Available online at www.sciencedirect.com
SCIENCE DIRECT®
Environmental Modelling & Software 21 (2006) 430–446
Environmental Modelling & Software
www.elsevier.com/locate/envsoft

Regression and multilayer perceptron-based models to forecast hourly O_3 and NO_2 levels in the Bilbao area

E. Agirre-Basurko [a,*], G. Ibarra-Berastegi [b], I. Madariaga [c]

[a] Applied Mathematics Department, Industrial Technical Engineering University School, University of the Basque Country, Plaza La Casilla, 3, 48012 Bilbao, Spain
[b] Fluid Mechanics Department, ESI Bilbao, University of the Basque Country, C/Alda, Urkijo, s/n, 48013 Bilbao, Spain
[c] Electronics Department, ESI Bilbao, University of the Basque Country, C/Alda, Urkijo, s/n, 48013 Bilbao, Spain

Received 13 August 2003; received in revised form 26 July 2004; accepted 28 July 2004
Available online 9 February 2005

Abstract

In this paper, we present the results obtained using three prognostic models to forecast ozone (O_3) and nitrogen dioxide (NO_2) levels in real-time up to 8 h ahead at four stations in Bilbao (Spain). Two multilayer perceptron (MLP) based models and one multiple linear regression based model were developed. The models utilised traffic variables, meteorological variables and O_3 and NO_2 hourly levels as input data, which were measured from 1993 to 1994. The performances of these three models were compared with persistence of levels and the observed values. The statistics of the Model Validation Kit determined the goodness of the fit of the developed models. The results indicated improved performance for the multilayer perceptron-based models over the multiple linear regression model. Furthermore, comparisons of the results of the multilayer perceptron-based models proved that the insertion of four additional seasonal input variables in the MLP provided the ability of obtaining more accurate predictions. The comparison of the results indicated that this model performance was more efficient in the forecasts of O_3 and NO_2 hourly levels k hours ahead ($k = 1, 4, 5, 6, 7, 8$), but not in the forecasted values 2 and 3 h ahead. Future research in this area could allow us to improve results for the above forecasts. The multilayer perceptron modelling was developed using the MATLAB software package.

Keywords: Air quality modelling; Neural networks; Multilayer perceptron; Multiple linear regression

1. Introduction

As in many other urban areas, nowadays ozone (O_3) and nitrogen dioxide (NO_2) are the most relevant air pollutants in Bilbao (North Central Spain) (European Communities, 1998; Basque Government, 2001). It is well-known that ozone is a secondary pollutant, produced owing to the interaction of meteorology, NO_x and volatile organic compounds (VOCs) (Finlayson-Pitts and Pitts, 1986; Saunders et al., 1997). In the same way, NO_2 is formed as a result of the reaction of NO with O_3.

Deterministic and statistical models have been developed for the purpose of forecasting photochemical smog levels. Deterministic models are based on mathematical relationships that describe the processes involved in the formation of air pollutants. They are also known as cause/effect models (Zanetti, 1990). For example, the Urban Airshed Model (UAM) has been widely recognized (Scheffe and Morris, 1993). The UAM is an eulerian tridimensional model, which has been applied in the USA, Greece, Germany, Austria, Holland, Italy and Japan. However, the UAM is not appropriate as a prognostic model in coastal zones,

* Corresponding author. Tel.: +34 94 601 4405; fax: +34 94 601 4300.
E-mail address: mapagbae@lg.ehu.es (E. Agirre-Basurko).

1364-8152/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2004.07.008

where the influence of the coastal breeze plays a significant role (Rye, 1995). It also has limitations in the forecasting of ozone levels (Chock and Chang, 1999). In general, cause/effect models are more suitable over extensive areas such as whole regions and large cities. Moreover, they require precise data from the emission and transportation of pollutants (deriving from traffic) and meteorological conditions. The lack of sufficient data is one of the causes of the uncertainty of the deterministic models in many instances. Statistical models are able to establish a relationship between the input variables (predictors) and the output variables (predictands) without detailing the causes and effects in the formation of pollutants. Models based in time series analysis (Simpson and Layton, 1983) (Kuang-Jung, 1992), multiple linear regression based models (Cassmassi, 1998; Cardelino et al., 2001), and more recently, neural network-based models (Gardner and Dorling, 1999; Elkamel et al., 2001) are excellent examples.

Since Boznar et al. (1991) presented the earliest paper based on the use of neural networks for the prediction of sulphur dioxide (SO_2), several authors have developed different neural network-based models to forecast air pollutant concentrations. A wide review of applications of the multilayer perceptron in atmospheric sciences, including forecasting, was presented by Gardner and Dorling (1998). Daily forecasts of maximum ozone concentrations were made using neural networks, and then, the results were compared with the forecasts obtained using multiple regression models (Yi and Prybutok, 1996) (Jorquera et al., 1998). Oxides of nitrogen (NO_x) and NO_2 concentrations were predicted by applying a multilayer perceptron-based model and other statistical models, and the comparison of the different results demonstrated the benefit of using a multilayer perceptron (Gardner and Dorling, 1999). The use of neural networks allowed the prediction of $PM_{2.5}$ concentrations in Santiago, Chile (Perez et al., 2000) and the prediction of PM_{10} concentrations in Helsinki (Kukkonen et al., 2003). Neural networks have proven their efficiency to describe nonlinear relationships, such as those involved in ozone formation (Gardner and Dorling, 2000). Recently, neural networks have been used in the prediction of SO_2 levels and have demonstrated greater efficiency than linear methods (Chelani et al., 2002). In general, the use of neural networks, and in particular multilayer perceptrons, has generally provided better results compared to statistical linear methods.

In the present study, for the first time, a neural network-based model was built to forecast O_3 and NO_2 concentrations in the Bilbao area. Furthermore, traffic variables were used as predictor variables in the developed models. The primary goal of the work was to build an accurate statistical model to forecast O_3 and NO_2 levels k hours ahead in the Bilbao area ($k = 1,\ldots,8$). Two techniques were applied to build the models: the multilayer perceptron and multiple linear regression. Based on these techniques, three different models were designed, and comparisons between them established the most efficient performer as a forecasting tool.

2. Techniques applied in modelling

Before describing the methodology followed to elaborate and evaluate the prognostic models, a brief review of the techniques of multiple linear regression and multilayer perceptrons is shown below.

2.1. Multiple linear regression

This is the general form of multiple linear regression:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i \tag{1}$$

where, for a set of i observations, Y_i is the predictand variable, β_0 is a coefficient, $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients of the $X_{i1}, \ldots, X_{ip}$ independent variables (predictors) and ε_i is the residual error (difference between observations and predicted values).

The hypotheses required to apply multiple linear regression are: (i) the predictor variables must be independent, and (ii) the residual errors ε_i must be independent and they must be normally distributed, with 0 mean and σ^2 constant variance.

The observations $\{X_{i1}, X_{i2}, \ldots, X_{ip}, Y_i\}_{i=1,2,\ldots,n}$ are helpful in the estimation of the parameters β and they form the calibration set. The least square method is the usual technique used to estimate the parameters. Hence, the equation for the predicted value is:

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \cdots + b_p X_{ip} \tag{2}$$

where, b_i are the estimations of the β_i parameters and $\hat{Y}_i$ is the predicted value.

The goal of the regression analysis is to determine the values of the parameters of the regression equation and then to quantify the goodness of the fit in respect of the dependent variable Y.

2.2. Multilayer perceptrons

The concept of artificial neural networks was established in 1943 (McCulloch and Pitts, 1943). Later, in 1958, the first practical artificial neural network was presented: the perceptron (Rosenblatt, 1958). Since 1986, neural networks have become more widely recognized (Rumelhart et al., 1986).

Neural networks are mathematical structures built in a similar way to the nervous system, where the neuron is the fundamental element. Neural networks are formed by artificial neurons (nodes) set in layers and connected with each other. Neural networks are capable of learning from the patterns presented to them and from the errors they commit in the learning process, so that finally they should identify patterns never seen before (generalization). Several types of neural networks can be determined depending on the type of connections, the neuron model and the methods for adjusting the weights (Hagan et al., 1996). The Multilayer Perceptron (MLP) is an artificial neural network which is well-known for its ability to represent any smooth measurable functional relationship between the inputs (predictors) and the outputs (predictands). It is the most commonly used type of feed-forward neural network in the atmospheric sciences (Gardner and Dorling, 1998).

The MLP is composed of at least three layers of neurons: the input layer, the hidden layer(s) and the output layer. Fig. 1 shows an MLP with N neurons in the input layer, S neurons in the hidden layer and L neurons in the output layer. It could be represented as an N–S–L network.

The output of the network can be expressed by Eq. (3) as follows:

$$y_k^o = f_k^o\left(b_k^o + \sum_{i=1}^{S} w_{ik}^o y_i^h\right) = f_k^o\left(b_k^o + \sum_{i=1}^{S} w_{ik}^o f_i^h\left(b_i^h + \sum_{j=1}^{N} w_{ji}^h x_j\right)\right), \quad k = 1, \ldots, L \tag{3}$$

where h denotes the elements of the hidden layer and o indicates the elements of the output layer, w_{ji}^h is the weight that connects the neuron j of the input layer with the neuron i of the hidden layer, w_{ik}^o is the weight that connects the neuron i of the hidden layer with the neuron k of the output layer, f_i^h is the transfer function of neuron i of the hidden layer, f_k^o is the transfer function of neuron k of the output layer, b_i^h is the bias of neuron i of the hidden layer, b_k^o is the bias of neuron k of the output layer, n_i^h is the excitation level (Eq. (4)) of neuron i of the hidden layer, n_k^o is the excitation level (Eq. (5)) of neuron k of the output layer, y_i^h is the output of neuron i of the hidden layer, y_k^o is the output of neuron k of the output layer and x_j is the input of neuron j of the input layer.

$$n_i^h = b_i^h + \sum_{j=1}^{N} w_{ji}^h x_j \tag{4}$$

$$n_k^o = b_k^o + \sum_{i=1}^{S} w_{ik}^o y_i^h \tag{5}$$

The most widely used nonlinear transfer functions are the logarithmic sigmoid function (logsig) and the hyperbolic tangent function (tansig):

$$\mathrm{logsig}(x) = \frac{1}{1 + \mathrm{e}^{-x}} \tag{6}$$

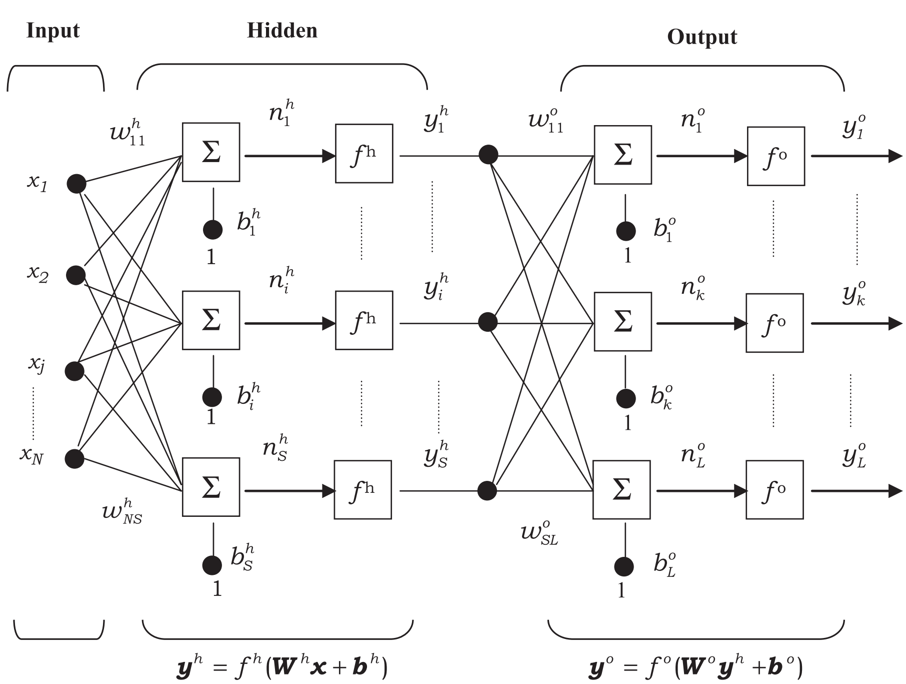

Fig 1. A multilayer perceptron-based model.

$$\mathrm{tansig}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{7}$$

Both are differentiable and bounded functions (the first between 0 and 1, the second between −1 and 1). Hence, after the normalization of the patterns $(x_1, x_2,\ldots,x_N, t_1, t_2,\ldots,t_L)^t$, the learning process does start. It could be considered as an optimization process, where the error function E has to be minimized:

$$E = \frac{1}{L}\sum_{k=1}^{L}\left(t_k - y_k^o\right)^2 \tag{8}$$

being $(t_1, t_2,\ldots,t_L)^t$ the target vector of the MLP.

When the MLP memorizes the patterns introduced to it and it is not capable of identifying new situations, overtraining does occur. The early stopping technique can be used to avoid overtraining (Sarle, 1995). In the early stopping technique the available data are separated into three sets: the training set, the validation set and the test set. The training set is used to update the network weights and biases. During the training, the validation set is used to guarantee the generalization capability of the model. Training should stop before the error on the validation set begins to rise. Finally, the test set is a new set used to check the generalization of the MLP.

Different training algorithms could be applied to minimize the error function, but the most widely used are the backpropagation algorithm and the algorithms derived from it.

3. Database

An air pollution network managed by the Basque Government since 1977 measures hourly meteorological parameters and air pollution variables at each station in Bilbao. In the same way, the traffic network managed by the Local Municipality of Bilbao measures two different and independent traffic variables at each station: the variable NV indicates the number of vehicles circulating every 10 min and the variable OP indicates the fraction of time for which the area of road is occupied by a vehicle. Both network measures are highly consistent.

The data used in this work were hourly current (at time t) data and historical (at time $t - z$, $z = 1,\ldots,15$) data from the air pollution network and the traffic network of Bilbao during years 1993–1994. The data selected jointly reduced the study to four stations in Bilbao, namely Elorrieta, Txurdinaga, Mazarredo and Deusto. These four stations, located in the central area of Bilbao, are close to each other – the greatest distance between any of them is <5 km. The selection of the variables of this study (Table 1) is based on earlier works (Ibarra et al., 2001a).

Table 1
Meteorological variables, air pollution variables and traffic variables used to develop the models

Classification of variables	Variables	Notation
MET	Wind speed (m/s)	Vx
	Wind direction (N°)	Vy
	Temperature (°C)	TEM
	Relative humidity (%)	HUM
	Pressure (kPa)	PRE
	Radiation (cal cm^{-2} h^{-1})	RAD
	Thermal gradient (°C)	GRAD
POL	Ozone (μg/m^3)	O_3
	Nitrogen dioxide (μg/m^3)	NO_2
TRAF	Number of vehicles (vehic./10 min)	NV
	Occupation percentage (%)	OP
	Velocity (km h^{-1} 100^{-1})	KH

The meteorological variables considered were wind speed and direction, thermal contrast between Feria and Banderas (two stations located at sea level and 200 m above sea level, respectively), relative humidity, pressure, temperature and radiation. In the same way, O_3 and NO_2 concentrations measured at the four stations were used. All these variables were measured hourly.

Finally, as several works have proven that traffic plays a significant role in the formation of ozone (Mayer, 1999; Borrego et al., 2000; Ibarra et al., 2001b), the database was completed with the mean hourly values of three traffic variables registered in Bilbao in the period 1993–1994: (i) the number of vehicles NV, (ii) the occupation percentage OP, and (iii) the variable KH = (NV/OP), which gives an idea of the velocity.

The matrix of data collected had 17.520 entries (hourly measures throughout 1993–1994 period) for each station. Some of the entries were missing values, and in such cases, the entire row was eliminated. The proportion of missing values ranged between 1 and 6% for all the specific data measured in the period 1993–1994 at Elorrieta, Mazarredo, Txurdinaga and Deusto. Tables 2 and 3 show the number of valid values, the number of missing values, average, standard deviation, maximum and the percentiles 50, 75 and 95 of O_3 and NO_2 at time t in each station in 1993 and 1994, respectively. The obtained values are very similar at the different stations. The worst case was registered in Deusto, where the number of valid data involved in the prediction of ozone was reduced to 70%.

4. Methodology

Two multilayer perceptron-based models (MLP1 and MLP2) and one multiple linear regression model (LR) were developed using the current and past values of the indicated variables measured in the Bilbao air pollution

Table 2
Statistics for NO_2 and O_3 at time t in Deusto, Elorrieta, Mazarredo and Txurdinaga in 1993

Station	Pollutant	Missing values	Mean	Std. Dev.	Max.	Percentile		
						50	75	95
Deusto	NO_2	829	53.358	27.71	210.20	50.20	69.80	103.80
	O_3	2633	22.72	18.77	136.30	19.70	34.50	56.40
Elorrieta	NO_2	127	54.06	36.93	404.00	46.50	68.50	120.00
	O_3	78	21.45	20.06	146.00	15.50	33.00	61.00
Mazarredo	NO_2	95	51.73	25.57	183.00	49.50	67.50	96.00
	O_3	164	28.69	28.60	155.50	18.00	46.50	85.00
Txurdinaga	NO_2	62	52.10	25.92	213.00	48.50	64.50	100.50
	O_3	359	26.45	27.75	307.00	14.50	41.50	82.00

and traffic networks during 1993 and 1994. After introducing the appropriate inputs, the outputs of the models were the forecasted O_3 or NO_2 levels at time $t+k$, $k = 1,\ldots,8$. Data from 1993 were used to build the models and data from 1994 were used to test the models.

4.1. Building the models

The technique applied to build the LR model is the multiple linear regression. The (9)–(10) equation system represents the LR model:

$$O_3(t+k) = f(NO_2(t+k), \mathrm{MET}(t-z), \mathrm{TRAF}(t-z), \mathrm{POL}(t-z)) \tag{9}$$

$$NO_2(t+k) = g(O_3(t+k), \mathrm{MET}(t-z), \mathrm{TRAF}(t-z), \mathrm{POL}(t-z)) \tag{10}$$

The $\mathrm{MET}(t-z)$ variables are current and historical ($z = 0, 1,\ldots,15$) values of temperature, pressure, wind, thermal gradient, relative humidity and global radiation. The $\mathrm{TRAF}(t-z)$ variables are current and historical ($z = 0, 1,\ldots,15$) values of the variables NV, OP and KH. The $\mathrm{POL}(t-z)$ are the current and historical ($z = 0, 1,\ldots,15$) values of O_3 and NO_2 in Elorrieta, Mazarredo, Txurdinaga and Deusto. These are the independent variables of the model. $O_3(t+k)$ and $NO_2(t+k)$, the forecasts of O_3 and NO_2 respectively, ($k = 1,\ldots,8$), are the dependent variables. Finally, the functions f and g are linear functions.

The fundamental characteristic of this model is that the variable $NO_2(t+k)$ is an independent variable in each equation designed to forecast $O_3(t+k)$, and at the same time, $O_3(t+k)$ is another independent variable in each equation designed to forecast $NO_2(t+k)$, $k = 1,\ldots,8$.

Data from year 1993 were used to calculate the coefficients of the regression equation system (9)–(10). Stepwise regression and tolerance filtering were applied to select the independent variables and to calculate the coefficients of the (9)–(10) equation system. Hence, with only five significant independent variables, the coefficients of Eqs. (9)–(10) were determined using the hourly data from the mentioned networks corresponding to year 1993. It is remarkable the role that the independent variable $NO_2(t+k)$ plays in the forecast of $O_3(t+k)$, and reciprocally, the role that the independent variable $O_3(t+k)$ plays in the forecast of $NO_2(t+k)$, $k = 1, 2,\ldots,8$, these being the most relevant variables in the majority of predictions. Once the coefficients of

Table 3
Statistics for NO_2 and O_3 at time t in Deusto, Elorrieta, Mazarredo and Txurdinaga in 1994

Station	Pollutant	Missing values	Mean	Std. Dev.	Max.	Percentile		
						50	75	95
Deusto	NO_2	648	47.16	21.89	168.00	45.40	59.50	85.34
	O_3	2456	32.48	23.60	135.80	29.85	49.80	72.30
Elorrieta	NO_2	178	40.57	19.89	140.00	38.00	53.00	76.50
	O_3	218	25.31	22.26	137.00	18.00	37.50	69.50
Mazarredo	NO_2	211	42.97	19.60	141.50	41.00	55.50	78.00
	O_3	224	30.64	30.46	189.50	19.00	49.87	90.57
Txurdinaga	NO_2	81	44.57	20.76	160.00	42.00	56.50	82.00
	O_3	109	27.88	25.34	168.50	17.00	43.00	78.50

Eqs. (9)–(10) were determined, the next step was to calculate the forecasts of $O_3(t+k)$ and $NO_2(t+k)$ at each station ($k = 1, 2,\ldots,8$), using the data from year 1994 or test data.

The multilayer perceptron-based models built in this work have the structure N–S–1, where it is accepted that the MLP-based model have one input layer with N neurons, one unique hidden layer with S neurons and one output layer with 1 neuron (output). The inputs required to forecast O_3 and NO_2 up to 8 h ahead must be determined. In the same way, the number of neurons in the hidden layer is unknown. However, the number of neurons in the output layer is known: one unique neuron, which corresponds to the forecasted value of O_3 or NO_2 at time $t+k$ ($k = 1,\ldots,8$).

The MLP neural networks used in this work were trained using the scaled conjugate gradient (SCG) algorithm. It has been proven that the SCG algorithm performs better than the standard backpropagation algorithm and it converges faster than other conjugate gradient algorithms (Moller, 1993). At the same time, in order to avoid overtraining, the early stopping technique was applied: the first 85% of the data from 1993 were the training set, the last 15% of the data from 1993 formed the validation set and the data from 1994 were chosen to be the test set. Finally, the transfer functions selected for the layers were the hyperbolic tangent (tansig) for the hidden layer and the linear function for the output layer.

Based on these criteria the MLP1 model and the MLP2 models were built. Although both models have a similar structure, the number of input variables selected for each is different. The inputs of the MLP1 model were determined by applying stepwise regression and tolerance filtering using data from 1993 in two equations that were similar to Eqs. (9) and (10), but without the variables $NO_2(t+k)$ and $O_3(t+k)$ as independent variables in the corresponding prognostic equations. The result was that five significant inputs were responsible for the total variance. Hence, the structure of the MLP1 model was 5–S1–1, where S1 is the number of neurons in the hidden layer. In addition to the inputs of the MLP1 model, in the MLP2 model four new input variables were inserted: $\sin(2\pi h/24)$, $\cos(2\pi h/24)$, $\sin(2\pi d/7)$ and $\cos(2\pi d/7)$, where h is the hour of the day ($h = 1, 2,\ldots,24$) and d is the day of the week ($d = 1, 2,\ldots,7$). Therefore, the structure of the MLP2 model was 9–S2–1, where S2 is the number of neurons in the hidden layer. S1 and S2 were calculated by applying the generalization rule proposed by Amari et al. (1997) in a trial and error procedure: "the model is optimally trained when the ratio of the number of training samples to the number of the connection weights exceeds 30".

Therefore, with the MLP1 model 16 multilayer perceptrons were built to forecast the values $O_3(t+k)$ and $NO_2(t+k)$ for each station ($k = 1,\ldots,8$), which made a total of 64 multilayer perceptrons. In the same way, using the MLP2 model, another 64 multilayer perceptrons were built to forecast $O_3(t+k)$ and $NO_2(t+k)$ for the four stations ($k = 1,\ldots,8$).

4.2. *Testing the models*

To reduce the uncertainty of applying the appropriate statistics to choose the best model, in 1991 it was decided to initiate a series of workshops. These workshops were supported by COST 710 and COST 615 and the European Association for the Science of Air Pollution (EURASAP). In 1993 the workshop took place in Manno (Switzerland), and it was dedicated to the establishment of objective criteria for comparing different models. Consequently, a data processing package known as the *Model Validation Kit* (European Commission, 1994) was created, which was improved in the following workshop in Mol (Belgium) in 1994. The kit was formed by criteria based on a previous work (Hanna et al., 1991). Although these measures were thought to compare the performance of cause/effect models, their application to statistical models is immediate. These statistics allow the comparison of the performance of different models, where C_p are the forecasted values and C_o are the observed values, σ indicates the standard deviation and *Mean* is the mean value.

The proposed measures in the Model Validation Kit are:

(i) The correlation coefficient between C_o and C_p, R, quantifies the global description of the model:

$$R = \mathrm{Mean}\left[(C_o - \mathrm{Mean}(C_o))\left(C_p - \mathrm{Mean}(C_p)\right)\right] / \left((\sigma_{C_o})(\sigma_{C_p})\right) \tag{11}$$

(ii) The Normalized Mean Square Error, NMSE, is a version of the mean square error, but normalized with the object of establishing comparisons among different models:

$$\mathrm{NMSE} = \frac{\mathrm{Mean}\left(C_o - C_p\right)^2}{\mathrm{Mean}(C_o)\mathrm{Mean}\left(C_p\right)} \tag{12}$$

(iii) The factor of two, FA2, which gives the percentage of forecasted cases in which the values of the ratio C_o/C_p are in the range [0.5, 2]:

$$0.5 \le C_o/C_p \le 2 \tag{13}$$

(iv) The Fractional Bias, FB, is a normalized measure that allows the comparison of the mean of the observed values and the mean of the predicted values. A model with FB = 0 is a model that represents perfectly the measured mean value:

$$FB = 2\frac{\mathrm{Mean}(C_o) - \mathrm{Mean}(C_p)}{\mathrm{Mean}(C_o) + \mathrm{Mean}(C_p)} \quad (14)$$

(v) The Fractional Variance, FV, is another normalized measure that allows the comparison of the difference between the predicted variance and the observed variance. A model with FV = 0 is a model whose variance is equal to the variance of the observed values:

$$FV = 2\frac{\sigma_{C_o} - \sigma_{C_p}}{\sigma_{C_o} + \sigma_{C_p}} \quad (15)$$

In this study, the calculation of the statistics of the Model Validation Kit on the test set determined the goodness of the fit of the LR, MLP1 and MLP2 models in a quantitative manner. These results were compared with the values of the statistics corresponding to observations and persistence of levels.

5. Results

Table 4(1a–4b) shows the values of the statistics included in the Model Validation Kit for the observation, LR, MLP1, MLP2 and persistence of levels on the test set (1994) in Elorrieta, Txurdinaga, Mazarredo and Deusto. The best forecast has NMSE, FV and FB values equal to zero and the corresponding values of R and FA2 equal to the unit.

Firstly, in the forecasts of $NO_2(t+1)$ and $O_3(t+1)$, a similar situation was observed at all the stations. The lowest values of NMSE were obtained with the MLP1 and MLP2 models, being notably lower than the corresponding values obtained for NMSE with the LR model or persistence. The correlation coefficients obtained using the MLP2 model demonstrated the efficacy of the performance of the MLP2 model, 0.905 being the correlation coefficient corresponding to Elorrieta, 0.889 the correlation coefficient corresponding to Mazarredo, 0.889 the correlation coefficient corresponding to Txurdinaga and 0.879 the correlation coefficient corresponding to Deusto in the forecast of $NO_2(t+1)$ and they took the values 0.916 for Elorrieta, 0.929 for Mazarredo, 0.931 for Txurdinaga and 0.879 for Deusto in the forecast of $O_3(t+1)$. In the same way, the greatest values of FA2 were obtained using the MLP2 model, so that the percentage of the ratio C_p/C_o that falls in the range [0.5, 2] was 97.2–98% for $NO_2(t+1)$ and 83.2–93.5% for $O_3(t+1)$. Finally, the values FB and FV of the model MLP1 and MLP2 were very low, although they were greater than the values corresponding to the LR model and persistence. Consequently, the more accurate forecasts of $NO_2(t+1)$ and $O_3(t+1)$ were provided by the MLP2 model.

Secondly, in the forecasts of $NO_2(t+2)$ and $O_3(t+2)$, the lowest values of NMSE were obtained with the MLP1 and MLP2 models, except for Deusto, where the values of NMSE obtained using the MLP1 and the MLP2 were very high, particularly in the case of the MLP2 model. The greatest values of the correlation coefficients (0.637–0.810) were obtained with the LR model, and in all cases the correlation coefficient for persistence was better than those for the MLP1 and the MLP2 models. Furthermore, the greatest values of FA2 were those corresponding to the LR model and persistence, and greatest values of FB and FV were for the MLP1 and the MLP2 models, being very high for Deusto. As a conclusion, the LR model was the most appropriate model for forecasting O_3 and NO_2 levels 2 h ahead at the four stations. The MLP1 and MLP2 models were not performing better in this case, and they were clearly less effective than the LR model in the forecasts of $NO_2(t+2)$ and $O_3(t+2)$ for Deusto.

As in previous forecasts, for $NO_2(t+3)$ and $O_3(t+3)$ the lowest values of NMSE were also obtained with the MLP1 and the MLP2 models. However, the LR model provided the greatest correlation coefficients (0.606–0.675) and the greatest values of FA2 in every case. Finally, values of FB and FV obtained using the MLP1 and the MLP2 models showed a greater dispersion than the LR model between the forecasted values and the observations. Consequently, in the forecasts of $NO_2(t+3)$ and $O_3(t+3)$ the LR model performed best at all stations.

The above tendency changed in the forecasts of $NO_2(t+4)$ and $O_3(t+4)$, when the MLP2 model began to improve its performance. The lowest values of NMSE were obtained with the MLP1 and the MLP2 models, and the greatest correlation coefficients and FA2 values were those corresponding to the MLP2 in the majority of cases. The values of FB were similar in all the models, although the values of FV showed greater dispersion in the MLP2 model. The MLP1 model was not as effective as the MLP2 model, but it demonstrated greater accuracy than persistence. The conclusion was that the MLP2 model was performing better than the LR model, except in the forecast of $O_3(t+4)$ at Txurdinaga, where the LR model performance was more efficient.

The results of the Model Validation Kit showed that the MLP2 model was the most effective in the forecasts of $NO_2(t+5)$ and $O_3(t+5)$. It had the lowest values of NMSE (around 0) and FB, and the greatest values of R (0.496–0.670) and FA2 (51.9%–87.9%), although the values of FV were not the lowest (particularly for Txurdinaga and Deusto). The forecasts were slightly less accurate in the case of Deusto; it could be due to the fact

Table 4
Values of the Model Validation Kit statistics

Predictand	Model	NMSE	*R*	FA2	FB	FV
(1a) Elorrieta on the test set ($k = 1, 2, 3, 4$)						
Elorrieta NO_2 ($t + 1$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.06	0.880	0.948	−0.017	0.080
	MLP1	0.0008	0.890	0.954	−0.029	0.106
	MLP2	0.0003	0.905	0.972	−0.017	0.051
	Persistence	0.06	0.890	0.965	0.000	0.004
Elorrieta O_3 ($t + 1$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.11	0.878	0.880	0.018	0.152
	MLP1	0.0006	0.914	0.917	0.025	0.106
	MLP2	0.0005	0.916	0.914	0.023	0.086
	Persistence	0.12	0.871	0.900	−0.015	0.012
Elorrieta NO_2 ($t + 2$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.40	0.637	0.654	0.366	0.059
	MLP1	0.05	0.315	0.690	0.222	−0.047
	MLP2	0.038	0.388	0.757	0.194	0.241
	Persistence	0.14	0.734	0.891	−0.013	0.000
Elorrieta O_3 ($t + 2$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.24	0.735	0.696	−0.184	0.189
	MLP1	0.05	0.243	0.393	−0.230	−0.138
	MLP2	0.03	0.240	0.432	−0.178	0.244
	Persistence	0.28	0.728	0.761	−0.025	0.030
Elorrieta NO_2 ($t + 3$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.17	0.675	0.823	−0.010	−0.014
	MLP1	0.003	0.529	0.871	0.056	0.435
	MLP2	0.05	0.623	0.827	−0.230	0.404
	Persistence	0.21	0.593	0.832	0.005	−0.016
Elorrieta O_3 ($t + 3$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.34	0.648	0.670	0.005	0.264
	MLP1	0.08	0.287	0.464	−0.288	0.445
	MLP2	0.125	0.327	0.462	−0.348	0.639
	Persistence	0.43	0.603	0.670	−0.041	0.017
Elorrieta NO_2 ($t + 4$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.18	0.652	0.805	−0.088	−0.027
	MLP1	0.02	0.622	0.845	−0.153	0.164
	MLP2	0.02	0.652	0.866	−0.154	0.213
	Persistence	0.27	0.479	0.788	−0.014	0.000
Elorrieta O_3 ($t + 4$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.48	0.556	0.628	0.134	0.331
	MLP1	0.000	0.453	0.525	0.001	0.550
	MLP2	0.002	0.577	0.590	0.045	0.404
	Persistence	0.54	0.488	0.628	−0.057	0.031
(1b) Elorrieta on the test set ($k = 5, 6, 7, 8$)						
Elorrieta NO_2 ($t + 5$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.19	0.663	0.796	0.096	0.010
	MLP1	0.02	0.543	0.838	−0.151	0.111
	MLP2	0.007	0.670	0.879	−0.085	0.108
	Persistence	0.33	0.318	0.743	−0.013	0.013
Elorrieta O_3 ($t + 5$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.46	0.536	0.603	−0.056	0.359
	MLP1	0.01	0.459	0.537	0.091	0.455
	MLP2	0.01	0.543	0.571	0.108	0.559
	Persistence	0.74	0.444	0.583	0.013	0.006
Elorrieta NO_2 ($t + 6$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.33	0.576	0.687	0.246	0.019
	MLP1	0.03	0.524	0.819	−0.166	0.241
	MLP2	0.02	0.648	0.860	−0.143	0.105
	Persistence	0.37	0.265	0.722	−0.042	−0.026

(*continued on next page*)

Table 4 (*continued*)

Predictand	Model	NMSE	R	FA2	FB	FV
Elorrieta O_3 ($t+6$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.50	0.444	0.560	−0.126	0.328
	MLP1	0.002	0.529	0.587	0.039	0.525
	MLP2	0.004	0.539	0.589	0.060	0.501
	Persistence	0.80	0.310	0.527	−0.036	0.042
Elorrieta NO_2 ($t+7$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.20	0.622	0.796	−0.097	−0.041
	MLP1	0.003	0.626	0.876	−0.056	0.226
	MLP2	0.006	0.669	0.889	−0.075	0.194
	Persistence	0.39	0.241	0.695	−0.016	0.006
Elorrieta O_3 ($t+7$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.68	0.394	0.542	0.155	0.464
	MLP1	0.004	0.445	0.534	0.062	0.576
	MLP2	0.003	0.500	0.576	0.054	0.568
	Persistence	0.81	0.261	0.505	−0.096	0.047
Elorrieta NO_2 ($t+8$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.16	0.643	0.847	−0.086	0.145
	MLP1	0.02	0.510	0.836	−0.138	0.545
	MLP2	0.02	0.663	0.858	−0.134	0.143
	Persistence	0.37	0.284	0.715	0.040	0.052
Elorrieta O_3 ($t+8$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.63	0.367	0.557	0.101	0.459
	MLP1	0.01	0.427	0.550	0.107	0.767
	MLP2	0.01	0.499	0.578	0.093	0.597
	Persistence	0.88	0.264	0.489	−0.010	−0.009
(2a) Mazarredo on the test set ($k = 1, 2, 3, 4$)						
Mazarredo NO_2 ($t+1$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.07	0.838	0.938	−0.022	0.072
	MLP1	0.0005	0.854	0.961	−0.022	0.067
	MLP2	0.0005	0.889	0.972	−0.022	0.077
	Persistence	0.06	0.865	0.966	−0.001	0.001
Mazarredo O_3 ($t+1$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.12	0.899	0.784	0.014	0.077
	MLP1	0.0002	0.922	0.787	0.015	0.062
	MLP2	0.0004	0.929	0.846	0.021	0.079
	Persistence	0.13	0.895	0.828	−0.015	−0.001
Mazarredo NO_2 ($t+2$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.12	0.723	0.890	−0.013	0.148
	MLP1	0.023	0.242	0.659	0.150	−0.121
	MLP2	0.047	0.326	0.658	0.216	−0.129
	Persistence	0.14	0.712	0.888	0.012	0.007
Mazarredo O_3 ($t+2$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.23	0.800	0.706	0.016	0.116
	MLP1	0.001	0.473	0.403	0.037	0.368
	MLP2	0.001	0.406	0.398	0.035	0.432
	Persistence	0.28	0.758	0.694	−0.024	0.012
Mazarredo NO_2 ($t+3$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.16	0.596	0.827	−0.105	0.159
	MLP1	0.31	−0.315	0.504	0.538	−0.151
	MLP2	0.04	0.380	0.834	0.210	0.177
	Persistence	0.21	0.557	0.831	−0.002	−0.003
Mazarredo O_3 ($t+3$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.43	0.669	0.589	0.112	0.198
	MLP1	0.02	0.580	0.470	0.153	0.843
	MLP2	0.03	0.669	0.499	0.166	0.694
	Persistence	0.50	0.604	0.599	−0.025	0.002
Mazarredo NO_2 ($t+4$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.15	0.519	0.864	−0.086	0.229

Table 4 (*continued*)

Predictand	Model	NMSE	R	FA2	FB	FV
	MLP1	0.0005	0.502	0.865	0.022	0.167
	MLP2	0.005	0.587	0.886	−0.068	0.321
	Persistence	0.25	0.356	0.806	0.010	−0.001
Mazarredo O_3 ($t+4$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.53	0.605	0.538	0.109	0.281
	MLP1	0.04	0.538	0.440	−0.189	0.497
	MLP2	0.01	0.610	0.457	−0.117	0.545
	Persistence	0.70	0.486	0.528	−0.048	0.001
(2b) Mazarredo on the test set (k = 5, 6, 7, 8)						
Mazarredo NO_2 ($t+5$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.21	0.497	0.812	−0.010	0.173
	MLP1	0.002	0.331	0.856	−0.039	0.347
	MLP2	0.0001	0.489	0.861	−0.011	0.319
	Persistence	0.33	0.316	0.747	0.030	0.011
Mazarredo O_3 ($t+5$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.51	0.539	0.558	0.029	0.301
	MLP1	0.0006	0.613	0.509	−0.025	0.375
	MLP2	0.000001	0.657	0.516	0.001	0.410
	Persistence	0.79	0.376	0.505	−0.063	−0.014
Mazarredo NO_2 ($t+6$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.16	0.465	0.859	−0.030	0.183
	MLP1	0.01	0.429	0.857	−0.100	0.489
	MLP2	0.007	0.590	0.877	−0.083	0.335
	Persistence	0.31	0.159	0.776	−0.008	−0.024
Mazarredo O_3 ($t+6$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.68	0.517	0.483	−0.048	0.284
	MLP1	0.0001	0.573	0.464	0.011	0.519
	MLP2	0.0003	0.631	0.486	0.018	0.386
	Persistence	1.16	0.352	0.465	−0.025	−0.019
Mazarredo NO_2 ($t+7$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.16	0.445	0.857	−0.044	0.322
	MLP1	0.007	0.442	0.847	−0.080	0.318
	MLP2	0.006	0.589	0.881	−0.076	0.334
	Persistence	0.34	0.105	0.745	0.004	0.003
Mazarredo O_3 ($t+7$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.69	0.474	0.490	0.022	0.327
	MLP1	0.00003	0.576	0.448	−0.005	0.465
	MLP2	0.000004	0.594	0.472	0.002	0.374
	Persistence	1.20	0.260	0.453	0.000	−0.010
Mazarredo NO_2 ($t+8$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.18	0.444	0.834	−0.038	0.371
	MLP1	0.009	0.439	0.848	−0.093	0.413
	MLP2	0.005	0.571	0.884	−0.072	0.408
	Persistence	0.37	0.189	0.719	0.050	0.006
Mazarredo O_3 ($t+8$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.66	0.418	0.489	0.040	0.373
	MLP1	0.001	0.557	0.491	0.034	0.459
	MLP2	0.0003	0.575	0.456	0.017	0.465
	Persistence	1.15	0.166	0.433	−0.066	−0.035
(3a) Txurdinaga on the test set (k = 1, 2, 3, 4)						
Txurdinaga NO_2 ($t+1$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.05	0.887	0.967	0.010	0.090
	MLP1	0.0001	0.888	0.976	−0.011	0.061
	MLP2	0.0002	0.889	0.980	−0.014	0.066
	Persistence	0.06	0.872	0.970	−0.004	−0.008
Txurdinaga O_3 ($t+1$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.09	0.914	0.865	0.011	0.034

(*continued on next page*)

Table 4 (*continued*)

Predictand	Model	NMSE	R	FA2	FB	FV
	MLP1	0.00004	0.906	0.935	0.007	0.047
	MLP2	0.00008	0.931	0.886	0.009	0.045
	Persistence	0.10	0.902	0.910	−0.002	0.014
Txurdinaga NO_2 ($t+2$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.11	0.743	0.909	0.075	0.209
	MLP1	0.039	0.417	0.780	0.197	0.221
	MLP2	0.061	0.341	0.732	0.245	0.214
	Persistence	0.14	0.694	0.910	0.006	−0.009
Txurdinaga O_3 ($t+2$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.20	0.810	0.767	−0.082	0.098
	MLP1	0.08	0.276	0.839	−0.273	0.008
	MLP2	0.03	0.481	0.447	−0.164	0.006
	Persistence	0.27	0.769	0.783	−0.010	0.006
Txurdinaga NO_2 ($t+3$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.13	0.617	0.889	−0.015	0.223
	MLP1	0.08	0.252	0.742	0.288	0.631
	MLP2	0.09	0.333	0.771	0.301	0.666
	Persistence	0.21	0.490	0.864	−0.010	−0.025
Txurdinaga O_3 ($t+3$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.39	0.674	0.633	0.049	0.141
	MLP1	0.002	0.559	0.639	0.044	0.765
	MLP2	0.008	0.563	0.553	0.088	0.654
	Persistence	0.46	0.645	0.712	0.011	−0.004
Txurdinaga NO_2 ($t+4$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.16	0.573	0.865	0.027	0.273
	MLP1	0.002	0.420	0.870	−0.047	0.454
	MLP2	0.007	0.598	0.887	−0.084	0.501
	Persistence	0.30	0.344	0.787	−0.021	−0.037
Txurdinaga O_3 ($t+4$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.45	0.562	0.620	0.016	0.229
	MLP1	0.01	0.464	0.493	−0.111	0.544
	MLP2	0.03	0.486	0.465	−0.184	0.368
	Persistence	0.58	0.495	0.637	−0.024	0.010
(3b) Txurdinaga on the test set ($k = 5, 6, 7, 8$)						
Txurdinaga NO_2 ($t+5$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.18	0.555	0.837	0.001	0.226
	MLP1	0.007	0.465	0.866	−0.083	0.495
	MLP2	0.008	0.558	0.870	−0.092	0.545
	Persistence	0.36	0.252	0.737	−0.013	−0.036
Txurdinaga O_3 ($t+5$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.52	0.516	0.570	0.046	0.226
	MLP1	0.003	0.565	0.674	0.057	0.468
	MLP2	0.0004	0.628	0.575	0.021	0.330
	Persistence	0.72	0.375	0.581	−0.072	−0.007
Txurdinaga NO_2 ($t+6$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.19	0.467	0.836	0.016	0.315
	MLP1	0.004	0.488	0.870	−0.066	0.416
	MLP2	0.010	0.555	0.867	−0.101	0.415
	Persistence	0.37	0.185	0.724	0.019	0.012
Txurdinaga O_3 ($t+6$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.58	0.461	0.569	0.057	0.237
	MLP1	0.0009	0.531	0.684	0.030	0.433
	MLP2	0.0005	0.574	0.561	0.023	0.344
	Persistence	0.88	0.244	0.526	−0.044	0.000
Txurdinaga NO_2 ($t+7$)	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.21	0.431	0.818	0.041	0.359
	MLP1	0.01	0.560	0.867	−0.097	0.497
	MLP2	0.008	0.583	0.881	−0.088	0.426
	Persistence	0.40	0.161	0.703	−0.002	−0.038

Table 4 (*continued*)

Predictand	Model	NMSE	R	FA2	FB	FV
Txurdinaga O_3 $(t + 7)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.62	0.365	0.519	−0.042	0.325
	MLP1	0.003	0.470	0.645	0.056	0.507
	MLP2	0.00007	0.541	0.558	−0.008	0.361
	Persistence	0.92	0.216	0.517	−0.123	−0.012
Txurdinaga NO_2 $(t + 8)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.19	0.426	0.836	−0.058	0.366
	MLP1	0.005	0.565	0.876	−0.071	0.452
	MLP2	0.007	0.589	0.882	−0.083	0.399
	Persistence	0.40	0.150	0.707	−0.001	−0.061
Txurdinaga O_3 $(t + 8)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.78	0.386	0.515	0.155	0.357
	MLP1	0.01	0.435	0.621	0.098	0.432
	MLP2	0.004	0.556	0.571	0.061	0.448
	Persistence	1.04	0.158	0.497	−0.123	−0.010
(4a) Deusto on the test set ($k = 1, 2, 3, 4$)						
Deusto NO_2 $(t + 1)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.06	0.868	0.948	−0.013	0.099
	MLP1	0.0002	0.875	0.976	−0.013	0.098
	MLP2	0.0002	0.879	0.976	−0.015	0.104
	Persistence	0.07	0.873	0.954	−0.001	−0.005
Deusto O_3 $(t + 1)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.11	0.886	0.811	0.035	0.115
	MLP1	0.20	0.833	0.787	0.140	0.292
	MLP2	0.003	0.896	0.832	0.053	0.127
	Persistence	0.12	0.878	0.822	0.003	0.003
Deusto NO_2 $(t + 2)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.12	0.693	0.901	−0.067	0.116
	MLP1	0.84	0.219	0.349	0.834	0.527
	MLP2	1.38	−0.223	0.279	1.23	0.363
	Persistence	0.15	0.683	0.901	0.008	−0.003
Deusto O_3 $(t + 2)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.22	0.755	0.727	0.094	0.222
	MLP1	0.34	0.561	0.403	0.562	0.165
	MLP2	1.32	0.521	0.262	0.997	0.896
	Persistence	0.26	0.713	0.706	0.013	0.000
Deusto NO_2 $(t + 3)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.15	0.606	0.871	−0.032	0.162
	MLP1	0.02	0.210	0.793	0.147	0.624
	MLP2	0.07	0.317	0.797	0.269	1.00
	Persistence	0.21	0.545	0.849	0.020	0.004
Deusto O_3 $(t + 3)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.31	0.631	0.683	0.092	0.304
	MLP1	0.009	0.523	0.652	−0.094	0.803
	MLP2	0.005	0.368	0.628	−0.069	1.031
	Persistence	0.41	0.542	0.618	0.016	0.002
Deusto NO_2 $(t + 4)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.17	0.550	0.850	−0.051	0.089
	MLP1	0.0004	0.545	0.881	−0.021	0.454
	MLP2	0.0005	0.581	0.896	−0.022	0.394
	Persistence	0.26	0.407	0.809	0.028	−0.023
Deusto O_3 $(t + 4)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.43	0.507	0.629	0.128	0.374
	MLP1	0.04	0.538	0.611	0.206	0.676
	MLP2	0.04	0.566	0.622	0.202	0.545
	Persistence	0.51	0.443	0.590	−0.018	0.014

(*continued on next page*)

Table 4 (*continued*)

Predictand	Model	NMSE	R	FA2	FB	FV
(4b) Deusto on the test set ($k = 5, 6, 7, 8$)						
Deusto NO_2 $(t + 5)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.20	0.485	0.819	−0.088	0.044
	MLP1	0.142	−0.118	0.567	0.370	0.017
	MLP2	0.00004	0.496	0.871	−0.006	0.484
	Persistence	0.30	0.322	0.779	0.045	0.012
Deusto O_3 $(t + 5)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.54	0.409	0.573	0.149	0.350
	MLP1	0.04	0.471	0.604	0.196	0.791
	MLP2	0.06	0.532	0.597	0.249	0.713
	Persistence	0.61	0.345	0.563	−0.049	0.002
Deusto NO_2 $(t + 6)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.19	0.490	0.839	−0.056	0.155
	MLP1	0.002	0.501	0.869	−0.042	0.459
	MLP2	0.002	0.528	0.876	−0.046	0.379
	Persistence	0.33	0.269	0.753	0.041	0.024
Deusto O_3 $(t + 6)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.52	0.359	0.585	0.114	0.423
	MLP1	0.01	0.399	0.605	0.112	0.671
	MLP2	0.03	0.468	0.576	0.178	0.556
	Persistence	0.69	0.253	0.530	−0.048	−0.007
Deusto NO_2 $(t + 7)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.20	0.452	0.819	−0.063	0.126
	MLP1	0.002	0.488	0.871	−0.041	0.483
	MLP2	0.002	0.520	0.873	−0.042	0.407
	Persistence	0.33	0.235	0.740	0.052	0.016
Deusto O_3 $(t + 7)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.58	0.332	0.568	0.139	0.413
	MLP1	0.03	0.360	0.573	0.178	0.787
	MLP2	0.02	0.452	0.603	0.134	0.713
	Persistence	0.75	0.219	0.511	−0.028	−0.003
Deusto NO_2 $(t + 8)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.19	0.472	0.817	−0.095	0.128
	MLP1	0.005	0.475	0.854	−0.069	0.517
	MLP2	0.004	0.501	0.864	−0.060	0.448
	Persistence	0.32	0.262	0.740	0.051	0.023
Deusto O_3 $(t + 8)$	Observation	0.00	1.000	1.000	0.000	0.000
	LR	0.60	0.326	0.550	0.204	0.518
	MLP1	0.03	0.358	0.579	0.174	0.837
	MLP2	0.065	0.426	0.551	0.253	0.843
	Persistence	0.79	0.165	0.499	−0.003	0.004

that there were less valid data measured at Deusto during the period 1993–1994.

In the same way, the MLP2 model produced the most accurate forecasts of $NO_2(t + 6)$ and $O_3(t + 6)$. This was shown by the values of the Model Validation Kit, where using the MLP2 model the lowest values of NMSE (around 0) and the greatest values of R (0.468, 0.648) and FA2 (48.6–87.7%) were obtained. Furthermore, the MLP1 model performed also better than the LR model and persistence. Similar situations were measured for the four stations, but the results for Deusto were not as accurate as those obtained at the other stations.

Finally, in the forecasts of $NO_2(t + 7)$ and $O_3(t + 7)$, the same situation was observed at all stations: as the best values of the statistics of the Model Validation Kit were those obtained for the MLP2 model, this model was the best performer in the forecasts of O_3 and NO_2 levels 7 h ahead. Furthermore, the MLP1 model was shown to be more efficient than the LR and persistence. Similarly, it could be concluded that the MLP2 is the most accurate prognostic model to forecast $NO_2(t + 8)$ and $O_3(t + 8)$.

In order to illustrate these results, Figs. 2 and 3 show graphically the performance of observation, LR, MLP1, MLP2 and persistence in the forecasts of $NO_2(t + k)$ and $O_3(t + k)$, $k = 1, 2,…,8$, by calculating the statistics of the Model Validation Kit on the test set at Elorrieta.

In the majority of cases it was demonstrated that the FB and FV values corresponding to the outputs of the

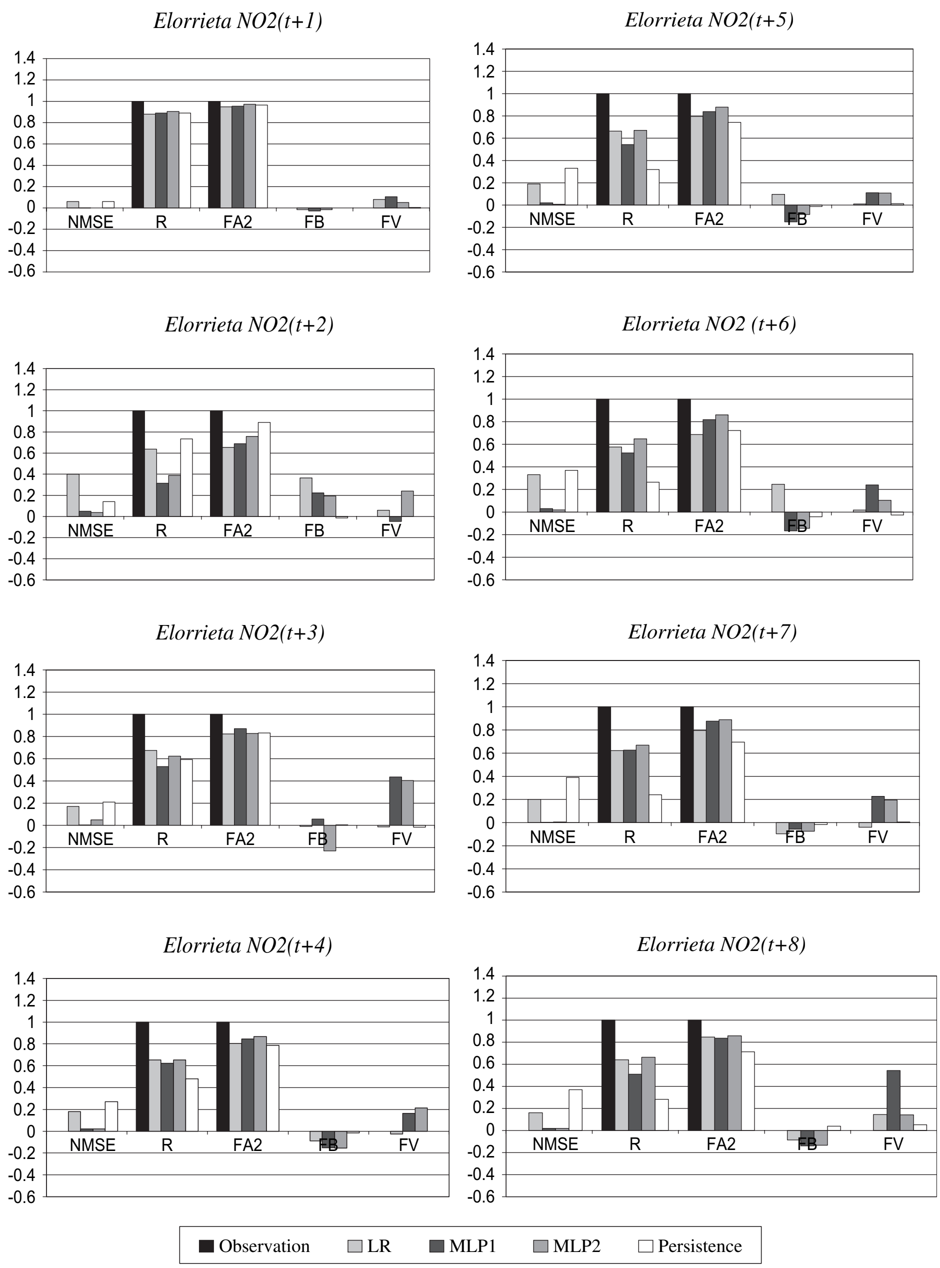

Fig. 2. Performance of observation, LR, MLP1, MLP2 and persistence in the forecast of $NO_2(t+k)$, $k = 1, 2,\ldots,8$ at Elorrieta (1994).

MLP2 model were greater than the corresponding values for the outputs of the LR model. This indicates a higher variability on the outputs obtained with the MLP2 model, and overprediction and underprediction phenomena could be expected.

Additionally, the response obtained by applying the MLP2 model was analysed. For example, Fig. 4 shows scatterplots of the forecasts and observed values of $O_3(t+1)$ during 1993–1994 at Txurdinaga, where A represents the forecasted value (output) obtained with the MLP2 model and T represents the corresponding observation (target). Similarly, Fig. 5 illustrates scatterplots of the forecasts and observed values of $O_3(t+8)$ during 1993–1994 at Txurdinaga. The two lines in each plot indicate the perfect correspondence between the observation and the forecasted values and the best linear fit through the data. In order to illustrate the behaviour of the performance on each subset of the early stopping technique, three regressions were performed: a first for the training set, a second for the validation set and a third for the test set. Although the values of the correlation coefficients differed slightly among the three sets, the fit was less accurate in the case of outputs from the validation set. This lesser accuracy in advanced time forecasts such as $O_3(t+8)$ could be explained by the selection of the validation set, which was formed by

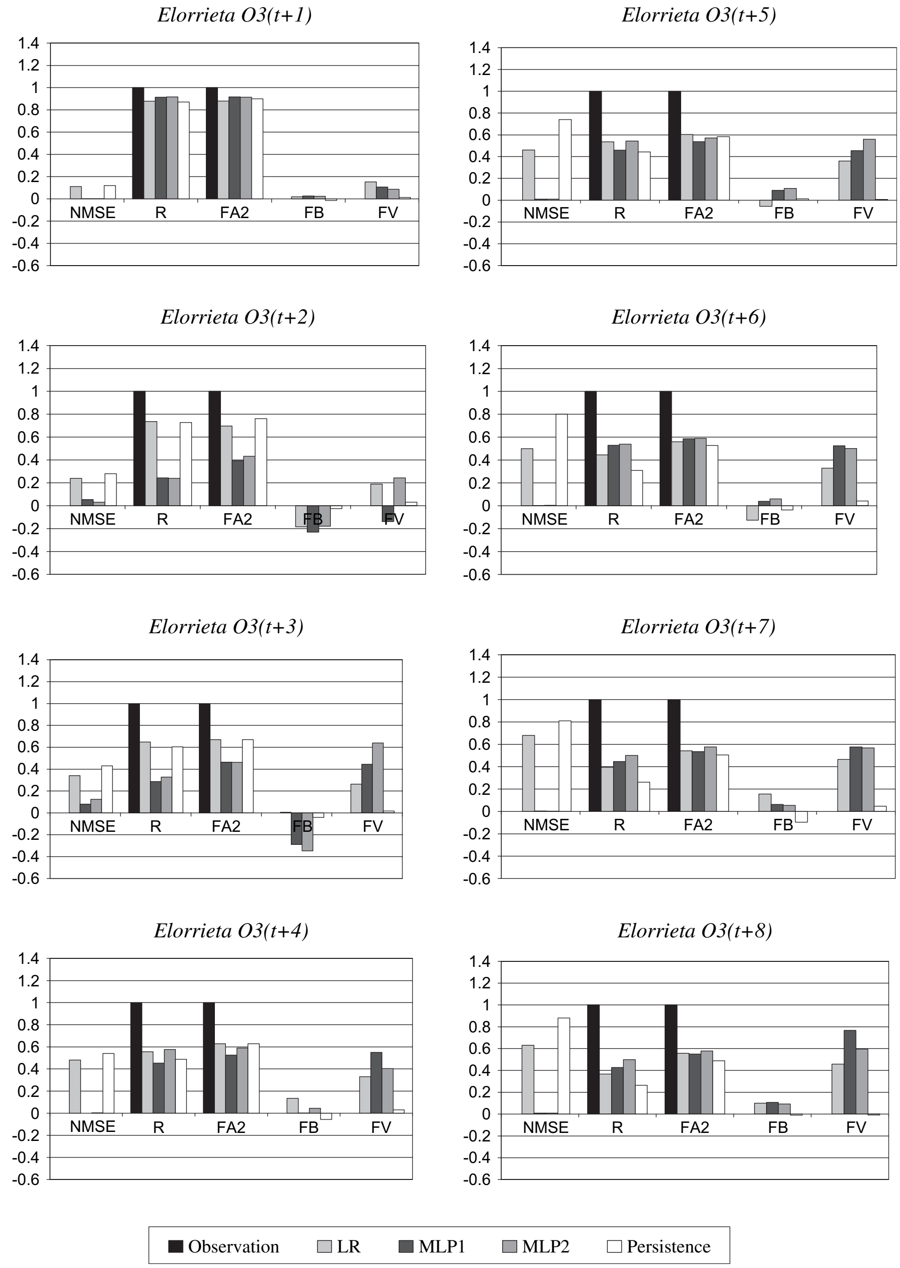

Fig. 3. Performance of observation, LR, MLP1, MLP2 and persistence in the forecast of $O_3(t+k)$, $k = 1, 2,\ldots,8$, at Elorrieta (1994).

the 15% of data from 1993. However, the values of the correlation coefficients obtained for the outputs corresponding to the training set and the test set were in agreement with the generalization of the model. Furthermore, the values of intercept and slope demonstrate that the lowest intercepts generally coincide with the highest slopes, which indicates that overprediction of low values and underprediction of high values are occurring.

The MLP2 model proved its ability in the forecasts of $NO_2(t+k)$ and $O_3(t+k)$ concentrations ($k = 1, 2,\ldots,8$), except in the forecasted values two and three hours ahead. In general, the MLP2 model performed consistently for the different air pollutants at the different stations, except at Deusto, where the performance was slightly worse. The reason for this could be explained by the number of valid cases for this study, which was reduced to 70% in the case of ozone at Deusto. In order to achieve better results, it would be advisable to replace the missing values in the datasets in future research.

6. Conclusions

A joint study of the values obtained from the statistics of the Model Validation Kit showed that multilayer perceptron-based models performed better

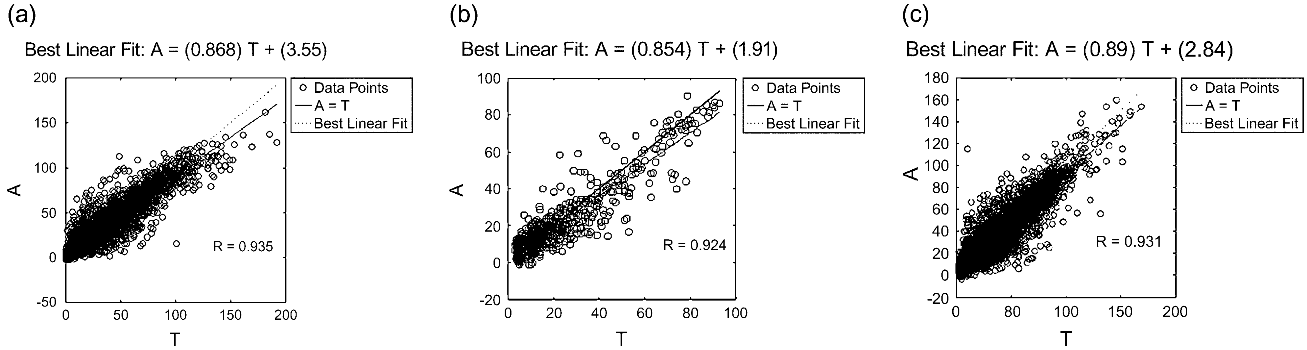

Fig. 4. Scatterplots of observed values (x-axis) and forecasted values (y-axis) of $O_3(t+1)$ at Txurdinaga on the (a) training data, (b) validation data and (c) test data, for the MLP2 model. A is the output or forecast using the MLP2 model, T is the target or observation and R is the correlation coefficient.

than the multiple linear regression based model in 75% of cases (in six forecasts up to eight) in the Bilbao area. Moreover, for $k = 1, 4, 5, 6, 7, 8$ h ahead the MLP2 model provided the most accurate forecasts of O_3 and NO_2 at time $t + k$ in the studied area. Nevertheless, for $k = 2, 3$ the most accurate forecasts of O_3 and NO_2 at time $t + k$ were provided by the LR model (Agirre, 2003).

On the one hand, the forecasts obtained using the MLP models were better than those forecasts obtained after the application of the LR model. Consequently, the multilayer perceptron-based models perform better than the linear regression based model in nonlinear relationships like that involving O_3, NO_2, meteorological variables and traffic variables. On the other hand, due to the introduction of the four additional input variables $\sin(2\pi h/24)$, $\cos(2\pi h/24)$, $\sin(2\pi d/7)$ and $\cos(2\pi d/7)$, $h = 1, 2,\ldots,24$, $d = 1, 2,\ldots,7$, the MLP2 model provided better results than the MLP1 model, which emphasises the importance of taking into account the seasonal character of O_3 and NO_2.

The causes of the uncertainties in the forecasts of O_3 and NO_2 2 and 3 h ahead are unknown. The application of the MLP2 model to a new dataset formed by data from two recent years would be interesting. In this way, the failure of the model in the forecasts of $O_3(t + k)$ and $NO_2(t + k)$, when $k = 2, 3$, could be resolved on a new database and, at the same time, the model could be performed on a recent dataset. Furthermore, in order to improve the performance of the MLP2 model, a wider dataset could be selected, and the new training, validation and test sets could be properly chosen. Finally, the application of the MLP2 model could be extended to the forecasts of NO and CO hourly levels in the Bilbao area. It could also be applied in different areas with similar air quality problems and meteorological and traffic influences. Future research will focus on the study of the development of the MLP2 model in other stations within the Basque Government air pollution network.

The advantages of neural networks are that they do not require very exhaustive information about air pollutants, reaction mechanisms, meteorological parameters or traffic flow and that they have the ability of allowing nonlinear relationships between very different

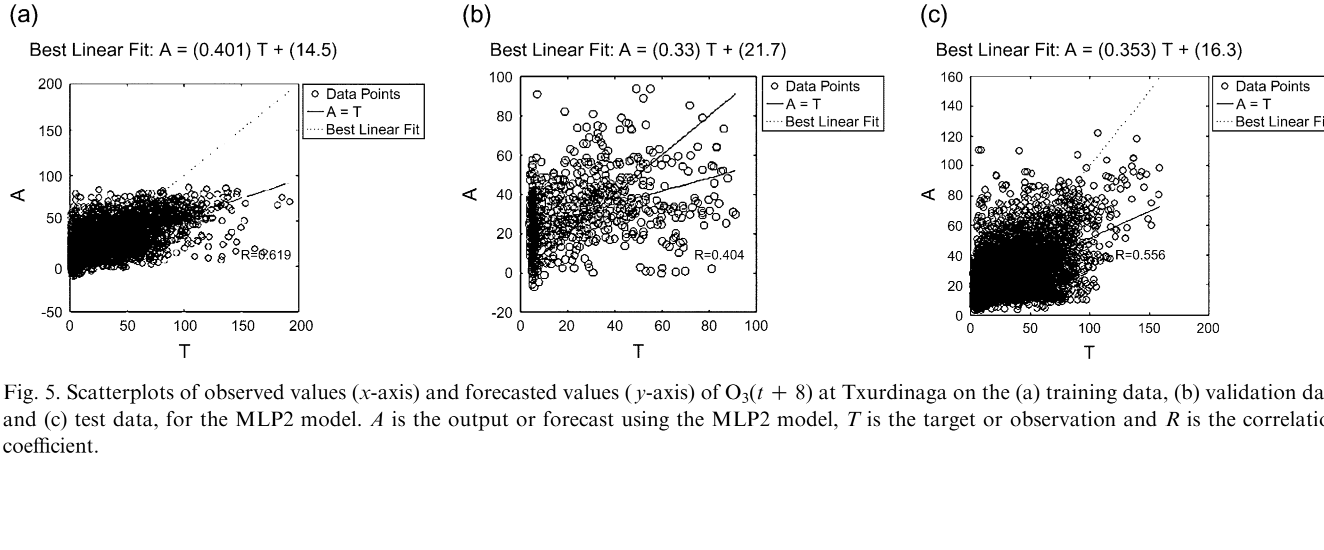

Fig. 5. Scatterplots of observed values (x-axis) and forecasted values (y-axis) of $O_3(t+8)$ at Txurdinaga on the (a) training data, (b) validation data and (c) test data, for the MLP2 model. A is the output or forecast using the MLP2 model, T is the target or observation and R is the correlation coefficient.

predictor variables. These facts and the quality of the results they have provided are the reasons that make them more attractive to apply than other models.

Acknowledgements

The authors wish to thank the Environmental Department of the Basque Government for providing data from their air pollution network and also the Local Municipality of Bilbao for making data from the traffic network available.

References

Agirre, E., 2003. Elaboration of a statistical prognostic model for short-term real-time prediction of O_3 and NO_2 levels in the city of Bilbao. PhD University of the Basque Country, Bilbao.

Amari, S., Murata, N., Müller, K., Finke, M., Yang, H.H., 1997. Asymptotic statistical theory of overtraining and cross-validation. IEEE Transactions on Neural Networks 85, 985–996.

Basque Government, 2001. Air quality in the Basque Country in the period 1996–2000. Publication Service of the Basque Government, Vitoria-Gasteiz.

Borrego, C., Tchepel, O., Barros, N., Miranda, A.I., 2000. Impact of road traffic emissions on air quality of the Lisbon region. Atmospheric Environment 34, 4683–4690.

Boznar, M., Lesjak, M., Mlakar, P., 1991. A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment 27B (2), 221–230.

Cardelino, C., Chang, M., St. John, J., Murphey, B., Cordle, J., Ballagas, R., Patterson, K., Stogner, J., Zimmer-Dauphinee, S., 2001. Ozone predictions in Atlanta, Georgia: analysis of the 1999 ozone season. Journal of the Air and Waste Management Association 51, 1227–1236.

Cassmassi, J.C., 1998. Objective ozone forecasting in the South Coast Air Basin: updating the objective prediction models for the late 1990s and southern California ozone study (SCOS97-NARSTO) application. Proceedings of the 12th Conference on Numerical Weather Prediction, Phoenix, AZ. American Meteorology Society, Boston, MA, pp. 54–58.

Chelani, A.B., Chalapati Rao, C.V., Phadke, K.M., Hasan, M.Z., 2002. Environmental Modelling and Software 17, 161–168.

Chock, D.P., Chang, T.Y., 1999. The impact of an 8 h ozone air quality standard rog and NO_x controls in Southern California. Atmospheric Environment 33, 2471–2485.

Elkamel, A., Abdul-Wahab, S., Bouhamra, W., Alper, E., 2001. Measurement and prediction of ozone levels around a heavily industrialized area: a neural network approach. Advances in Environmental Research 5, 47–59.

European Commission, 1994. The evaluation of models of heavy gas dispersion. Model evaluation group seminar. Office for Official Publications of the European Communities, L-2985, Luxemburg.

European Communities, 1998. Europe's environment: the second assessment. European Environment Agency, Office for Official Publications of the European Communities.

Finlayson-Pitts, B.J., Pitts, J.N., 1986. Atmospheric chemistry: fundamentals and experimental techniques. Wiley, New York.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. Atmospheric Environment 32 (14/15), 2627–2636.

Gardner, M.W., Dorling, S.R., 1999. Neural network modelling and prediction of hourly NO_x and NO_2 concentrations in urban air in London. Atmospheric Environment 33, 2627–2636.

Gardner, M.W., Dorling, S.R., 2000. Statistical surface ozone models: an improven methodology to account for non-linear behaviour. Atmospheric Environment 34, 21–34.

Hagan, M.T., Demuth, H.B., Beale, M., 1996. Neural network design. PWS Publishing Company, Boston.

Hanna, S.R., Strimaitis, D.G., Chang, J.C., 1991. Hazard response modeling uncertainty (a quantitative method). User's Guide for Software for Evaluating Hazardous Gas Dispersion Models. American Petroleum Institute, Washington.

Ibarra, G., Elias, A., Agirre, E., Uria, J., 2001a. Short-term real-time forecasting of hourly ozone, NO_2 and NO levels by means of multiple linear regression modelling. Gate to EHS. Ecomed Publishers, pp. 1–7.

Ibarra, G., Madariaga, I., Elias, A., Agirre, E., Uria, J., 2001b. Long-term changes of ozone and traffic in Bilbao. Atmospheric Environment 35, 5581–5592.

Jorquera, H., Perez, R., Aldo, C., Espejo, A., Letelier, M.V., Acuña, G., 1998. Forecasting ozone daily maximum levels at Santiago, Chile. Atmospheric Environment 32 (20), 3415–3424.

Kuang-Jung, H., 1992. Time series analysis of the interdependences among air pollutant. Atmospheric Environment 26B, 491–503.

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

Mayer, H., 1999. Air pollution in cities. Atmospheric Environment 33, 4029–4037.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5, 115–133.

Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6, 525–533.

Perez, P., Trier, A., Reyes, J., 2000. Prediction of $PM_{2.5}$ concentrations several hours in advance using neural networks in Santiago, Chile. Atmospheric Environment 34, 1189–1196.

Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review 65, 386–408.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning internal representations by error propagation. MITT Press, Cambridge.

Rye, P.J., 1995. Modelling photochemical smog in the Perth region. Mathematical Computational Modelling 21, 117–127.

Sarle, W.S., 1995. Stopped training and other remedies for overfitting. Proceedings of the 27th Symposium on the Interface.

Saunders, S.M., Jenkin, M.E., Derwent, R.G., Pilling, M.J., 1997. WWW site of a Master Chemical Mechanism for use in tropospheric chemistry models. Atmospheric Environment 31, 1249.

Scheffe, R.D., Morris, R.E., 1993. A review of the development and application of the urban airshed model. Atmospheric Environment 27B, 23–39.

Simpson, R.W., Layton, A.P., 1983. Forecasting peak ozone levels. Atmospheric Environment 17, 1649–1654.

Yi, J., Prybutok, V.R., 1996. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. Environmental Pollution 92, 349–357.

Zanetti, P., 1990. Air pollution modelling: theories, computational methods and available software. Computational Mechanics Publications, Southampton, Boston.