Development of a computational software to forecast ozone levels

Elena Agirre, Albaro Anta and Luis Javier R. Barron
University of the Basque Country
Spain

1. Introduction

Nowadays, the pollution caused by photochemical oxidants is one of the major problems in air quality. Tropospheric ozone (O_3) is a photochemical oxidant, which is produced as a consequence of chemical reactions of nitrogen oxides (NOx) with volatile organic compounds (VOCs) under the influence of solar radiation. It is well known that high ozone concentrations can be harmful to human health and the environment (WHO, 2000), particularly when these concentrations are high and exceed the thresholds established by the European ozone directive (European Communities, 2002). This directive requires that national authorities to inform the public – hourly or daily – about any incidence of ozone pollution above 180 $\mu g/m^3$ (information threshold). Table 1 summarizes the ozone thresholds and target values for the protection of human health.

Thresholds and targets	Value ($\mu g/m^3$)	Reference period
Population alert threshold	240	1 h average
Population information threshold	180	1 h average
Human health protection target value	120	8 h average

Table 1. Human health protection ozone thresholds and target (Directive 2002/3/EC).

Therefore, the use of prognostic models designed with the purpose of forecasting those ozone levels would be a very helpful tool for national authorities to inform the population about such situations in which the ozone thresholds for the protection of human health could be exceeded.

This chapter presents a computational software named *airEsan*, which has been elaborated aimed to forecast daily maximum ozone levels at several stations of the Air Quality Monitoring Network of the Basque Country (North Central Spain). The mathematical prognostic model in use in this computational software was derived from the application of the multilayer perceptron as the base of the model.

Several types of prognostic models have been designed in order to forecast air pollutant levels. Eulerian models (Scheffe and Morris, 1993), models based in series analysis (Kuang-

Jung, 1992), multiple linear regression based models (Cardelino et al., 2001) and artificial neural network based models (Gardner et al., 1999) are some examples. The use of artificial neural networks has generally provided better results in forecasting ozone concentrations (Comrie, 1997; Cobourn, 2000; Gardner et al., 2000). Artificial neural networks have proved their efficiency in describing non-linear relationships such as those involved in ozone formation. In this way, the research team of air quality *AireKal* formed by E. Agirre, A. Anta and L.J.R. Barron has developed a line of research focussed on the study of prognostic models applied to air quality, based mainly on the use of artificial neural networks with the purpose of forecasting in real time hourly ozone levels up to eight hours ahead (Agirre et al., 2005) and maximum daily ozone levels (Agirre et al., 2006a) at several stations in the Air Quality Monitoring Network of the Basque Country. The initial technique used by this team was the multivariate linear regression, but having proven the superior efficiency of the artificial neural networks (Agirre, 2003), the creation of models based on the use of the multilayer perceptron was tackled in subsequent works with the aim of increasing the prediction ability of the ozone concentrations at new locations of the Air Quality Monitoring Network of the Basque Country. As a result, a mathematical prognostic model was determined to forecast the daily maximum ozone values.

Section 2 provides a description of the mathematical model, indicating the topology of the multilayer perceptron. This mathematical model has been elaborated and validated with the database formed by the hourly values of some air pollutants (ozone and nitrogen dioxide) and meteorological parameters (temperature, relative humidity, pressure, solar radiation, wind direction and wind speed) registered in the Air Quality Monitoring Network managed by the Environmental Department of the Basque Government. Previous research had already proved that the mentioned mathematical model was an effective prognostic model (Agirre et al., 2006b).

Once the mathematical model was validated, a computational software named *airEsan* was developed. The computational software *airEsan* was exclusively installed in the Environmental Department of the Basque Government and it is currently being updated in ongoing research.

2. The mathematical model in *airEsan*

The artificial neural networks are mathematical-computational structures that emulate the way that neurons work in the human brain. The artificial neural networks are formed by neurons set in layers and connected with each other. After the establishment of the concept of artificial neural network (McCulloch and Pitts, 1943), a great range of studies have used different types of artificial neural networks, the multilayer perceptron being the most commonly used in air quality (Gardner and Dorling, 1998). The multilayer perceptron is a well-known artificial neural network due to its ability to represent any smooth measurable functional relationship between the inputs and the outputs (Hornik et al., 1989). Hence, it was the type of artificial neural network selected to elaborate the mathematical model of the computational software *airEsan*.

The multilayer perceptron is formed by the input layer, one or more hidden layers and the output layer. Fig.1 shows a multilayer perceptron with four neurons in the input layer, three neurons in the unique hidden layer and one neuron in the output layer.

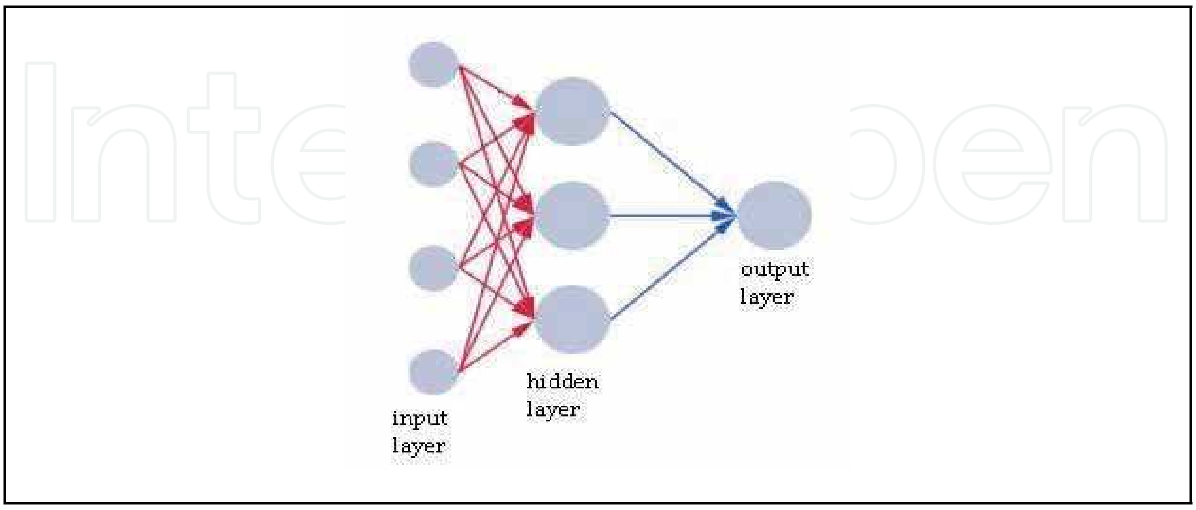

Fig. 1. A mulitlayer perceptron.

The input vector is introduced in the input layer. The value of each neuron in the input layer is multiplied by a synaptic weight and the addition of the resulting weighted values is calculated. A transference function is applied to this addition and the result is the output in the hidden layer. The same method is applied to produce the final output in the output layer.

Equation 1 represents an abbreviated form of the output of the multilayer perceptron:

$$\boldsymbol{y} = f^2(\boldsymbol{w}^2(f^1(\boldsymbol{w}^1\boldsymbol{x} + \boldsymbol{b}^1)) + \boldsymbol{b}^2) \qquad (1)$$

where $\boldsymbol{x}$ is the input vector, $\boldsymbol{y}$ is the output vector, f^1 and f^2 are the transference functions, $\boldsymbol{w}^1$ and $\boldsymbol{w}^2$ are the weight matrixes and $\boldsymbol{b}^1$ and $\boldsymbol{b}^2$ are the bias vectors.

The multilayer perceptron has the capability to learn from the patterns presented to it and from the errors it commits in the learning process. The learning process is similar to an optimization process, in which the error function E has to be minimized:

$$E = \sum_{k=1}^{L} (y_k - t_k)^2 \qquad (2)$$

where $(t_1, t_2, \ldots, t_L)^t$ is the target vector, $(y_1, y_2, \ldots, y_L)^t$ the output vector and L is the number of neurons in the output layer. The output of the multilayer perceptron is compared to the target, and if the corresponding error is not the minimum, the output is feed forwarded, generating an adjustment in the network weights so that the difference between the output of the network and the target output is minimum in the next iteration. This process is known as *backpropagation*. After a finite number of iterations, the learning process is completed when the global minimum of function E is reached. After the learning process, the multilayer perceptron has to be capable of identifying patterns that have never been presented to it before.

The multilayer perceptrons designed in the computational software *airEsan* consist of three layers: the input layer, the middle layer and the output layer. On the one hand, the input layer is made up of historic hourly values of the ozone, nitrogen dioxide, temperature, pressure, relative humidity, solar radiation, wind direction and wind speed. All these data are recorded in the Air Quality Monitoring Network managed by the Environmental Department of the Basque Government. The seasonality of ozone is taken into account by introducing as input variables the components $sin(2\,\pi\, d/N)$ and $cos(2\,\pi\, d/N)$, being d the day of the year (d = 1,2,…,N and N = 365 or 366). On the other hand, the output layer is made up of the forecast of the daily maximum ozone concentration. Finally, the number of neurons in the hidden layer is determined following the next rule in a trial and error procedure: "the number of training examples must be at least 30 times the number of parameters of the multilayer perceptron" (Amari et al., 1997).

The training algorithm utilized was an algorithm derived from backpropagation, known as the *Scaled Conjugate Gradient* algorithm (Moller, 1993). The hyperbolic tangent function was used as the transference function between the input layer and the hidden layer and the linear function was utilized to connect the hidden layer and the output layer. The early stopping technique was used to avoid overtraining (Sarle, 1995), separating the database into three subsets: the training set, the validation set and the test set. In the last update of the mathematical model contained in the computational software *airEsan*, data from the period 2001-2005 were used to train the model, the validation set was formed by data from 2006 and finally the model was tested with data from 2007.

The goodness of the fit of the mathematical model was measured in a quantitative way by the calculation of the values of the Model Validation Kit on the test set (Hanna et al., 1991). This kit is formed by the following statistics:

(i) the correlation coefficient

$$R = Mean\left[(C_o - Mean(C_o))(C_p - Mean(C_p))\right] / \left((SC_o)(SC_p)\right) \tag{3}$$

(ii) the Normalized Mean Square Error

$$NMSE = \frac{Mean((C_o - C_p)^2)}{Mean(C_o)Mean(C_p)} \tag{4}$$

(iii) the factor of two FA2 which shows in which proportion are the values of the forecasted/observation proportion in the interval [0.5, 2]

$$0.5 \leq \frac{C_p}{C_o} \leq 2 \tag{5}$$

(iv) the Fractional Bias

$$FB = 2\frac{Mean(C_o) - Mean(C_p)}{Mean(C_o) + Mean(C_p)} \tag{6}$$

(v) the Fractional Variance

$$FV = 2\frac{SC_o - SC_p}{SC_o + SC_p} \tag{7}$$

where C_p is the forecasted value, C_o is the observation, *Mean* indicates the mean value and S is the standard deviation. The best performance is obtained by the values $R = FA2 = 1$ and $NMSE = FB = FV = 0$.

The values of these five statistics guaranteed the accuracy of the mathematical-computational model in use in *airEsan* aimed to forecast the daily ozone maximum concentrations for one day ahead at the studied stations.

3. The computational software *airEsan*

3.1 Origin of the computational software *airEsan*

Our research group has been working for over six years with the Environmental Department of the Basque government, who expressed interest in acquiring a software application to predict ozone concentrations a day in advance. In this way, they would have an utility model to forecast the exceedances of ozone thresholds for the protection of human health. Therefore, the computational software *airEsan* was developed to predict the ozone maximum levels for the next day. The mathematical model included in *airEsan* was designed with the guidelines described in section 2 and it is annually updated taking into account the most recent records of the Air Quality Monitoring Network and the correspondant adjustments. The computational model *airEsan* is being utilized exclusively in the Environmental Department of the Basque Government.

3.2 Utilities of the computational software *airEsan*

Obviously the main usefulness of the computational software *airEsan* is the forecasting capability of daily maximum concentrations of ozone one day in advance. But as the option *Queries* in Fig. 2 indicates, the ozone prediction system *airEsan* is also capable of giving the maximum ozone level forecast for the following day, previous forecasts of the maximum ozone concentrations, the daily maximum ozone levels registered in the past, the exceedances of the ozone thresholds for the protection of human health and the comparison between the observations and the forecasts of the daily maximum ozone concentrations at different stations in the Air Quality Monitoring Network of the Basque Country.

Fig. 2. Main utilities of the computational software *airEsan*.

If the user's query is *Daily maximum O_3 forecast for the next day*, Fig. 3 will show the forecast of the maximum ozone concentration for the next day at the selected stations.

Fig. 3. Maximum O_3 forecast for the next day.

The computational software *airEsan* also gives the forecast of the air quality index with respect to the daily ozone maximum concentration. This year the colours and values for each index will be in accordance with table 2 (Basque Government, 2009).

Color	Description of air quality	Ozone (µg/m^3)
	Good	0-90
	Acceptable	90.1-160
	Moderate	160.1-180
	Bad	180.1-270
	Very bad	270.1-360
	Dangerous	>360

Table 2. Air quality index related to ozone in the Basque Country (2009).

Fig. 4 illustrates the good air quality with respect to ozone in the study period at Mundaka station. In the same way, the evolution of the observations and the forecasts of the daily maximum ozone levels in the time period selected at each station is depicted in Fig. 4.

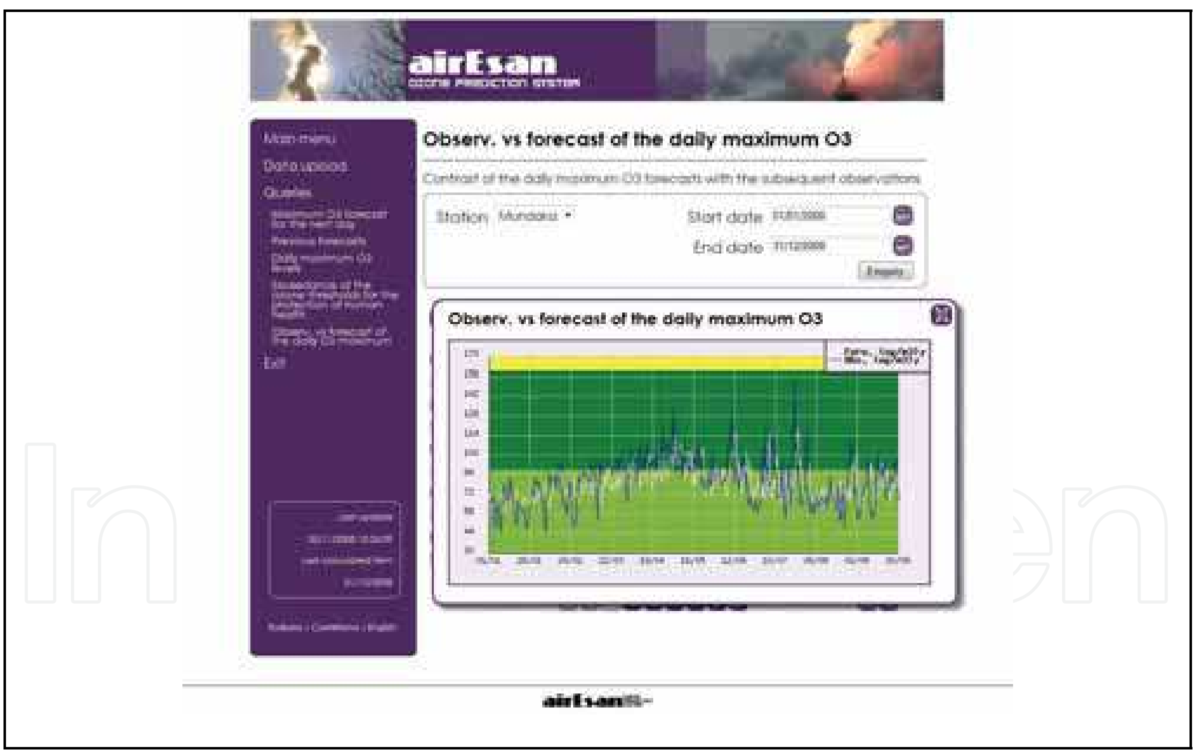

Fig. 4. Graphical display of the observed values vs. the forecast of the daily maximum O_3.

In order to study the trend of the past forecasts of daily maximum ozone, the *Previous forecasts of daily maximum ozone* option could be selected. Fig. 5 shows the window to select the station and the dates to obtain the corresponding past forecasts of the daily maximum ozone.

Fig. 5. Previous forecasts of daily maximum ozone.

Moreover, Fig. 6 shows the options of the *Exceedances of the ozone thresholds for the protection of human health* window.

Fig. 6. Exceedances of the ozone thresholds for the protection of human health.

This option provides the dates and the values of the daily maximum ozone concentrations that have exceed the population information threshold (180 $\mu g/m^3$) and the alert population threshold (240 $\mu g/m^3$) in a specified date at the selected stations.

4. Conclusions

Computational software used for forecasting purposes can be a very useful tool in decision making, planning and evaluation of air quality, especially if it provides the ability to inform or alert population about the possible exceedances of ozone thresholds for the protection of human health. This chapter presents the most interesting options of the computational software *airEsan*, which was designed in order to obtain the forecasts of daily maximum ozone concentrations at several stations in the Basque Country. The software contains an efficient mathematical model based on the use of the multilayer perceptron, which was designed to predict the daily maximum ozone concentrations one day in advance at several stations of the Air Quality Monitoring Network in the Basque Country. The mathematical model was developed using the Neural Network Toolbox of Matlab. At this moment, the computational software *airEsan* is exclusively being used within the Environmental Department of the Basque Government.

Finally, future research will study the elaboration and validation of the mathematical prognostic model to forecast the daily maximum ozone concentrations at new stations of the Air Quality Monitoring Network in the Basque Country, using a new database with the past values and the most recent values registered at the aforementioned Network. The mathematical results will be included in the computational software *airEsan,* which will also be updated. The computational software could be extended for the prediction of different air pollutants, but before that, the corresponding mathematical model should be built.

5. Acknowledgments

The authors gratefully acknowledge data provision and financial support from the Environmental Department of the Basque Government.

6. References

Agirre, E. (2003). Elaboration of a real-time statistical model for predicting short-term levels of O_3 and NO_2 in the city of Bilbao. *Ph. D*, University of the Basque Country, Bilbao.

Agirre, E.; Anta, A. & Barron, L.J.R. (2005). Development and evaluation of a prediction model based on the use of multilayer perceptron, in real time, to forecast ozone levels in Rioja Alavesa and Kostaldea. *Environmental Department of the Basque Government*, project number 175P/2005.

Agirre, E.; Anta, A. & Barron, L.J.R. (2006a). Forecasting of daily maximum ozone levels in Kostaldea and Northern Alava. Determination of the input variables of the prediction model. *Environmental Department of the Basque Government,* project number 102P/2006.

Agirre, E.; Ibarra-Berastegi, G. & Madariaga, I. (2006b). Regression and multilayer perceptron-based models to forecast hourly O_3 and NO_2 levels in the Bilbao area. *Environmental Modelling & Software,* 21*, 4,* 430-446.

Amari S-I.; Murata N.; Müller K.R. & Finke M., Yang H.H. (1997). Asymptotic statistical theory of overtraining and cross-validation. *IEEE Transactions on Neural Networks,* 8, 5, May 1997, p. 985-996.

Basque Government (2009). http://www.ejgv.euskadi.net/r53-2291/es/contenidos/informacion/red_calida_aire_capv/es_975/indice_calidad_c.html

Cardelino, C.; Chang, M. & John J.S. (2001). Ozone predictions in Atlanta, Georgia: analysis of the 1999 ozone season. *Journal Of the Air & Waste Management Association,* 51, 8, August 2001, p. 1227-1236.

Cobourn, W.G.; Dolcine, L.; French, M. & Hubbard, M.C. (2000). A comparison of nonlinear regression and neural network models for ground-level ozone forecasting. *Journal of the Air and Waste Management Association,* 50, 11, November 2000, p. 1999-2009.

Comrie, A.C. (1997). Comparing neural networks and regression models for ozone forecasting. *Journal of Air and Waste Management Association,* 47,6, June 1997, p. 653-663.

European Communities (2002). Directive 2002/3/EC of the European Parliament and of the Council of 12 February 2002 relating to ozone in ambient ai*r. Official Journal of the European Communities,* Luxembourg.

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, August 1998, p. 2627-2636.

Gardner, M.W. & Dorling, S.R. (1999). Neural network modelling and prediction of hourly NOx and NO_2 concentrations in urban air in London. *Atmospheric Environment,* 33, 5, February 1999, p. 709-719.

Gardner, M.W. & Dorling, S.R. (2000). Statistical surface ozone models: an improved methodology to account for non-linear behaviour. *Atmospheric Environment,* 34, 1, January 2000, p. 21-34.

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.

Hornik, K.; Stinchcombe, M. & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks,* 2, 5, 1989, p. 359-366.

Kuang-Jung, H. (1992). Time series analysis of the interdependences among air pollutants. *Atmospheric Environment,* 26B, 4, December 1992, p. 491-503.

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. (1991). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6, 4, 1993, p. 525-533.

Sarle, W.S. (1995). Stopped training and other remedies for overfitting. *Proceedings on the 27th Symposium on The Interface of Computing Science and Statistics*, pp. 352–360, Pittsburgh.

Scheffe, R.D. & Morris R.E. (1993). A review of the development and application of the Urban Airshed model. *Atmospheric Environment* 27B, Mar 1993, p. 23-39.

WHO (2000). Guidelines for Air Quality. *World Health Organization,* Geneva, 190 pp.

Technology Education and Development

Edited by Aleksandar Lazinica and Carlos Calafate

ISBN 978-953-307-007-0
Hard cover, 528 pages
Publisher InTech
Published online 01, October, 2009
Published in print edition October, 2009

The widespread deployment and use of Information Technologies (IT) has paved the way for change in many fields of our societies. The Internet, mobile computing, social networks and many other advances in human communications have become essential to promote and boost education, technology and industry. On the education side, the new challenges related with the integration of IT technologies into all aspects of learning require revising the traditional educational paradigms that have prevailed for the last centuries. Additionally, the globalization of education and student mobility requirements are favoring a fluid interchange of tools, methodologies and evaluation strategies, which promote innovation at an accelerated pace. Curricular revisions are also taking place to achieved a more specialized education that is able to responds to the society's requirements in terms of professional training. In this process, guaranteeing quality has also become a critical issue. On the industrial and technological side, the focus on ecological developments is essential to achieve a sustainable degree of prosperity, and all efforts to promote greener societies are welcome. In this book we gather knowledge and experiences of different authors on all these topics, hoping to offer the reader a wider view of the revolution taking place within and without our educational centers. In summary, we believe that this book makes an important contribution to the fields of education and technology in these times of great change, offering a mean for experts in the different areas to share valuable experiences and points of view that we hope are enriching to the reader. Enjoy the book!