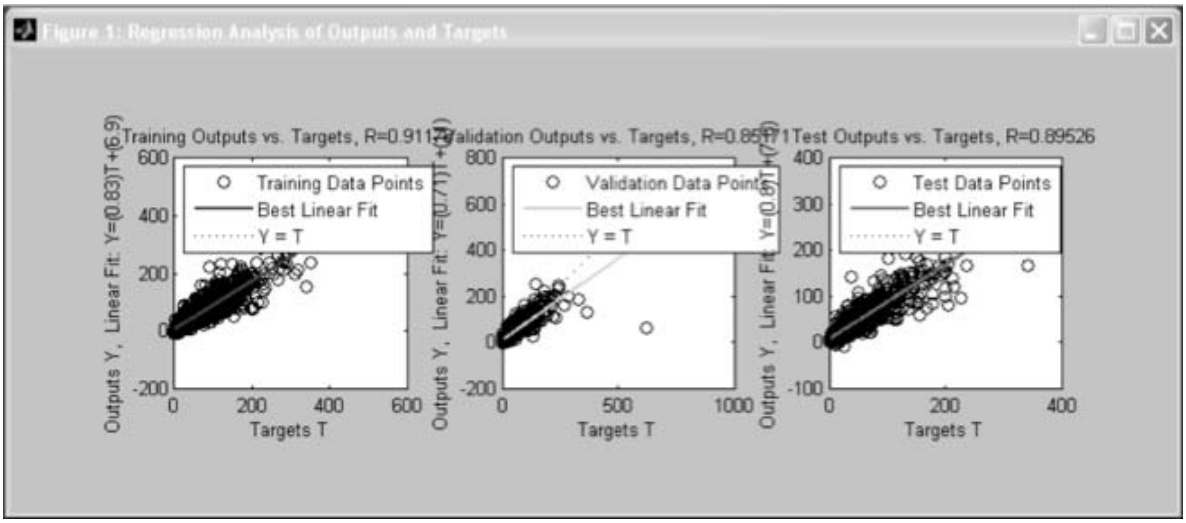

Fig. 27 Experiment #6 – Regression analysis and output-target correlation

G. NARX Experiments

The comparison between results obtained in the NARX experiments carried on SO_2 and NO_x with normalized time is synthesized in Table VIII. The comparison between results obtained in the NARX experiments carried on SO_2 (NO_x) with and without time is shown in Table IX (Table X).

TABLE VIII COMPARISON OF NARX EXPERIMENTS WITH NORMALIZED TIME

#	Name	Pollutant	ρ-Training set	ρ-Validation set	ρ-Test set
1	NARX 1	SO_2	0.936143	0.856899	0.914534
2	NARX 2	SO_2	0.837379	0.804149	0.839577
3	NARX 3	SO_2	0.844331	0.810828	0.813115
4	NARX 1	NO_x	0.899700	0.900660	0.910463
5	NARX 2	NO_x	0.907814	0.885201	0.905966
6	NARX 3	NO_x	0.911783	0.851712	0.895261

TABLE IX COMPARISON OF NARX EXPERIMENTS ON SO_2 WITH AND WITHOUT TIME

Time	#	Name	Pollutant	ρ-Training set	ρ-Validation set	ρ-Test set
Yes	1	NARX 1	SO_2	0.936143	0.856899	0.914534
Yes	2	NARX 2	SO_2	0.837379	0.804149	0.839577
Yes	3	NARX 3	SO_2	0.844331	0.810828	0.813115
No	1	NARX 1	SO_2	0.924203	0.906917	0.914810
No	2	NARX 2	SO_2	0.817806	0.823199	0.830544
No	3	NARX 3	SO_2	0.830124	0.824901	0.812203

TABLE X COMPARISON OF NARX EXPERIMENTS ON NO_x WITH AND WITHOUT TIME

Time	#	Name	Pollutant	ρ-Training set	ρ-Validation set	ρ-Test set
Yes	1	NARX 1	NO_x	0.899700	0.900660	0.910463
Yes	2	NARX 2	NO_x	0.907814	0.885201	0.905966
Yes	3	NARX 3	NO_x	0.911783	0.851712	0.895261
No	1	NARX 1	NO_x	0.927858	0.918192	0.910560
No	2	NARX 2	NO_x	0.923797	0.903678	0.903597
No	3	NARX 3	NO_x	0.915838	0.900658	0.896476

H. Evaluations on the Absence of the Time Component in NARX Experiments

The experiments outlined above show how the absence of the time variable in the input data has no significant impact in terms of the correlation coefficient's value, which remains stable and at levels similar to those observed in cases where such component is present. This leads us to consider that the forecasting model is stable.

I. Six-Hour Forecasts on the Best NARX Experiments without the Time Variable

The six-hour forecasts made on the best NARX experiments without the time variable are detailed in Table XI.

TABLE XI SIX-HOUR FORECASTS OF THE BEST NARX EXPERIMENTS ON SO_2 AND NO_X WITHOUT TIME

Experiment	Pollutant	Time	ρ-Training set	ρ-Validation set	ρ-Test set	Best
#1 NARX 1	SO_2	t+1	0.778581	0.759041	0.754776	X
#1 NARX 1	SO_2	t+2	0.708686	0.680791	0.619671	X
#1 NARX 1	SO_2	t+3	0.630116	0.612488	0.648440	X
#3 NARX 3	SO_2	t+4	0.587474	0.580585	0.595234	
#3 NARX 3	SO_2	t+5	0.569433	0.529438	0.557897	
#3 NARX 3	SO_2	t+6	0.580470	0.512588	0.511362	
#4 NARX 1	NO_x	t+1	0.790732	0.763441	0.747740	X
#4 NARX 1	NO_x	t+2	0.671889	0.633358	0.638014	X
#4 NARX 1	NO_x	t+3	0.554362	0.508112	0.545336	
#4 NARX 1	NO_x	t+4	0.491690	0.451437	0.465050	
#4 NARX 1	NO_x	t+5	0.400250	0.403929	0.374476	
#4 NARX 1	NO_x	t+6	0.424576	0.396898	0.409401	

VII. CONCLUSIONS

The analysis of the problem of air pollution shows that most of the pollutants in the atmosphere are due to emissions caused by anthropogenic factors. In this work we developed a model using artificial neural networks to forecast short-term rate of air pollution for supporting environmental policy decisions.

The chosen forecasting model is of NARX type, and has demonstrated its validity for short-term previsions. In fact, network training undergoes a decay process shown by the values of the correlation coefficient, which decreases for forecasts of more than one hour.

The NARX model provides good results (relative to the correlation coefficient ρ) both for the pollutants SO_2 and NO_x. Its application shows that it is possible to make predictions up to three hours for the pollutant SO_2 and up to two hours for the NO_x. It is not possible to predict at a longer time period given the low correlation existing between network output and desired target.

In addition, the adopted model is stable or time-independent, in the sense that the analyzed phenomenon does not depend on the time instant in which it occurs, but on a combination of meteorological factors present in that place and both the chemical and physical processes acting at that precise moment and earlier.

REFERENCES

[1] Haroldo F. Campos Velho, Reinaldo R. Rosa, Fernando M. Ramos, Roger A. Pielke, Gervásio A. Degrazia, Camilo Rodrigues Neto, and Ademilson Zanandrea. Multifractal model for eddy diffusivity and counter-gradient term in atmospheric turbulence. *Physica A*, 295:219–223, 2001.

[2] Kolumban Hutter and Klaus Jöhnk. *Continuum Methods of Physical Modeling - Continuum Mechanics, Dimensional Analysis, Turbulence*. Springer-Verlag, Berlin Heidelberg, 2004.

[3] John H. Seinfeld and Spyros N. Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. Wiley- Interscience, U.S.A., second edition, 1998.

[4] T.S. Ursell. The diffusion equation: A multi-dimensional tutorial. Technical report, Department of Applied Physics, California Institute of Technology, Pasadena, CA 91125, 2007.

[5] H. Eugene Stanley. Turbulence: The legacy of A.N. Kolmogorov. *Journal of Statistical Physics*, 88.1:521–523, 1997.

[6] A.K.M.F. Hussain and W.C. Reynolds. Measurements in fully developed turbulent channel flow. *ASME J. Fluids Engng*, 97:568–580, 1975.

[7] C. Liu, R. Bammer, B. Acar, and M.E. Moseley. Characterizing non-gaussian diffusion by using generalized diffusion tensors. *Magnetic Resonance in Medicine*, 51:924–937, 2004.

[8] Simon Haykin. *Neural Networks: a comprehensive foundation*. Prentice Hall, New Jersey, (3rd ed) edition, 2007.

[9] K. Hornik, M. Stinchcombe, and H. White. Multi layer feed forward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.

[10] G.P. Zhang and M.Y.H. Patuwo. Forecasting with artificial neural networks: The state of the art. *Int. Archives of Photogrammetry and Remote Sensing*, 1(14):35–62, 1998.

Crescenzio Gallo (IEEE Member, ACM Professional Member) received his B.Sc. in Computer Science from the University of Bari, Italy (1978) and the M.Sc. in Information Science and Technologies from the University of Milan, Italy (2007). He has been with the Institute of Computer Science at the University of Bari as a Research Assistant in Information Systems, and with Telespazio (Rome, Italy) participating to the Landsat satellite projects of NASA and ESA. Since Jan 2004 he is a Research Professor in Information Processing Systems at the University of Foggia, Italy. His primary research interests include bioinformatics, artificial neural networks and its environmental applications.