Available online at www.sciencedirect.com
ScienceDirect
Chemosphere 70 (2008) 1190–1195
CHEMOSPHERE
www.elsevier.com/locate/chemosphere

Prediction of CO maximum ground level concentrations in the Bay of Algeciras, Spain using artificial neural networks

M.L. Martín [a,*], I.J. Turias [a], F.J. González [a], P.L. Galindo [a], F.J. Trujillo [a], C.G. Puntonet [b], J.M. Gorriz [b]

[a] *EPS de Algeciras, Universidad de Cádiz, Avda Ramón Puyol s/n, 11202 Algeciras (Cádiz), Spain*
[b] *E.T.S.I.I.T., Universidad de Granada, Spain*

Received 10 May 2007; received in revised form 14 August 2007; accepted 16 August 2007
Available online 24 October 2007

Abstract

The region of the Bay of Algeciras is a very industrialized area where very few air pollution studies have been carried out. The main objective of this work has been the use of artificial neural networks (ANNs) as a predictive tool of high levels of ambient carbon monoxide (CO). Two approaches have been used: multilayer perceptron models (MLPs) with backpropagation learning rule and k-Nearest Neighbours (k-nn) classifiers, in order to predict future peaks of carbon monoxide. A resampling strategy with twofold cross-validation allowed the statistical comparison of the different topologies and models considered in the study. The procedure of random resampling permits an adequate and robust multiple comparisons of the tested models and allow us to select a group of best models.

Keywords: Air pollution forecasting; Multiple comparison; k-nn; Multilayer perceptron

1. Introduction

Bay of Algeciras region has one of the main industrial estates of Andalusia and Spain, with urban areas that reach more than 250 000 inhabitants ('agglomeration' in the sense of 96/62 EU Directive). This important industrial activity requires the necessary control of its environmental impact. Thus, 96/62 EU Directive states that Member States shall take the necessary measures to ensure compliance with the limit values and shall draw up action plans indicating the measures to be taken in the short term where there is a risk of the limit values and/or alert thresholds being exceeded, in order to reduce that risk and to limit the duration of such an occurrence. Therefore, the prediction of air pollutants has become an important task in the recent years, because the estimation of ground level concentrations gives valuable information for air pollution reduction policies. The dispersion mechanism of atmospheric pollutants in urban areas is quite complex and it depends on different factors: meteorological conditions, orography, 'heat island' effect, etc. All these factors, some of which are out of human control, are decisive in air quality conditions. It is well-known that air pollution episodes can cause important toxic effects especially on high-risk population. Then, the ability to predict the occurrence of peak concentrations will contribute to reduce these effects, either reducing emissions or warning population.

The main purpose of the present work is the application of classification techniques, such as artificial neural networks (ANNs) and k-nn method to identify peaks of carbon monoxide (CO) concentrations in the Bay of Algeciras region, for the period 1999–2001. The importance of input weights has been determined in order to study the relevance of input variables to neural models (Garson, 1991).

A procedure of resampling simulation was designed to avoid variation coming from different sources, thus

* Corresponding author. Tel.: +34 956 028015; fax: +34 956 028014.
E-mail address: luz.martin@uca.es (M.L. Martín).

0045-6535/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chemosphere.2007.08.039

independence and randomness was guaranteed. In order to estimate the error produced by different models, we adopt the procedure of twofold cross-validation.

A detailed overview of the application of neural techniques on atmospheric sciences can be found in the work of Gardner and Dorling (1998). These authors conclude that models based on ANNs give better results than linear methods, especially when the problem under study is non-linear. Different methods have been used in the prediction of atmospheric pollution: persistence models (Perez et al., 2000), regression models (Comrie and Diem, 1999), ANNs (Jorquera et al., 1998; Elkamel et al., 2001; Kolehmainen et al., 2001; Kukkonen et al., 2003; Brunelli et al., 2007; Dutot et al., 2007). Comrie and Diem (1999) studied the factors that influence CO concentrations in the area under study (meteorological variables, previous concentrations, traffic density) in order to develop a multiple regression model. Pelliccioni and Poli (2000) used multilayer perceptron models (MLP) for the prediction of CO and nitrogen dioxide (NO_2). Viotti et al. (2002) used an ANN model with a hidden layer to predict short-term and medium-term air pollutant concentrations (CO, ozone and benzene) in an urban area. Kolehmainen et al. (2001) compared different prediction models from meterological and NO_2 concentration data measured in the city of Stocolm for the period 1994–1998. These authors conclude that MLP models predict with a better accuracy than linear and Self Organising Map (SOM) models. Perez (2001) introduced a study to predict hourly mean sulphur dioxide (SO_2) concentrations in the city of Santiago de Chile. They compared the predictions made by persistence models, linear regression models and ANNs. The input variables were SO_2 concentrations, temperature, relative humidity and wind speed. Prediction errors were between 30% and 60%, for short-term and long-term predictions, respectively.

2. Meteorological and air pollution data

CO is a primary air pollutant produced by incomplete combustion processes. In urban areas, traffic is its main source, while its reaction with the OH radical is the main removal mechanism. CO concentrations are usually well correlated with traffic density. 2000/69 EU Directive sets the maximum 8-h CO concentration in 10 000 μg m^{-3}. CO concentration in a given time t, CO(t), depends on three main factors: dispersion capacity of the atmosphere, emissions, and previous CO concentrations (accumulated CO) (Maffeis, 1999). CO emissions depend on ambient temperature and source features. This gaseous pollutant exhibits a clear cyclical behaviour, with a characteristic variation in the daytime and night-time and during the week (González Gallero, 2003). Two significant peaks can be observed during working days, which correspond with the hours of higher traffic intensity, the first between 8:00 and 10:00 am and the second one between 7:00 and 9:00 pm (Fig. 1a). In contrast, weekends are characterized

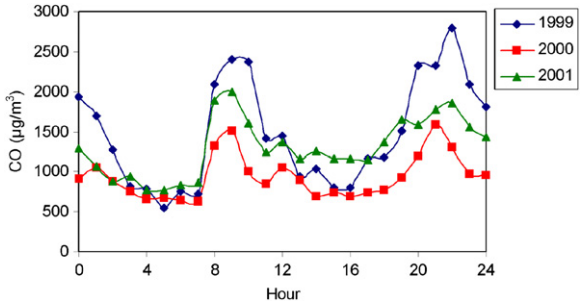

Fig. 1a. CO daily mean concentration (1999, 2000 and 2001).

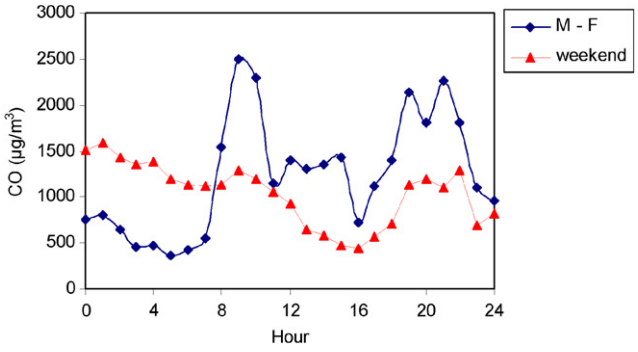

Fig. 1b. CO daily mean concentration for working days and weekends (1999).

by lower CO concentration levels during day-time and higher levels during night-time (Fig. 1b).

Data used in this work have been provided by the 'Consejería de Medio Ambiente' from the Andalusian Regional Government for a period of three years (1999–2001). CO concentration data are from the monitoring station located at the town of Algeciras, which gives data every half an hour. Only the exogen variables measured at the same monitoring station have been used in the present study: wind speed (V) and wind direction (D), temperature (T) and NO concentration. In order to avoid discontinuities, wind direction was transformed using the expression $1 + \sin(D + \pi/4)$ (Chelani et al., 2002).

The maximum 8-h CO concentration of 10 000 μg m^{-3} established by the 2000/69 EU Directive has not been surpassed during the period of analysis. However, local authorities are very interested in assessing the impact of urban traffic on air quality in Algeciras.

In a first step, data have been validated removing data measured during periods of malfunction of the monitoring station. Linear interpolation was used to fill data series when gap was not too large.

3. Prediction models

The classification analysis allows the development of models that are able to predict if one object belongs to a specific class or category, considering different features of the object. The data matrix has at least one categorical

variable, which sets the category the object belongs to and that is the variable to be predicted, and one or more variables that describe other features of the objects and are used as predictive variables. In the present study, the category represents the occurrence (or not) of an episode with maximum levels of a given air pollutant. The two techniques applied here are described below.

3.1. k-nn method

k-nn method is a non-parametric classification technique, based on the calculation of the distance of the pattern one wants to classify (here, the occurrence or not of a peak of concentration) from the k nearest neighbours, each of which are grouped or classified. k is generally a small even number (1, 3, or 5). Finally, the pattern will be associated with the group with more neighbour patterns. This algorithm strongly depends on the initial classification that is a known input. If these patterns are sufficiently representative of what is happening in the universe of patterns, the algorithm will perform better. This method does not require of a specific training stage (Duda et al., 2001), but it needs a sufficient number of reference points for the different classes. The Euclidean distance between the pattern and the reference points is calculated in order to find the nearest neighbours. Once the distances are ordered, the classes with the smallest distances are finally selected. A usually high computation time is the main drawback of this method.

3.2. Multilayer neural network with backpropagation

Multilayer perceptron (MLP) with learning based on error backpropagation (Backpropagation neural network, BPNN) is the most widely used neural method. This learning model can be applied to regression or classification problems (Bishop, 1995), in which there is no a priori knowledge of the model. In this model, the network is a dynamical system that changes as the learning algorithm sets. This algorithm tries to find the weights that codify the knowledge the network must has. Once this knowledge is learnt, the network is used with new input vectors to which the system will give a response with different outputs. The network will then have a generalization capability that must be measured. The network used in the present study has been trained using the Levenberg–Marquardt algorithm, minimizing the mean squared error (MSE) of the difference between the value obtained by the network in each step and the objective value.

4. Methodology

Two non-parametric classification techniques have been applied, k-nn and ANNs, based on autoregressive schemes, as shown in Table 1. The models tested have been described in Table 2.

Table 1
Autoregressive scheme (T: temperature, V: wind speed, D: wind direction, NO: nitric oxide concentration, h: hour of day , d: weekday)

Input variables	Output
$T(t-n)$, $V(t-n)$, $D(t-n)$, NO$(t-n)$, $h(t-n)$, $d(t-n)$, CO$(t-n)$...$T(t-1)$, $V(t-1)$,	Peak$_{(t)}$
$D(t-1)$, NO$(t-1)$, $h(t-1)$, $d(t-1)$, CO$(t-1)$	Yes/No

Table 2
Prediction models used in this study (k: number of neighbours; *nhiddens*: number of units in the hidden layer).

	k-means	ANNs 3 layers
Model	1-knn1, 2-knn3, 3-knn5	4-BP1, 5-BP2, 6-BP4, 7-BP10, 8-BP20
	$k = 1, 3, 5$	*nhiddens* $= 1, 2, 4, 10, 20$

The aim of this study is to present the capability to predict peaks higher than an established threshold, doing different experiments with different values of the following factors: (i) Size (lag) of the autoregressive window; (ii) selection of the input variables: meteorological variables and other air pollutant concentrations.

4.1. Analysis of input relevance

The study of the relevance of the input variables for the forecasting of CO peaks has been done using correlation analysis and the calculation of the relative importance of each input to the neural models through the network weights.

From the correlation analysis it can be observed that the highest correlation values are found for a lag which equals 1, in contrast to the values 3 and 6, also calculated, as shown in Table 3. CO concentration shows a negative correlation with temperature ($R = -0.307$ and -0.274 for $t-1$ and $t-3$, respectively) and wind speed ($R = -0.307$ and -0.274 for $t-1$ and $t-3$, respectively). Similarly, it is found a positive correlation between 8 h CO mean concentration values and the previous CO and NO concentration values ($t-1$) ($R = 0.350$ and 0.270, respectively). Although the correlation values are not very high, it seems reasonable to include these variables as model input variables due to the tendencies found. However, the correlation analysis results should not be determinant for the selection of variables as they only show the individual relationships between each pair of variables (Elkamel et al., 2001).

The connection weights of the artificial neural network were used to determine the relative importance of

Table 3
Correlation coefficient (R) values between CO concentration at time t and lagged imput variables

Lag	CO	T	V	D	NO	h	d
1	0.350	−0.307	−0.315	0.306	0.270	0.260	0.004
3	0.560	−0.274	−0.307	0.044	0.003	−0.166	−0.011
6	0.370	−0.013	0.043	0.002	−0.025	−0.003	−0.007

the meterological variables (T, V, and D) and the pollutant concentrations (NO and CO) lagged at different times. Additionally, two new variables have been included: time of the day (h) and day of the week (d).

To assess the relative importance of the different predictor variables quantitatively, the connection weights of the prepared neural network model are used according to the procedure developed by Garson (1991). The procedure essentially involves partitioning of the connection weights of each hidden neuron into components associated with each input neuron (Elkamel et al., 2001). The equation proposed by Garson for determining the relative importance of an input is given by

$$\mathrm{IM}(X_p) = \frac{\sum_{j=1}^{N}[(|W|p_j \sum_{k=1}^{N_p}|W|p_{j,k})|O|_j]}{\sum_{i=1}^{N_p}\{\sum_{j=1}^{N}[(|W|p_{i,j}\sum_{k=1}^{N_p}|W|p_{i,j,k})|O|_j]\}}$$

where $\mathrm{IM}(X_p)$ is the importance measure for the pth input variable X_p. N_p is the number of input variables to the neural network. The term $|W|p_j$ is the absolute value of the weight in the neural network corresponding to the pth input variable and the jth hidden layer ($j = 1, 2,..., N$). The term $|O|_j$ is the absolute value of the output layer weight corresponding to the jth hidden layer.

4.2. Selection procedure of the best model

In order to assess the generalization capability of the different models tested, the size of the autoregressive window has been chosen from the autocorrelation function of the data series. As shown in Fig. 2, the higher autocorrelation values are found for lags equal to 1, 3 and 6. A threshold of 1500 $\mu g/m^{-3}$ has been considered. This election assures a minimum 10% of data over the threshold. Then, a first experiment was developed to prove the model efficiency when only the lagged CO signal for $n = 1$, 3 and 6 was used as input variable. Afterwards, a second experiment was done, in which models were tested using the available exogen variables (T, D, V and NO) as additional input variables. As mentioned before, h and d were also included. These two variables were transformed to cyclic ones using sine components in order to avoid any non-physical discontinuities.

Statistical methods analyze how the models behave on average. Furthermore, it is necessary to consider the performance of a model over all the training sets that might be drawn from the underlying distribution. Only a limited sample of data is available and a complete strategy of experiments must be described to guarantee the independence of the results. In accordance with the work of Pizarro et al. (2002), a randomised procedure has been designed to control the different sources of variation in order to compare the different models.

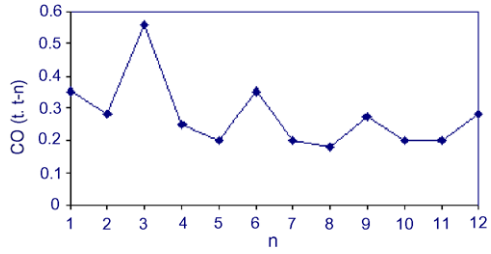

Fig. 2. Autocorrelation function for 8-h CO concentration series.

There are many methods to estimate the generalization performance of a model (hold-out, cross-validation, leave-one-out, penalization strategies,...). The complete strategy used in this paper repeats 30 times a similar process: random splitting of data into a pair of equal sized portions (training and test sets) and twofold cross-validation for the estimation of the generalization classification success index (CSI), i.e. the total sum of peaks/no peaks successfully classified for each case. The parameters of each model were estimated using one of the groups (the training set) while its performance was tested using the remaining one (the test set). Therefore, the performance was measured over test data not used in the training or design of the model. This process was repeated twice each time, swapping the sets and averaging the results over these two runs. For a given training and test set, each algorithm is trained for 10 times (to avoid internal randomness).

Analysis of variance (ANOVA) (Scheffe, 1959) has been used to test the null hypothesis taking into account certain assumptions, i.e. all populations are approximately normal and the samples from each population should be random and independent. The first assumption has been verified using the Kolmogorov–Smirnov test. Bonferroni multiple comparison procedure (Jobson, 1991) has been used to decrease the probability of making at least one Type I error. This approach is a follow-up method to ANOVA based on the following: if c comparisons are to be made, each one with confidence coefficient $(1 - a/c)$, then the maximum value of overall probability of making one or more Type I errors is a (statistical significance is fixed at 0.1). If the difference between the sample means exceeds a critical threshold, there is sufficient evidence to conclude that the population means differ.

5. Results and discussion

Table 4 shows the results obtained for the 8 hours ahead prediction of the 8-h CO concentrations higher than the threshold 1500 $\mu g\ m^{-3}$. It gives the values from the different models tested, without the information provided by the exogenous variables. On the other hand, Table 5 shows the results for 8 h ahead prediction of the 8-h CO concentrations but using the different exogenous variables selected (wind speed and wind direction, temperature and NO concentration). In both cases, results are shown for lags $n = 1$, 3 and 6.

A significant improvement is obtained when exogenous information is used as input variables in the models tested. Regarding the autoregressive window (previous CO values), the best results are obtained for $n = 3$, that is, if the

Table 4
Results of the different models without exogenous information (prediction horizon: 8 h)

Model	$n = 1$		$n = 3$		$n = 6$	
	CSI	σ	CSI	σ	CSI	σ
1-knn1	0.5593	0.0114	0.6891	0.0112	0.7180	0.0148
2-knn3	0.5695	0.0135	0.7242	0.0108	0.7547	0.0094
3-knn5	0.5878	0.0143	0.7389	0.0148	0.7689	0.0151
4-BP1	0.6329	0.0093	0.7389	0.0110	0.7411	0.0221
5-BP2	0.6320	0.0126	0.7405	0.0201	0.7493	0.0209
6-BP4	0.6255	0.0112	0.7528	0.0141	0.7479	0.0194
7-BP10	0.6098	0.0170	0.7394	0.0174	0.7265	0.0175
8-BP20	0.6055	0.0146	0.7117	0.0266	0.6918	0.0215

CSI: classification success index and σ: Standard deviation of CSI.

Table 5
Results of the different models with exogenous information (prediction horizon: 8 h)

Model	$n = 1$		$n = 3$		n = *6*	
	CSI	σ	CSI	σ	CSI	σ
1-knn1	0.6982	0.0108	0.7287	0.0156	0.7160	0.0129
2-knn3	0.7393	0.0146	0.7662	0.0167	0.7488	0.0164
3-knn5	0.7533	0.0161	0.7826	0.0138	0.7602	0.0144
4-BP1	0.7724	0.0136	0.7887	0.0151	0.7732	0.0176
5-BP2	0.7647	0.0133	0.7847	0.0175	0.7592	0.0150
6-BP4	0.7552	0.0148	0.7572	0.0169	0.7361	0.0277
7-BP10	0.7270	0.0161	0.7284	0.0165	0.,6924	0.0347
8-BP20	0.6846	0.0214	0.6471	0.0284	0.6735	0.0418

CSI: classification success index and σ: standard deviation of CSI.

values of the latter three 8-h CO concentration values are considered. Table 6 shows the results for the models that use the former exogenous variables as well as the cyclical daily and weekly behaviour of the CO concentrations. Better results are obtained in this case for $n = 1$, that is, considering the preceding 8-h CO concentration value of the prediction (peak/no peak occurrence). Results of the ANOVA test applied to the eight models described in Table 6 ($n = 1$), show that mean values are significantly different ($F = 55.63$). Then, the Bonferroni criteria were used (Table 7).

It can be inferred that the best system of regression is obtained for models 4, 6 and 5, which have the higher correlation coefficients. As the Bonferroni criterion sets, once the models are not significantly different, the simplest one must be selected. Then, model 4 has been selected, which is a neural network with one hidden unit. Models 2 and 3 in Table 2, which are not significantly different from model 4, have a higher computation time.

Table 6
Results of the different models with exogenous information and cyclical behaviour of the CO concentrations (prediction horizon: 8 h)

Model	$n = 1$		$n = 3$		$n = 6$	
	CSI	σ	CSI	σ	CSI	σ
1-knn1	0.7591	0.0246	0.7500	0.0174	0.7325	0.0174
2-knn3	0.7800	0.0215	0.7711	0.0115	0.7514	0.0119
3-knn5	0.7774	0.0226	0.7855	0.0106	0.7672	0.0153
4-BP1	0.7917	0.0139	0.7791	0.0156	0.7820	0.0219
5-BP2	0.7879	0.0213	0.7604	0.0160	0.7717	0.0271
6-BP4	0.7886	0.0138	0.7706	0.0219	0.7596	0.0287
7-BP10	0.7733	0.0201	0.7428	0.0201	0.7355	0.0338
8-BP20	0.7169	0.0160	0.6698	0.0406	0.7048	0.0364

CSI: classification success index and σ: standard deviation of CSI.

Table 7
Results of the Bonferroni criteria

Model	Mean	No significantly different
4-BP1	0.7917051	7 3 2 5 6 4
6-BP4	0.7885945	7 3 2 5 6 4
5-BP2	0.7879032	7 3 2 5 6 4
2-knn3	0.7799539	1 7 3 2 5 6 4
3-knn5	0.7774194	1 7 3 2 5 6 4
7-BP10	0.7732719	1 7 3 2 5 6 4
1-knn1	0.7591014	1 7 3 2
8-BP20	0.7169355	8

Table 8
Relative importance of the input variables to the ANN model

Variable	Relative importance
CO($t - 1$)	34.40
T($t - 1$)	30.44
V($t - 1$)	11.00
NO($t - 1$)	10.17
h($t - 1$)	6.51
D($t - 1$)	5.12
d($t - 1$)	2.35

Once the best neural model is selected, that is, the one that gets the best generalization for test patterns, an analysis of the relevance of the input variables in terms of the neural weights has been developed. The results are shown in Table 8. As it can be seen, the variables with a higher relative contribution are CO (34.40%) and T (30.44%) for lag $n = 1$. The reason is that the highest CO concentration values are likely to occur during the coldest periods of the year, when traffic emissions are usually higher and thermal inversions are stronger. The relative contribution of wind speed (11.00%) seems to confirm the dispersive effect of mechanically induced turbulence on CO levels. Furthermore, the effect of NO concentration series (10.17%) over the prediction of CO peaks can be explained considering that they are both primary gaseous pollutants being urban traffic their main emission source. Wind direction shows a significantly smaller contribution (5.12%), though it should not be removed from the models as it seems to show the slight directional dependence on CO emission sources. Cyclic variables (h and d) have also been included to capture the daily and weekly variation of CO concentration induced by traffic volume fluctuations (Maffeis, 1999). Besides, better results are obtained when these two variables are considered, as shown in Tables 4–6.

6. Conclusions

This study describes a suitable classification methodology for the prediction of maximum CO atmospheric

concentration levels. It is shown how the models can be improved by including the exogenous information of different variables and the cyclical behaviour of the CO concentrations. Errors are about 20% for the best models based on ANNs. The random procedure applied in the study presented here leads to safer results, minimizing the uncertainty associated to the natural data randomness. When these strategies are adequately applied to the obtained error rates in a well designed experiment, and the needed assumptions are verified, it is possible to determine the optimal complexity, o even, to determine which model fits better the samples. These results show how statistical methods can be successfully employed for the topology determination of neural networks architectures. Finally, the methodology presented here, together with a suitable selection of the informative variables, can also be applied to the prediction of the maximum levels of other atmospheric pollutants.

Acknowledgements

We would like to thank Consejería de Medio Ambiente from the Andalusian Regional Government which provided us with all the monitoring data.

References

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University, Oxford.

Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., Vitabile, S., 2007. Two-days ahead prediction of daily maximum concentrations of SO_2, O_3, PM10, NO_2, CO in the urban area of Palermo, Italy. Atmos. Environ. 41, 2967–2995.

Chelani, A.B., Chalapati Rao, C.V., Phadke, K.M., Hasan, M.Z., 2002. Prediction of sulphur dioxide concentration using artificial neural networks. Environ. Modell. Softw. 17, 161–168.

Comrie, A., Diem, J.E., 1999. Climatology and forecast modeling of ambient carbon monoxide in Phoenix, Arizona. Atmos. Environ. 33, 5023–5036.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, 2nd ed. Wiley.

Dutot, A.-L., Rynkiewicz, J., Steiner, F.E., Rude, J., 2007. A 24-h forecast ozone peaks and exceedance levels using neural classifiers and weather prediction. Environ. Modell. Softw. 22, 1261–1269.

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. Adv. Environ. Res. 5, 47–59.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences. Atmos. Environ. 32, 2627–2636.

Garson, G.D., 1991. Interpreting neural connection weights. Artif. Intell. Exp. 6, 47–51.

González Gallero, F.J., 2003. Characterization and source apportionmet of TSP Atmospheric concentration in the Campo de Gibraltar region, Spain. PhD University of Cadiz, Spain.

Jobson, J.D., 1991. In: Applied Multivariate Data Analysis Springer Texts in Statistics, vol. 1. Springer-Verlag, New York.

Jorquera, H., Pérez, R., Aldo, C., Espejo, A., Letelier, M.V., Acuña, G., 1998. Forecasting oxone daily maximum levels at Santiago, Chile. Atmos. Environ. 32 (20), 3415–3424.

Kolehmainen, M., Martikainen, H., Ruuskanen, J., 2001. Neural networks and periodic components used in air quality forecasting. Atmos. Environ. 35, 815–825.

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Gavin, C., 2003. Extensive evaluation of neural networks models for the prediction of NO_2 and PM10 concentrations, compared with a deterministic modelling system and measurement in central Helsinki. Atmos. Environ. 37, 4539–4550.

Maffeis, G., 1999. Prediction of carbon monoxide acute air pollution episodes. Model formulation and first application in Lombardy. Atmos. Environ. 33, 3859–3872.

Pelliccioni, A., Poli, U., 2000. Use of neural net models to forecast atmospheric pollution. Environ. Monit. Assess. 65, 297–304.

Perez, P., 2001. Prediction of sulphur dioxide concentrations at a site near downtown Santiago, Chile. Atmos. Environ. 35, 4929–4935.

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

Pizarro, J., Guerrero, E., Galindo, P., 2002. Multiple comparison procedures applied to model selection. Neurocomputing 48, 155–173.

Scheffe, H., 1959. Analysis of Variance. Wiley, New York.

Viotti, P., Liuti, G., Di Genova, P., 2002. Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia. Ecol. Model. 148, 27–46.