DTD 5
ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Ecological Modelling xxx (2004) xxx–xxx

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

# Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling

O. Pastor-Bárcenas[a], E. Soria-Olivas[b,*], J.D. Martín-Guerrero[b], G. Camps-Valls[b], J.L. Carrasco-Rodríguez[a], S. del Valle-Tascón[a]

[a] *Dpto. Biología Vegetal, Universitat de València, c/ Dr. Moliner 50, 46100 Burjassot, València, Spain*
[b] *G.P.D.S, Digital Signal Processing Group, Electronic Engineering Department, E.T.S.E, Universitat de València, c/ Dr. Moliner 50, 46100 Burjassot, València, Spain*

Received 13 June 2003; received in revised form 8 June 2004; accepted 4 July 2004

**Abstract**

This paper presents the use of artificial neural networks (ANNs) for surface ozone modelling. Due to the usual non-linear nature of problems in ecology, the use of ANNs has proven to be a common practice in this field. Nevertheless, few efforts have been made to acquire knowledge about the problems by analysing the useful, but often complex, input–output mapping performed by these models. In fact, researchers are not only interested in accurate methods but also in understandable models. In the present paper, we propose a methodology to extract the governing rules of trained ANN which, in turn, yields simplified models by using unbiased sensitivity and pruning techniques. Our proposal has been evaluated in thousands of trained ANNs under different conditions to establish a relationship between present contaminants (or several atmospheric variables) and surface ozone concentrations. The technique presented has demonstrated to be unbiased and stable with regard to the interpretability of the models and the good results obtained.


*Keywords:* Neural networks; Sensitivity analysis; Non-linear models; Tropospheric ozone; Forecasting

## 1. Introduction

Artificial neural networks (ANNs) have proven to be as efficient classification and regression tools, which are presently used in many modelling tasks (Haykin, 1999). Their appeal consists basically in their ability to perform complex and flexible input–output mappings. However, these excellent properties can turn into dramatic problems when one tries to understand the established relationships between the input and the output variables. In fact, researchers from many areas are even more interested in the interpretability issue than in the accuracy of the learning machine. This is particularly interesting in biomedical engineering, bioinformatics, or ecology, where the relative relevance (ranking) of the input variables in the problem is sometimes more

* Corresponding author. Tel.: +34 963160459; fax: +34 963160466.
*E-mail address:* emilio.soria@uv.es (E. Soria-Olivas).

0304-3800/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2004.07.015

important than the accuracy obtained by the classifier when discriminating a cardiac disease, when predicting the efficacy of an antisense oligonucleotide, or when modelling the surface ozone concentrations, respectively.

The multilayer perceptron (MLP) is the most commonly used ANN and many methods have been proposed to analyse the information encoded in the model parameters. Basically, two approaches are found in the literature: *sensitivity measures* (Pal, 1999; Orr and Müller, 1998) and *pruning methods* (LeCun et al., 1990; Guo and Uhrig, 1992; Hassibi and Stork, 1993; Hsu et al., 1995; Leray and Gallinari, 1999; Tresp et al., 1997; Zurada et al., 1994). In the field of environmental modelling, a recent paper (Olden and Jackson, 2002) has developed a randomisation approach and benchmarked three methods for network analysis. An extensive comparison of methods to study the contribution of variables in ANNs was recently published (Gevrey et al., 2003). These are certainly effective approaches to the problem. However, these sensitivity measurements should be calculated in many situations, i.e. for many neural networks with different topologies and training conditions. In fact, if we use these measures with only one network, we might arrive at *biased* solutions or what is even worse, we might even draw unreasonable conclusions. This issue is intimately related to the fact that neural networks are very flexible and thus they perform mappings that are too complex, which are difficult to encode in a set of understandable and easy rules.

In this work, we present the use of ANNs for surface ozone modelling. In addition, we propose a methodology for knowledge discovery in trained ANNs, which, in turn, yields simplified models by using *unbiased* sensitivity and pruning techniques. Our proposal has been tested in thousands of models that are trained to establish a relationship between both present contaminants and several atmospheric variables and surface ozone concentrations. Both the interpretability of neural models and the good results obtained validate the techniques presented and circumvent the bias problem of many previously proposed methods.

Many studies have shown the adverse effect of tropospheric ozone on health (Larsen et al., 1991; Spekton et al., 1991; Kinney et al., 1996). Moreover, high levels of ozone affect crops and forest vegetation (Krupa et al., 1994; Legge et al., 1995; Hogsett et al., 1997). Damage to agriculture results in the expenditure of more than one billion euros per year (Krupa et al., 1993; Slaughter et al., 1993). Tropospheric $O_3$ is formed from complex and non-linear photochemical reactions, and thus, it is a secondary contaminant. Precursor contaminants (Stockwell et al., 1997; Chameides and Lodge, 1992), which are emitted to the atmosphere (mainly NO, $NO_2$, and volatile organic composites), interact photochemically in the atmosphere (Alsthuller, 1978; Colbeck, 1985). At the same time, ozone is destroyed by oxidation. In laboratory conditions, $O_3$ concentrations are absolutely predictable from the precursors but this becomes a more difficult problem in field conditions, when ozone concentration (emissions) depend upon a great many uncontrolled and unpredictable factors (Comrie, 1997; Spellman, 1998).

In order to reduce the negative effects on public health and the economy, governments in USA and EU have established maximum $O_3$ levels that should not be exceeded (EPA, 1996; EU, 2002). Consequently, it might be very useful to develop mathematical models capable of predicting $O_3$ levels as a function of precursors. This would certainly avoid high emission levels and would aid in developing control strategies of ozone levels. Unfortunately, no general or mechanicistic models are available to achieve this goal. Besides, deterministic models would also require knowing the atmospheric stability, the diffusion coefficients in lateral and vertical direction, the reaction mechanisms and the kinetic data, and these are usually not well understood.

Linear models are not well suited to this problem since they yield relatively poor predictions due to the complexity induced by the presence of a great variety of underlying processes (Hadjiiski et al., 1999; McCollister and Wilson, 1975; Roberson and Steyn, 1990). This situation encourages the use of non-linear regression models such as neural networks. Neural models do not assume any distribution or relationship between variables, nor do they need a priori knowledge of the problem, and they have proven to be effective techniques in a wide range of applications. The use of ANNs in ecological modelling is well known and their superiority with respect to linear models has been extensively demonstrated in this field (Simpson et al., 1992; Lek et al., 1996; Paruelo and Tomasel, 1997; Wise et al., 1995). Recently, state-of-the-art neural-based prediction models of $O_3$ levels have been

developed (Nunnari et al., 1998; Balaguer et al., 2001). Nevertheless, as explained before, the main limitation of these approaches is that little attention has been paid to the structure of the model, and thus, it is not possible to ascertain the influence of the input descriptors on the formation of the tropospheric ozone. This paper presents a strategy to extract knowledge from the model which, in turn, leads to pruned models with a lower number of inputs and, consequently, better generalisation capabilities. These are certainly desirable characteristics that make the approach different from the previously adapted ones.

The rest of the paper is organized as follows. Section 2 describes Material and methods. Section 3 presents and discusses the results. Finally, Section 4 presents some concluding remarks and a proposal for further work.

## 2. Material and methods

### 2.1. Data description

The data used in this research was obtained from the Conselleria d' Agricultura, Peixca i Alimentació (Department of Agriculture) of the Generalitat Valenciana (Regional Government), València, Spain. The measuring site was located in the "Centre de Capacitació Agraria de Carcaixent" (València, Spain; 39°7′N, 0°27′W; 21 m above sea level; thermo-Mediterranean bio-climate). The sampling site is adjacent to the small town of Carcaixent, which is 40 km south of the city of València.

The input variables to the networks were NO and $NO_2$ levels as well as meteorological parameters such as wind speed (WS), temperature (T), atmospheric pressure (P), solar irradiance (SI) and relative humidity (RH). The data set for developing the models was obtained during 20 consecutive days (1 April–20 April 2002) from the rural monitoring station. Models were applied in order to forecast $O_3$ levels, exclusively in the ozone season (the periods with high $O_3$ values), which are spring and summer. In these periods, typical daily profiles present a maximum $O_3$ level at noon and a minimum $O_3$ level at night.

Once the input variables were defined, time series were processed to give zero mean and unit variance. The next step in training a neural model was to select the number of hidden nodes and layers, the learning rate, the number of epochs, and the initialisation range of synaptic weights.

### 2.2. Neural networks

Neural networks are extensively used in non-linear modelsm which have been widely applied to pattern recognition, system identification, and time series prediction. The MLP is the most commonly used neural network for pollution prediction (Gardner, 1999; Kolehmainen et al., 2001). MLP is fully described in (Haykin, 1999), and therefore, only a brief outline is provided here. The traditional model of a multilayer feed forward neural network is shown in Fig. 1.

An MLP is composed of a layered arrangement of artificial neurons in which each neuron of a given layer feeds all neurons the of the next layer. A single neuron extracted from the $l$-th layer of an $L$-layer network is also represented in Fig. 1. The inputs $x_i^l$ to the neuron are multiplied by adjustable coefficients $w_{ij}^l$ called *weights*, which represent the synaptic

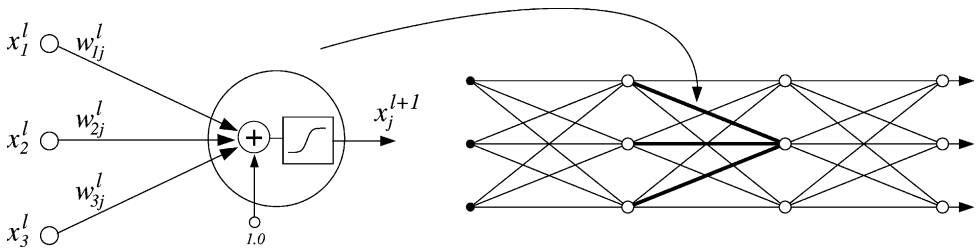


Fig. 1. Neuron model and feedforward network: each neuron passes the weighted sum of its inputs through a sigmoid function. The output of a neuron in a given layer acts as an input to neurons in the next layer. In the network illustration, each line represents a synaptic connection. Figure adopted from Wan, (1993).

connectivity between neuron $i$ in $l$-1 layer and neuron $j$ in layer $l$. The output of a neuron, $x_j^{l+1}$, is usually taken to be a sigmoid-shaped function of the weighted sum of its inputs:

$$x_j^{l+1} = f\left(\sum_i w_{ij}^l x_i^l\right) \quad (1)$$

A *bias* input to the neuron is achieved by fixing $x_0^l$ to +1. The network structure is completely defined by taking $x_l^0$ to be the external inputs, and $x_i^L$ to be the final outputs of the network. The MLP used in this work incorporates the hyperbolic tangent as a non-linear activation function $f$ of hidden neurons.

The back propagation algorithm was used to train the networks. There are two possible ways to apply this algorithm: on the one hand, the synaptic weights can be updated after each training pattern is shown to the network (on-line learning), and on the other hand, the update can be carried out after all the patterns are showed (batch learning).

In addition, a critical issue when training a neural model is the risk of falling in local minima, which produces poor results in the validation (generalization) set. In order to alleviate this problem, we also varied the initialisation range of the weights and incorporated the momentum term in the weight update.

### 2.3. Sensitivity analysis

Knowledge Discovery is defined as the process of identifying valid, novel, potentially useful, and ultimately understandable structure in data (Bradley et al., 1998). The scientific community is not only searching for methods that provide accurate estimations of the underlying system function, but also for methods that explain those complex, and often non-linear, relationships from the input–output mapping performed by the models. In this paper, sensitivity analysis for the MLP has been used in order to acquire knowledge about the problem.

Sensitivity analysis is used to study the influence of input variables on the dependent variable and consists of evaluating the changes in training error that would be obtained if an input were removed from the model. This measure, commonly known as *delta error* in the literature, produces a valuable ranking of the relevance of the variables (Pal, 1999). An additional sensitivity measure, called average absolute gradient (AAG) has been computed. This measurement is based on perturbing an input, and then monitoring model outputs (Orr and Müller, 1998). These sensitivity measures have been selected for their simplicity, their induced computational cost, and the good results achieved in other applications (Orr and Müller, 1998). The authors' experience in other applications (Martín et al., 2003) has demonstrated that the selected methods are especially well suited to this case study.

### 2.4. Methodology

In previous work (Balaguer et al., 2001; Balaguer et al., 2002), effective predictive models were developed to provide 24-h in-advance forecasts of the hourly ozone concentration. In the present paper, a similar methodology has been followed which is based on the data collected in the year 2002. The collected variables described in Section 2.1 were monitored continuously. The data was averaged over 60-min intervals. The training set consisted of data collected over 20 consecutive days (480 samples per variable) and the following 10 days constituted the validation set.

We used the cross-validation method to select the best model. The criterion used to select a candidate model for the final system was based on the model predictive performance in the validation data set. This performance was evaluated using several measures for the estimation of the ozone concentration.

Once the model is trained, two strategies were followed to determine the relative importance of the variables in the model:

1. *Sensitivity analysis.* We perform a sensitivity analysis of the model and a *re-training* of the network without taking into account the less significant predictors. Three direct consequences are obtained: (i) a valuable ranking of input variables is obtained, (ii) simplicity of the method is ensured, and (iii) unbiased rankings are provided.
2. *Pruning*. At this point, we can discard the less relevant predictors heuristically or by any fixed threshold, as in our case study. This allows us, in turn, to trace possible strategies to control $O_3$ emissions by acting only on the most relevant variables.

Table 1
Architecture, learning algorithm and results ($r$, $R^2$, MAE; RMSE, $d_2$) achieved by the best neural networks in the validation set

| Mode | Layers | Hidden neurons | $r$ | $R^2$ | MAE (ppb) | RMSE (ppb) | $d_2$ |
|---|---|---|---|---|---|---|---|
| Batch | 1 | 5 | 0.88 | 0.77 | 7.36 | 9.32 | 0.93 |
| Batch | 1 | 5 | 0.88 | 0.78 | 7.19 | 9.22 | 0.93 |
| On-line | 1 | 11 | 0.86 | 0.75 | 7.71 | 10.06 | 0.93 |
| On-line | 2 | 5 | 0.85 | 0.73 | 8.33 | 11.01 | 0.92 |
| On-line | 1 | 5 | 0.85 | 0.73 | 8.20 | 10.40 | 0.92 |
| Batch | 1 | 6 | 0.89 | 0.79 | 7.03 | 8.86 | 0.90 |
| Batch | 1 | 5 | 0.89 | 0.80 | 6.93 | 8.92 | 0.89 |
| Batch | 1 | 12 | 0.89 | 0.80 | 7.03 | 8.85 | 0.89 |
| Batch | 1 | 5 | 0.89 | 0.79 | 7.08 | 9.05 | 0.88 |
| Batch | 2 | 14 | 0.76 | 0.59 | 12.48 | 14.64 | 0.69 |

The row of the best model is grey-shaded.

## 3. Results and discussion

### 3.1. Performance indexes

Neural networks were developed using the techniques explained in Section 2. We varied the network parameters and found more than 1000 models for each kind of network (batch, on-line, and one or two hidden layers). In order to evaluate network performance, the following indexes were taken into account: Pearson's coefficient ($r$), determination coefficient ($R^2$), mean absolute error (MAE), root mean square error (RMSE) and index of agreement ($d_2$) (Gardner and Dorling, 1999). RMSE and MAE are error measures, which indicate level of accuracy; $r$, $R^2$ and $d_2$ are measures that indicate the similarity between the model tendency and the observed one. The importance of MAE and RMSE is relative, because their value depends on the original data. In Table 1, MAE presented values between 7.03 and 12.48 ppb,[1] and RMSE presented values between 8.85 and 14.64 ppb. Since the highest values of observed ozone were between 50 and 70 ppb, the values achieved in the error measures indicated that the networks fitted the observed values correctly. Although MAE is more intuitive, its strong sensitivity to extreme values (Wilmott, 1981) makes RMSE the most commonly used index to compare results and to develop a posterior statistical analysis.

Table 1 shows network architectures and performance indexes for the best results in the validation set. These networks were used for input selection, thus decreasing the complexity of the models. The number of neurons was varied between 5 and 16, in order to avoid the overfitting problem. The results with only one hidden layer were more efficient than the results with two hidden layers, since more neurons were necessary to obtain similar results with two layers and the risk of overfitting was evident.

It is worth noting that the indexes that appear in Table 1 are global for the generalization set. Nevertheless, the goal results were also obtained for individual prediction days. An example of individual predictions is given in Fig. 2.

Fig. 3 shows results obtained using one hidden layer with 12 units in the validation set formed by patterns, which were not previously presented to the network. This network offered the best performance indexes, except for the $d_2$ measurement and the MAE. The $r$ and $R^2$ values were 0.89 and 0.80, respectively, which are better than those obtained in previous works using neural networks for tropospheric ozone prediction (Spellman, 1999; Soja and Soja, 1999). It is important to highlight that the correlation values were high, which indicated fitted estimates.

Fig. 4 shows the first 4 days of the validation set; it magnifies part of the validation set to better show the excellent fitting. Good accuracy in the prediction is observed, with a trend and performance indexes that are similar to those obtained using a 10-day period. Statistical parameters are better since time prediction is shorter.

The results obtained indicate good generalization capabilities. Identical conclusions could be drawn from the other networks, as shown in Table 1. In conclusion,

[1] 1 ppb = 1.963 μg m$^{-3}$ at 25 °C.

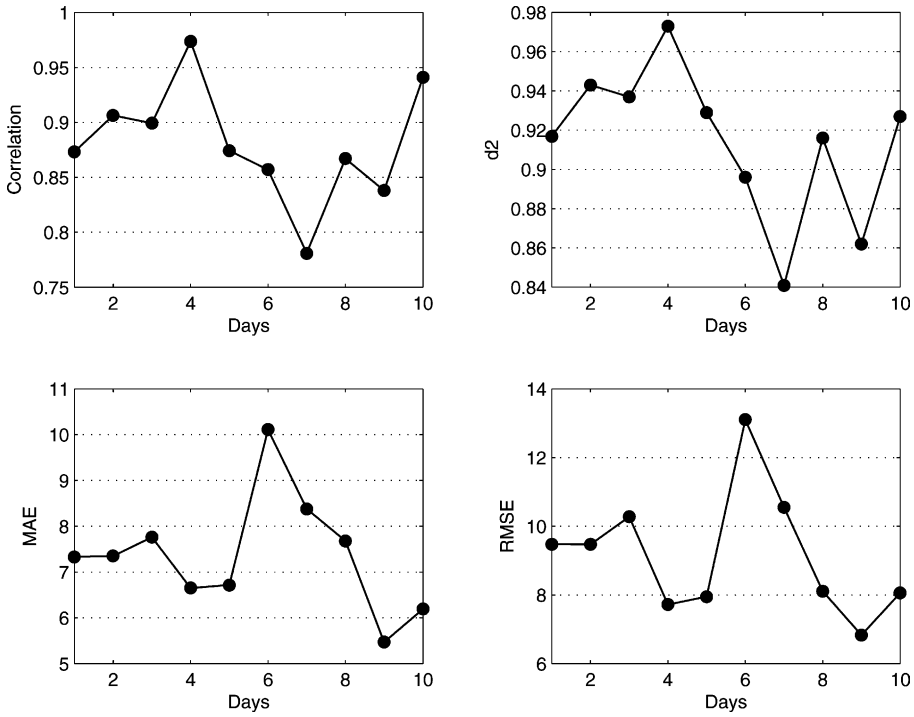


Fig. 2. Statistical indexes in the validation set using a neural network with five hidden nodes and batch back propagation algorithm.

the analysis of the results indicates that the best ANN models provided accurate predictions on the test data, indicating that model can extrapolate to unseen data.

### 3.2. Sensitivity analysis

As the achieved results were accurate, several sensitivity analyses were performed. Sensitivity analysis is based on studying the influence of input variables on the dependent variable, and consists of evaluating the changes in the error committed by the network that would result if an input were removed. In order to avoid biased results, 10 different models (those shown in Table 1) were considered (Ruck et al., 1990).

Table 2 shows the relative importance of inputs; the column labelled as one refers to the most important input and the column labelled as seven refers to the least important one.

It is important to highlight that the good results obtained by the models in the validation set ensure good generalization capabilities. This also implies that the networks have captured the underlying system that generated the data and therefore extracting rules in these cases is reliable. The most important variables for $O_3$ concentration are temperature and solar irradiance. This fact is extensively referenced in the literature

Table 2
Inputs ordered according to their relative importance for different neural models

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| T | SI | WS | RH | P | $NO_2$ | NO |
| T | SI | WS | RH | P | $NO_2$ | NO |
| T | SI | WS | RH | P | $NO_2$ | NO |
| T | SI | RH | $NO_2$ | WS | NO | P |
| T | SI | RH | WS | P | $NO_2$ | NO |
| T | SI | WS | RH | P | $NO_2$ | NO |
| T | SI | RH | $NO_2$ | WS | NO | P |
| T | SI | RH | WS | P | $NO_2$ | NO |
| T | SI | WS | RH | P | $NO_2$ | NO |
| T | SI | RH | WS | NO | P | $NO_2$ |

T, temperature; SI, solar irradiance; WS, wind speed; RH, relative humidity; and P, atmospherical pressure.

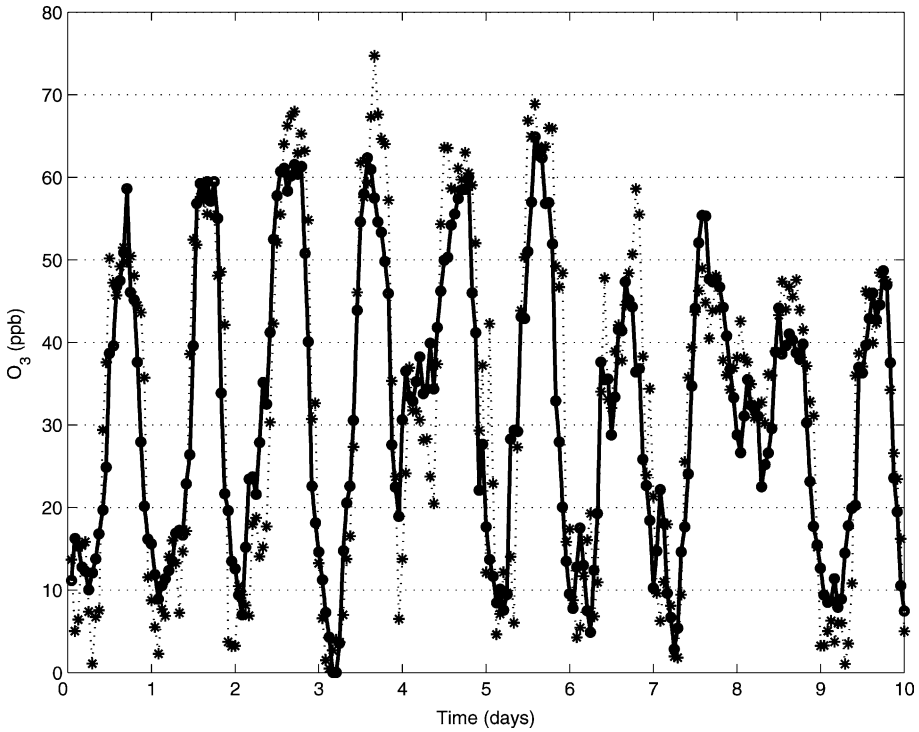


Fig. 3. Tropospheric ozone prediction in the validation set (10 days) carried out by a neural network with one hidden layer formed by 12 units (solid), and desired values (dashed). The statistical parameters of fit were $d_2$ = 0.987, $r$ = 0.899, $R^2$ = 0.808, MAE = 7.039 ppb, RMSE = 8.854 ppb.

(Chameides and Lodge, 1992). Moreover, most of the models consider NO and $NO_2$ levels to be the least important inputs.

In fact, the sensitivity analysis shows that precursors (NO, $NO_2$) are not very important in ozone concentration, a conclusion which is confirmed by their concentration values, which are very low. It should be taken into account that this fact corresponds to the model development in the period of time and area studied. These precursors do not present the highest relative importance; this suggests that a transport phenomenon is the main cause of the $O_3$ concentration. This is due to the fact that nitrogen oxide is transported a short distance, but $O_3$ can be carried a longer distance (Chameides and Lodge, 1992). In fact, both nitrogen oxides and $O_3$ were likely transported by wind to our study area.

Since temperature (T) and solar irradiance (SI) are the most important variables, high level of $O_3$ is formed in the areas surrounding the monitoring station. Furthermore, primary pollution agents (nitrogen oxides) are also transported to the area, but in a lower quantity because T and SI provoke the ozone formation.

### 3.3. Pruning

Pruning consists of removing those variables whose influence is less important in order to predict $O_3$ concentration, using sensitivity analysis (Pal, 1999; Ruck et al., 1990; Sarle, 2000). This procedure allows us to determine which models are more accurate and what the optimal architecture is. In order to carry out the pruning, we removed the $NO_2$ and the NO, and after that, we predicted the next $O_3$ concentration by taking into account only the four environmental variables mentioned. Table 3 shows the best results obtained and the corresponding architecture of the neural model. The goodness of the results justifies the input removal. The best networks are formed when there is one hidden layer with a high number of neurons; the results were

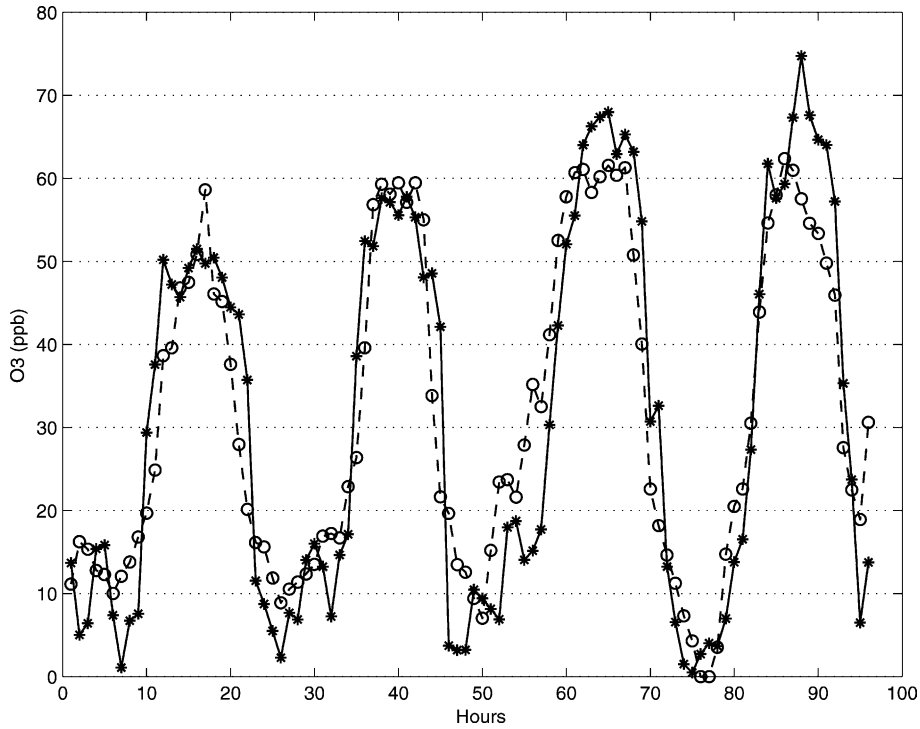


Fig. 4. Tropospheric ozone prediction in four days of the validation set carried out by a neural network with one hidden layer formed by 12 units (solid), and desired values (dashed). Statistical parameters of fit were $d_2$ = 0.987, $r$ = 0.899, $R^2$ =0.808, MAE = 7.039 ppb, RMSE = 8.854 ppb.

Table 3
Indexes for the pruned networks presenting the best results

| Mode | Layers | Hidden neurons | $r$ | $R^2$ | MAE (ppb) | RMSE (ppb) | $d_2$ |
|---|---|---|---|---|---|---|---|
| On-line | 1 | 6 | 0.86 | 0.76 | 8.07 | 10.18 | 0.94 |
| Batch | 1 | 6 | 0.88 | 0.78 | 7.73 | 9.54 | 0.93 |
| Batch | 2 | 12 | 0.76 | 0.59 | 17.03 | 21.03 | 0.57 |

not as good when networks with two hidden layers were used.

## 4. Conclusions

We have developed a methodology to build neural networks that accurately predict the current $O_3$ concentration by using information from environmental air pollutants and atmospheric data. A good model of $O_3$ concentration has been developed in terms of a high correlation coefficient between actual and predicted signals, and low RMSE and MAE values. Moreover, the robustness of the models can be guaranteed, since forecasting is accurate throughout the entire prediction period.

Sensitivity analysis has shown that the most relevant factors for the $O_3$ concentration are temperature and solar irradiation, while the least important variables are $O_3$ precursors (NO and $NO_2$). This fact could indicate that the $O_3$ was not generated at the monitoring area, but that it was transported by wind from other places.

The methodology could be interesting for studying the influence of precursors and the formation of tropospheric ozone in other sites. Our method only requires a few input variables, so it could be considered as an alternative to the physical and chemical models, which need a huge amount of data. Moreover, the technique

employed is an easy way to determine the relative importance of factors in $O_3$ concentration.

The main drawback of our neural model is related to the number of patterns used, which is very low due to the limitations encountered in the measuring site. If more data or more variables were available, the model precision would have been greater. Future work should consider *bootstrap* techniques for model analysis and testing.

## Acknowledgment

The authors gratefully acknowledge the financial support provided by the Conselleria de Agricultura, Peixca i Alimentació (Regional Government, València, Spain; Project GV-CAPA 00-06).

## References

Alsthuller, A.P., 1978. Association of oxidant episodes with warm stagnating anticyclones. J. Air Pollut. Control Assoc. 28, 152–155.

Balaguer, E., Soria, E., Carrasco, J.L., del Valle, S., 2001. Prediction if surface ozone concentrations 24 h in advance using neural networks. In: Advanced in Neural Networks and Applications. World Scientific and Engineering Press, EEUU, pp. 212–214.

Balaguer, E., Camps, G., Carrasco, J.L., Soria, E., del Valle, S., 2002. Effective 1-day ahead prediction of hourly surface ozone concentrations in eastern Spain using linear models and neural networks. Ecol. Modell. 156, 27–41.

Bradley, P.S., Fayyad, U.M., Mangasarian, O.L., 1998. Mathematical programming for data mining: formulations and challenges. Informs J. Comput. 3 (11), 217–238.

Chameides, W.L., Lodge, J.P., 1992. Tropospheric ozone: formation and fate. In: Lefohn, A.S. (Ed.), Surface Level Ozone Exposures and their Effects on Vegetation, pp. 5–30.

Colbeck, I., 1985. Photochemical ozone pollution. Weather 40, 241–245.

Comrie, A.C., 1997. Comparing neural networks and regression models for ozone forecasting. J. Air Waste Manag. Assoc. 47, 653–663.

EU, 2002. European Parliament, and the European Council. 2002/3/CE O.D. C 67 E. 9-3-2002, p. 14–20.

Gardner, M.W., Dorling, S.R., 1999. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmospheric Environ. 34, 21–34.

Gevrey, M., Dimopoulos, L., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Modell. 160, 249–264.

Guo, Z., Uhrig, R.E., 1992. Using modular neural networks to monitor accident conditions in nuclear power plants, SPIE. In: Applications of Artificial Neural Networks III. 1709, pp. 505–516.

Hadjiiski, L., Geladi, P., Hopke, P., 1999. A comparision of modeling nonlinear systems with artificial neural networks and partial least squares. Chemom. Intell. Lab. Syst. 49, 91–103.

Hassibi, B., Stork, D.G., 1993. Second order derivatives for network pruning: optimal brain surgeon. In: Hanson, S.J., Cowan, J.D., Giles, C.L. (Eds.), Advances in Neural Information Processing Systems, vol. 5, pp. 164–171.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice Hall, New Jersey.

Hogsett, W.E., Weber, J.E., Tingey, D., Herstrom, A., Lee, E.H., Laurence, J.A., 1997. An approach for characterizing tropospheric ozone risk to forests. Environ. Manag. 21, 105–120.

Hsu, C.T., Tzeng, Y.M., Chen, C.S., Cho, M.Y., 1995. Distribution feeder loss analysis by using an artificial neural network. Electric Power Syst. Res. 34, 85–90.

Kinney, P.L., Thurston, G.D., Raizenne, M., 1996. The effects of ambient ozone on lung function in children: a reanalysis of six summer camp studies. Environ. Health Perspect. 104, 170–174.

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

Krupa, S.V., Manning, W.J., Nosal, M., 1993. Use of tobacco cultivars as biological indicators of ambient ozone pollution: An analysis of exposure-response ralationships. Environ. Pollut. 81, 137–146.

Krupa, S.V., Nosal, M., Legge, A.H., 1994. Ambient ozone and crop loss: establishing a cause-effect relationship. Environ. Pollut. 83, 269–276.

Larsen, R.I., McDonell, W.F., Hortsman, D.H., 1991. An air quality data analysis system for interrelating effects, standards, and needed source reductions. Part II: A log-normal model relating human lung function decrease to ozone exposure. J. Waste Manag. Assoc. 41 (4), 455–459.

LeCun, Y., Denker, J., Solla, S., Howard, R.E., Jackel, L.D., 1990. Optimal Brain Damage. In: Touretzky, D.S. (Ed.), Advances in Neural Information Processing Systems (II).

Legge, A.H., Grünhage, L., Nosal, M., Jägger, H.J., Krupa, S.V., 1995. Qambient ozone and adverse crop response: an evaluation of North American and European data as they relate to exposure indices and critical levels. J. Appl. Bot. 69, 192–205.

Lek, S., Delacoste, M., Baran, P., Dimopoulus, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling non-linear relationship in ecology. Ecol. Modell. 90 (1), 39–52.

Leray, P., Gallinari, P., 1999. Feature selection with neural networks. Behaviormetrika, p. 26.

Martín, J.D., Soria, E., Camps, G., Serrano, A.J., Pérez, J.J., Jiménez, N.V., 2003. Dosage individualization of erythropoietin using a profile-dependent support vector regression. IEEE Trans. Biomed. Eng. 50, 1136–1142.

McCollister, G.M., Wilson, K.R., 1975. Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. Atmospheric Environ. 9, 417–423.

Nunnari, G., Nucifiora, A.F.M., Radineri, C., 1998. The application of neural techniques to the modelling of time series of atmospheric pollution data. Ecol. Modell. 111, 187–205.

Olden, D.O., Jackson, D.A., 2002. Illuminating the "black-box": a randomisation approach for understanding variable contributions in artificial neural networks. Ecol. Modell. 154, 135–150.

Orr, G.B., Müller, K.R., 1998. Neural Networks: Tricks of the Trade. Springer-Verlag, Berlin, Heidenberg.

Pal, N.R., 1999. Soft computing for feature analysis. Fuzzy Sets Syst. 103, 201–221.

Paruelo, J.M., Tomasel, F., 1997. Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. Ecol. Modell. 98 (2–3), 173–186.

Roberson, S.M., Steyn, D.G., 1990. Evaluation and comparison of statistical forecast models for daily maximum concentrations. Atmospheric Environ. 24B, 303–312.

Ruck, D.W., Rogers, S.K., Kabrisky, M., 1990. Feature selection using a multilayer perceptron. J. Neural Network Comput. 2 (2), 40–48.

Sarle, W.S., 2000. How to measure importance of inputs? SAS Institute Inc., Cary, NC, USA. Available at ftp://ftp.sas.com/pub/neural/importance.html.

Simpson, R., Williams, R., Ellis, R., Culverhouse, P.F., 1992. Biological pattern recognition by neural networks. Mar. Ecol. Prog. 79, 303–308.

Slaughter, L.H., Mulchi, C.L., Lee, E.H., 1993. Wheat-kernel growth characteristics during exposure to chronic ozone pollution 80, 73–79.

Soja, G., Soja, A.M., 1999. Ozone indices based on simple meteorological parametres: potencial and limitations of regression and neural network models. Atmospheric Environ. 33, 4229–4307.

Spekton, D.M., Thurston, G.D., Mao, J., He, D., Hayes, C., Lippman, M., 1991. Effects of single and multiday ozone exposures on respiratory function in active normal children. Environ. Res. 55, 107–122.

Spellman, G., 1998. Analysing air pollution meteorology. Weather 53, 34–42.

Spellman, G., 1999. An application of artificial neural networks to the prediction of surface ozone concentrations in the United Kingdom. Appl. Geography 19 (2), 123–136.

Stockwell, W.R., kramm, G., Scheel, H.E., Mohnen, V.A., Seiler, W., 1997. Ozone formation, destruction and exposure in Europe and the United Estates. In: Sandermann, H., Wellburn, A.R., Heath, R.L. (Eds.), Forest Decline and Ozone, pp. 1–38.

Tresp, V., Neuneier, R., Zimmermann H.G., 1997. Early brain damage. In: Mozer, M.C., Jordan, M.I., Petsche, T (Eds.), Advances in Neural Information Processing Systems, vol. 9, p. 669.

U.S. Environmental Protection Agency, 1996. Air quality criteria for ozone and related photochemical oxidants. EPA 600/8-84/020cF. U.S. Environmental Protection Agency, North Caroline.

Wan, E., 1993. Time series prediction using a neural network with embedded tapped delay-lines. In: Weigend, A., Gershenfeld, N. (Eds.), Predicting the Future and Understanding the Past, SFI Studies in the Science of Complexity, Addison-Wesley.

Wilmott, C.J., 1981. On the validation of models. Phys. Geography 2, 184–194.

Wise, B.M., Holt, B.R., Gallagher, N.B., Lee, S., 1995. A comparison of neural networks, non-linear biased regression and a genetic algorithm for dynamic model identification. Chemom. Intell. Lab. Syst. 30 (1), 81–89.

Zurada, J.M., Malinowski, A., Cloete, I., 1994. Sensitivity análisis for minimization of input data dimension for feedforward neural network. In: ISCAS'94 IEEE International Symposium on Circuits and Systems, vol. 6, IEE Press, London, pp. 447–450.