Artificial Neural Network Models for Prediction of Ozone Concentrations in Guadalajara, Mexico

Ignacio García[1], José G. Rodríguez[2] and Yenisse M. Tenorio[2]
[1]Centro Mexicano para la Producción Más Limpia (IPN), Departamento de Posgrado,
Barrio La Laguna, Col. Ticomán, Delegación Gustavo A. Madero,
[2]Escuela Superior de Ingeniería y Arquitectura (IPN-ESIA Zacatenco), Sección de
Investigación y Posgrado, U. Prof. "Adolfo López Mateos",
Zacatenco, Del. Gustavo A. Madero,
México,

1. Introduction

Advances in mathematical models to describe the formation, emission, transport and disappearance of air pollutants have led to a greater understanding of the dynamics of these pollutants. However, the more complex the model, the more information is required for their application to have sufficient certainty that the results will have technical or scientific value (Russell & Dennis, 2000). These deterministic models require much information that is not always possible to obtain; the data available have not always resulted in successful outcomes upon application of the model (Roth, 1999), or the cost of obtaining reliable data can be prohibitive (Pun & Louis, 2000).

There are other methods requiring less information that can be used to study air pollution in some areas. These methods generally make use of statistical techniques such as regression or other data-fitting methods using numerical techniques to establish the respective relationships between the various physicochemical parameters and variable of interest based on routinely-measured historical data.

The main objectives of these methods include investigating and assessing trends in air quality, making environmental forecasts and increasing scientific understanding of the mechanisms that govern air quality (Thompson et al., 2001).

Among the techniques being examined to relate air quality in a given area to measured physical and chemical parameters, the three that have been used most often are i) multivariate regression (Hubbard & Cobourne, 1998, Comrie & Diem, 1999 , Davis & Speakman, 1999; Draxler, 2000, Gardner & Dorling, 2000), ii) artificial neural networks (ANN) (Perez & Reyes, 2006; Brunelli et al., 2006; Thomas & Jacko, 2007; Grivas & Chaloulakou, 2005; Gardner & Dorling, 1999), and iii) time series and spectral analysis (Raga & Moyne, 1996, Chen et al., 1998; Milanchus et al., 1998, Salcedo et al., 1999, Sebald et al., 2000).

Artificial neural networks have greater flexibility, efficiency and accuracy, since they have a large number of features similar to those of the brain; i.e., they are capable of learning from experience, of generalizing from previous cases to new cases, and of abstracting essential features from inputs containing irrelevant information; they use adaptive learning, one of

the most attractive features of ANN, as well as the ability to learn to perform tasks based on training or initial experience. ANN do not need an algorithm to solve a problem because they can generate their own distribution of the weights of the links through learning and are easily inserted into the existing technology. Because of these characteristics, ANN generally has low computational requirements and their construction is less complex.

The pollutant of interest in this study is tropospheric ozone, as it is the main component of a type of air pollution known as smog or photochemical smog. According to the National Ecology Institute (NEI), the Metropolitan Zone of Guadalajara, Mexico (GMA) is in second place in Mexico in exceeding the NOM-020-SSA1-1993 Mexican air pollution standard. Tropospheric ozone is one of the five major pollutants with harmful effects on human health, causing respiratory problems and ailments such as headaches, and eye irritation as well as affecting vegetation, metals and construction materials, dyes and pigments.

1.1 Tropospheric ozone formation

Photochemical smog is formed through a photochemical process from a combination of gases in the troposphere, such as nitrogen oxides (NO_X, i.e., NO and NO_2), volatile organic compounds (VOCs) and carbon monoxide (CO), as has been documented (Seinfeld, 1978; Boubel, 1994 & Godish, 1991, as cited scientist in Comrie, 1997).
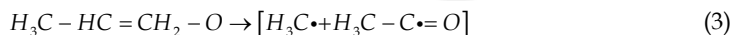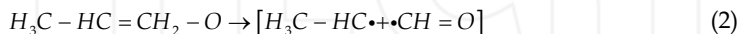
The sequence of events begins in the early hours of the morning when a heavy emission of hydrocarbons (HC) and nitrogen monoxide (NO) is produced at the start of human activity in large cities (heaters are turned on, and traffic density increases). Nitric oxide (NO) is oxidized to nitrogen dioxide (NO_2), increasing the concentration of the latter in the atmosphere. Higher concentrations of NO_2 together with increasing solar radiation as the morning wears on starts the photolytic NO_2 cycle, generating atomic oxygen which, as it is transformed into ozone, leads to an increase in the concentration of oxygen and hydrocarbon free radicals. These, when combined with significant amounts of NO, cause NO in the atmosphere to decrease.

This impedes completion of the photolytic cycle, rapidly increasing the ozone (O_3) concentration (Comrie, 1997).

These relationships can be expressed conceptually; the polluted urban atmosphere contains approximately one hundred different hydrocarbons, olefins being the most reactive. The result of the atomic oxygen attack on the olefin produces two free radicals. In the case of propylene, the first stage of the reaction is the addition of oxygen to the double bond to give a reactive complex (1)

$$H_3C - HC = CH_2 + O \rightarrow \left[H_3C - HC = CH_2 - O \right] \qquad (1)$$
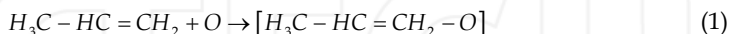
which can break up in two different ways (reactions 2 and 3)

$$H_3C - HC = CH_2 - O \rightarrow \left[H_3C - HC\bullet + \bullet CH = O \right] \qquad (2)$$

$$H_3C - HC = CH_2 - O \rightarrow \left[H_3C\bullet + H_3C - C\bullet = O \right] \qquad (3)$$

The more likely reaction is (2), since it implies less regrouping of the activated complex than (H). $CHO\bullet$ and $CH_3CO\bullet$ radicals quickly form formaldehyde and acetaldehyde, respectively. Reactions (2) and (3) are the initial stages of a chain process

$$CH_3\bullet + O_2 \rightarrow CH_3O_2\bullet \qquad (4)$$

$$CH_3O_2\bullet+NO \rightarrow CH_3O\bullet+NO_2 \tag{5}$$

$$CH_3O\bullet+O_2 \rightarrow HCHO + HO_2\bullet \tag{6}$$

$$HO_2\bullet+NO \rightarrow OH\bullet+NO_2\bullet \tag{7}$$

$$C_3H_6 + OH\bullet \rightarrow CH_3CH_2\bullet+H_2O \tag{8}$$
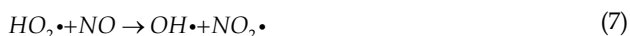
The chain reaction enables rapid oxidation of NO to NO_2 by alkoxyl radicals ($RO\bullet$) and peroxyacyl ($RO_2\bullet$) without the intervention of atomic oxygen and O_3, which provides some explanation for the changes observed in the concentration of gaseous pollutants during the day. When atmospheric concentrations of hydrocarbons increase because of motor vehicle activity, the photolytic cycle of NO_2 is disturbed and NO is oxidized to NO_2 by the chain reaction involving the hydrocarbon radical (equations 2–8). As a result, the constant low O_3 concentration found in the photolytic cycle of NO_2 grows, and ozone is not consumed in the oxidation of NO to NO_2 (Seinfeld, 1978).

As the morning advances, solar radiation promotes the formation of photochemical oxidants, increasing their concentration in the atmosphere. When concentrations of precursors (NO_X and HC) in the atmosphere are lowered, the formation of oxidants stops and their concentrations decrease as the day progresses. Hence, photochemical pollution in cities builds up mainly in the mornings.

Due to industrial development in the GMA in recent years, there has been an urban–green–industrial zone imbalance, leading to the generation of various kinds of pollutants that alter the quality of the environment and exceed the assimilative capacity of the ecosystem.

Given this situation, it is vital to have a mathematical model that correctly predicts ozone concentrations at any given time, as this will help determine preventive measures and/or corrective actions to prevent exposure to high ozone concentrations. These models are able to relate air quality to certain other specific parameters of the air shed, such as emission levels and weather conditions.

2. Data sources

From an analysis of reports from 2002–2005, it was determined that the highest ozone concentrations were in the southern area of the GMA, so specific data for meteorological and chemical variables were obtained from the Miravalle weather station, located in the south. These are shown in Table 1.

Station \ Year	2002	2003	2004	2005
Las Águilas	0.169	0.165	0.164	0.131
Atemajac	0.152	0.185	0.165	0.144
Centro	0.166	0.171	0.157	0.137
L. Dorada	0.225	0.195	0.197	0.215
Miravalle	0.232	0.225	0.226	0.154
Tlaquepaque	0.142	0.149	0.138	0.109
Vallarta	0.171	0.217	0.175	0.096

Table 1. Peak ozone concentrations (ppm) for the years 2002, 2003, 2004 and 2005 (Semades, 2005)

2.1 Meteorological and chemical variables

Meteorological data for the period April 1999 to June 2005 were obtained from the Mexican National Weather Service (MNWS). These data consist of averages over time intervals ranging from 0 to 23 hrs.

The meteorological variables are Wind Direction (average and maximum average) (degrees), wind speed (average and maximum average) (km/h) Average Temperature (°C), Relative Humidity (%). Barometric Pressure (mbar), Precipitation (mm) and Solar Radiation (W/m^2). The data were obtained from the Chapala station, which belongs to the Automatic Monitoring Stations (AMS) system.

Data on the following chemical variables were provided by the National Ecology Institute (NEI) for the Miravalle station; Ozone, Nitrogen Oxides— NO_X and NO_2, as shown in Figure 1.

Fig. 1. Distribution of GMA Atmospheric Monitoring Automatic Network (Semarnat & INE, 2009).

3. Selection of meteorological and chemical variables

Meteorological and chemical variables used to carry out ground-level ozone forecasts were selected based on existing knowledge from the scientific literature and an analysis of correlations between different variables, and on availability of data from monitoring stations.

3.1 Analysis of meteorological variables
3.1.1 Wind speed

Atmospheric movements of the air (i.e. winds) are responsible for the spread of high concentrations of pollutants (in this case the O_3 and its precursors) through the atmosphere, but this may or not occur quickly, because if the winds are calm, i.e., the wind speed is low and the topography traps the air mass, pollutants can not disperse. More pollutants continue to accumulate and their concentration can reach very high levels. In contrast, if wind speeds are high, the pollutants tend to disperse quickly (Melas et al., 2000).

3.1.2 Temperature

This variable has shown a strong correlation with the concentration of ozone. The basic reasoning is that photochemical reaction rates are sensitive to temperature, so that increasing the temperature in the troposphere stimulates a series of interlinked reactions that contribute to ozone formation. (Garcia, 2003)

3.1.3 Relative humidity

Water vapor is one of the most basic components of the atmosphere. Its amount can be quite variable. It is important because it is one of the atmospheric elements which most absorbs solar radiation, preventing it from interacting with the primary pollutants and forming secondary pollutants such as ozone (Ayllón, 1996).

3.1.4 Precipitation

This process is one of the main ways that pollutants are removed from the atmosphere, but as a result, pollutants removed from the air then contaminate the earth's surface, which in some cases results in their becoming even more active due to their effects on surface water, plants and materials (acid rain) (Melas et al., 2000).

3.1.5 Pressure

The relationship between temperature and pressure is that the vertical motion of air is determined by vertical variation in temperature in the troposphere; temperature decreases at a rate of 0.64 °C per 100 m of altitude. Thus, the earth's surface warms the air parcel next to it, and this hot air expands, becoming less dense than the cooler air above it. The warm air rises and cool air takes its place to then be heated in turn, making contact with the surface, and subsequently also rises. This creates air currents (vertical mixing) that contribute to the dispersion of pollutants (Rodríguez & Tenorio, 2006).

3.1.6 Solar radiation

This is the factor that has the greatest effect on photochemical reactions, i.e., it is involved in the formation and destruction of the various compounds involved in the increase of tropospheric ozone (Melas et al., 2000).

Photochemical dissociation in the atmosphere can be considered as a two-step process. The dissociation energy of a photon by a molecule causes it to be in an excited state, and the excited product disassociates into new products that can be highly reactive, generating photochemical smog (Wark & Warner, 2000), as explained in Section 1.1.

3.2 Analysis of chemical variables

Many pollutants are highly persistent, and it is generally accepted that the probability of pollution episodes is increased if the previous day's pollution levels were higher than normal.

In this study, the previous day's maximum O_3 and NO_X are used as chemical input variables (Melas et al., 2000).

3.2.1 Previous day's ozone

Even when the tropospheric ozone photolytic cycle is considered to be in equilibrium (generation and degradation of ozone in equilibrium with NOx), when hydrocarbons are involved (equations 2–8) the ozone generated is not consumed in the oxidation of NO to NO_2 (Seinfeld, 1978). There is ozone remaining from the previous day, which should be

input to the structure of the neural network model as an initial ozone concentration on the day of interest.

3.2.2 Oxides of nitrogen

These variables are the main precursors to ozone formation. Oxides that are present in the atmosphere in significant quantities are nitrogen monoxide and nitrogen dioxide (NO_x = $NO + NO_2$); approximately 90% of them are destroyed by photolysis in the formation of ozone (Wark & Warner, 2000).

These variables experience higher photolytic breakdown between 11 a.m. and 3 p.m.; i.e., when there is a higher incidence of light, after which time the levels start to gradually rise. Thus the concentrations which remain following the photolytic period participate as raw material for the formation of ozone the next day. Therefore, the maximum concentration between 3 p.m. and 11 p.m. on the previous day is used as the input variable (Seinfeld, 1978).

3.3 Statistical analysis

In addition to the analysis described in 2.1 and 2.2, the selection of variables was based on 1) individual regressions between the variable of interest (maximum ozone) and the various parameters (temperature, humidity, etc.), selecting only those with correlation coefficients (r) 0.3 and greater (Garcia, 2003); and 2) the t-ratio (t), used to obtain the degree of importance of each of the variables with respect to the dependent variable, namely ozone (see Table 2).

This analysis indicates that the higher the absolute value of *t*, the more necessary the variable (Miller et al., 1992) in the ANN model. All values of meteorological variables correspond to the day in question, so if models are to be used in real time, the values for that day are required. The values of the chemical variables correspond to the day previous to the day of interest.

Year / Variable	1999	2000	2001	2002	2003	2004
Constant	-2.344	-0.242	-10.033	-0.591	-1.295	-0.795
Wind speed	-0.049	-0.365	1.859	-0.719	-2.617	-3.490
Temperature	7.847	4.637	2.519	10.951	7.031	5.036
Humidity	1.340	1.583	-3.777	-0.437	-5.512	-3.978
Pressure	2.261	0.197	10.228	0.499	1.358	0.806
Precipitation	-0.071	-0.905	.635	-1.721	-1.692	-1.092
Solar radiation	-2.959	-0.778	3.132	-2.973	-2.699	0.316
NO_2 previous day	0.243	-0.875	1.293	0.105	1.213	3.654
NO_x previous day	-1.143	1.223	-0.950	-0.296	0.842	-0.152

Table 2. t-ratio values by variable and year.

With the results shown in Table 2, the database used had daily maximum temperature values (°C) Solar Radiation (W/m²), nitrogen dioxide (NO₂) on the previous day and oxides of nitrogen (NO$_X$) on the previous day, the latter two measured from 3 p.m. to 11 p.m.; and daily average values of wind speed (km/h), humidity (%) pressure (mbar) and precipitation (mm).

4. Artificial neural networks

The artificial neural network employed was a multilayer backpropagation network, which has been used successfully in several studies (Garcia & Shigidi, 2005, Kuo et al., 2003, Helle et al., 2001; Yesilnacar et al., 2007; Yetilmezsoy & Demirel, 2007).

The important feature of this network is its ability to self-adapt the weights of neurons in intermediate layers to learn the relationship between a set of patterns given as examples and their corresponding outputs, so that after having been trained, it can apply the same relationship to new input vectors and produce appropriate outputs from inputs that the system has never seen before, a feature known as the generalizability of an ANN (Mehrotra et al., 1997).
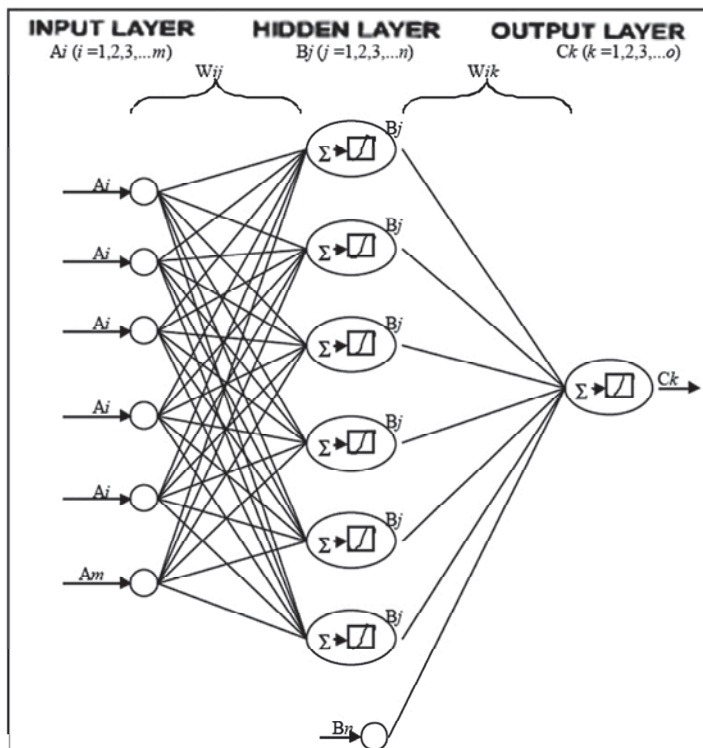

Fig. 2. Schematic example of an *m × n × o* artificial neural network, showing a multilayer perceptron with a 4 × 6 × 1 structure (additional shaded circles indicate bias nodes), which each contain an activation function ($\sum$) and a nonlinear transfer function (Comrie, 1997).

This type of network consists of three layers. There is an input layer A_i with *m* neurons, an output layer C_k with *o* neurons and at least one hidden layer B_j with *n* internal neurons. Each neuron in a layer (except the input layer) receives inputs from all neurons of the previous layer and sends its output to all neurons of the next layer (except the output layer), as shown in Figure 2 (Comrie, 1997).

The learning algorithm involves a forward propagation step in which the input pattern is presented to the network and propagated through the layers until it reaches the output layer, followed by a backward propagation step in which errors are passed from the output layer back to all the neurons in the intermediate layer and then back to the input layer, adjusting the synaptic weights so that the system converges (Gurney, 2003). In this way the network learns to recognize different features of the input patterns (Freeman et al., 1993, Lek et al., 2000). The important point is that each iteration of the ANN decreases the error between the actual data and forecast values.

To develop the ANN, we used the MATLAB version 7.4.0.287 computer program (R2007a), specifically the Neural Network Toolbox toolkit (Wang et al., 2006, Yetilmezsoy & Demirel et al., 2008, Garcia et al., 2008).

4.1 Parameters used to build artificial neural network models

The training algorithm selected for use in the ANN models was the Levenberg-Marquardt algorithm, because it achieves rapid convergence (TRAINLM) (Beale et al., 2010, Yetilmezsoy & Demirel et al., 2008 Yesilnacar et al., 2007, Wang et al., 2006) with a learning rate of 0.001. It is worth noting that when the learning rate increases or decreases, the performance of the models are neither improved nor impeded, so the 0.001 value was maintained. The ANN models were trained with 10,000 iterations on the training data (Comrie, 1997; Guardani et al., 1999).

To evaluate the results of the ANN models, three performance features were considered; the mean square error (MSE), mean squared error with regularization (MSEREG) and the error sum of squares (SSE). In the hidden layer, a log-sigmoidal function (*LOGSIG*) was used, and in the output layer, the transfer function was linear (PURELINE).

4.2 Number of hidden layers and hidden layer neurons

ANN models generally have acceptable performance with three layers; input, hidden and output (Del Brio & Sanz, 2001, Yetilmezsoy & Demirel et al., 2008, Helle et al., 2001). Deciding the number of neurons in the hidden layer is usually not so obvious, so the decision was based on the rules suggested by Goethals et al., (2007). The number of hidden neurons is based on the number of input variables (N_i) and output nodes (N_o) as shown in Table 3.

$2/3 * N_i$	$2/3 * (9) = 6$
$0.75 * N_i$	$0.75 * (9) = 6.75 \approx 7$
$0.5 * (N_i + N_o)$	$0.5 * (9 + 1) = 5$
$2 * N_i + 1$	$(2 * (9)) + 1 = 19$
$2 * N_i$	$2 * (9) = 18$

Table 3. Rules suggesting the number of hidden neurons based on the number of input (N_i) and/or output (N_o) nodes Goethals et al., (2007).

With these rules, and various studies predicting tropospheric ozone (Comrie, 1997; Guardani et al., 1999; Hooyberghs et al., 2005, Jiang et al., 2004; McKendry et al., 2002; Melas et al., 2000, David & Speakman, 1999, Hubbard & Cobourn, 2001), four arrays were created for training and testing the ANN models, each with a different number of neurons in the hidden layer as shown in Table 4. This is because there is no rule for the optimal number of neurons in the hidden layer (Yetilmezsoy and Demirel et al., 2008, Helle et al., 2001). In each problem, different arrangements should be tried for organizing the internal representation, selecting the one that gives the best results according to the stated objectives.

5. Development of artificial neural network models

Four different ANN structures were created; 9×6×1 (nine input signals, six nodes in the transfer layer and one node in the output layer) (R-6); 9×10×1 (R-10), 9×12×1 (nine input signals, twelve nodes in the transfer layer and one node in the output layer) (R-12) and 9×15×1 (R-15).

In training, the three performance criteria mentioned above were used; mean square error (MSE), mean squared error with regularization (MSEREG) and error sum of squares (SSE).

As a source of data for the training process, a database was used which contained daily maximum temperature values (°C), Solar Radiation (W/m²), nitrogen dioxide (NO_2) on the previous day, and nitrogen oxides (NO_X) on the previous day; the latter two for the period 3 p.m. to 11 p.m., and daily average values of Wind Speed (km/h), Humidity (%) Pressure (mbar) and Precipitation (mm) for the period 1999–2004. The database contained 2065 validated data for each variable.

5.1 Training of ANN models

The models were trained on 1990–2004 data. Real and predicted O_3 values were classified into three concentration ranges according to the NOM-020-SSA1-1993 standard, quantifying the number of data falling within each range (percentage of correct answers) in order to estimate the performance of each model.

This was done under the assumption that it would be very difficult to try to get specific concentrations from models. The concentration ranges were: low <0.06 ppm, intermediate 0.06–0.11 ppm and high >0.11 ppm, the latter corresponding to the maximum allowable range of the rule.

The number of observed values and estimated values in each of these ranges were counted to find how many times the model made a correct estimate.

In order to refine the 12 trained models, an analysis based on the correlation coefficient (r) was carried out. A linear regression was performed between real ozone and estimated ozone as shown in Table 4. Based on the results, the models with 12 and 15 neurons in the hidden layer were chosen, as they had the best correlation between the calculated and actual data, in both cases using MSE to evaluate performance.

No. of neurons in hidden layer	Performance criterion		
	MSE	MSEREG	SSE
6	0.740	0.664	0.744
10	0.735	0.688	0.752
12	0.753	0.722	0.743
15	0.744	0.735	0.739

Table 4. Correlation coefficients of the ANN with each performance criterion.

Figure 3 and 4 shows correlation plots of the networks with 12 and 15 neurons in the hidden layer, respectively, during the training phase using 1999–2004 data. The dispersion of clouds of points is very similar in the two cases, corresponding to correlation coefficients close to each other.
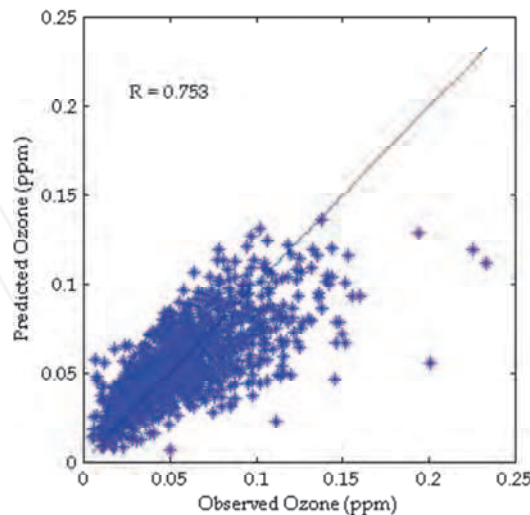

Fig. 3. Scatterplot of the predicted vs. observed ozone concentrations (ppm) for the model with 12 neurons in the hidden layer in the training phase for the years 1999–2004.
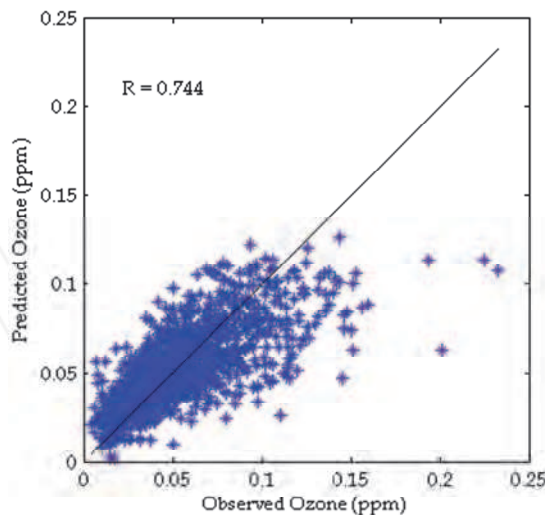

Fig. 4. As in Figure 3, but for the model with 15 neurons in the hidden layer.

5.2 Performance of ANN models

Once training was completed, performance was evaluated with 2005 data. This second database has 173 validated values for each parameter under the same conditions as the training database, containing daily data from January to June 2005.

Table 5 shows the number and percentage of times that correct concentrations were obtained by the networks with respect to the observed values, where R-12 and R-15 refer to networks with 12 and 15 neurons in the hidden layer respectively.

	Low (<0.06 ppm)	Medium (0.06–0.11 ppm)	High (>0.11 ppm)
Total days	79	77	17
R-12	41 (52%)	33 (43%)	0 (0%)
R-15	50 (63%)	48 (62%)	2 (12%)

Table 5. Number and percentage of correct estimated O3 values with respect to observed values, at the performance stage.

Figure 5 shows the ANN regression models corresponding to 12 neurons in the hidden layer (R-12), and Figure 6 shows the ANN model with 15 neurons in the hidden layer (R-15), both models with data from January to June 2005.

Figure 5 shows the regression with the R-12 network, with a regression coefficient of 0.575, and Figure 6 shows the correlation of the network with 15 neurons and a regression coefficient of 0.545.
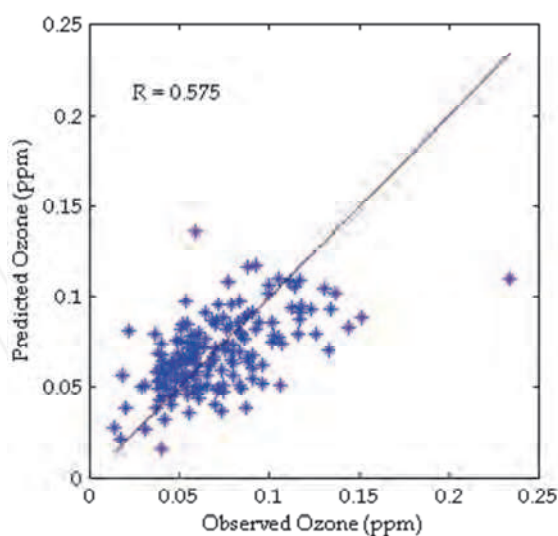

Fig. 5. Scatterplot of the predicted concentrations with observed ozone concentrations for the model with 12 neurons in the hidden layer in the test phase for the year 2005.
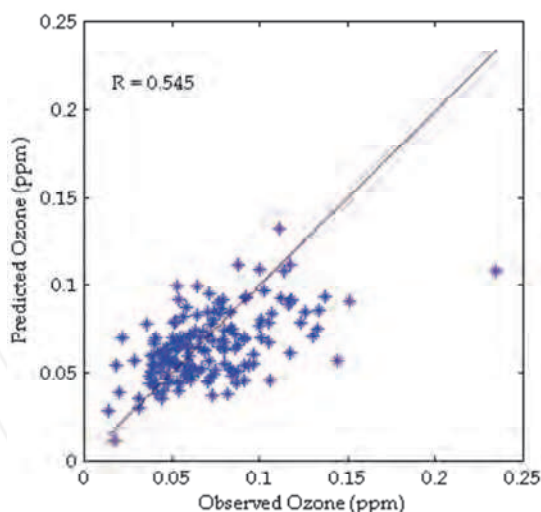
Fig. 6. As in Figure 5, but for the model with 15 neurons in the hidden layer.

The performance of the two networks is very similar. However, it is clear that it is difficult for both models to detect ozone concentrations exceeding the standard, which is important for this study.

In order to remedy the estimation problem in the high concentration range, it was decided to scale the value estimated by the network. Thus, the final estimated value of ozone (O_3) is calculated as

$$\hat{O}_3 = \alpha \hat{O}_3 \qquad\qquad (9)$$

where α is the scaling factor and O_3 is the ozone concentration (ppm) estimated by the neural network model. The value of α was obtained by an incremental search of values that when the equation was applied, reached the number of times that the observed concentration standard for each year was exceeded, without knowing (until then) if the times when estimated concentration exceeded the standard also corresponded to the days when this actually occurred. This process yielded an average value of α for both models of 1.21. Thus ozone values estimated by each network were multiplied by 1.21, giving the results in Table 3, which shows that efficiency is lost in the lower range at the expense of a gain in the intermediate and high ranges (which are of greatest interest).

	Low (<0.06 ppm)	Intermediate (0.06-0.11 ppm)	High (>0.11 ppm)
Total days	79	77	17
R-12 (α=1.21)	18 (23%)	57 (74%)	11 (65%)
R-15 (α=1.21)	23 (29%)	57 (74%)	8 (47%)

Tabla 6. Number and percentage of correct O3 values estimated by different networks using the scaling.

Table 6 shows that the overall performance of the networks is 50%, and 64% for the detection of elevated ozone concentrations with R-12, and 47% with R-15. It may be noted that using 12 neurons in the hidden layer fails to detect a greater number of days in the high range.

With the results, it was decided to work with the model with 12 neurons in the hidden layer; this model was selected in the present study for predicting tropospheric ozone concentrations in the metropolitan area of Guadalajara, Jalisco. Figure 7 shows a comparison with the previously selected model without the scaling factor. As can be seen graphically, the model shows a good trend in the low and intermediate range of tropospheric ozone concentrations but a poor performance for the higher concentrations.
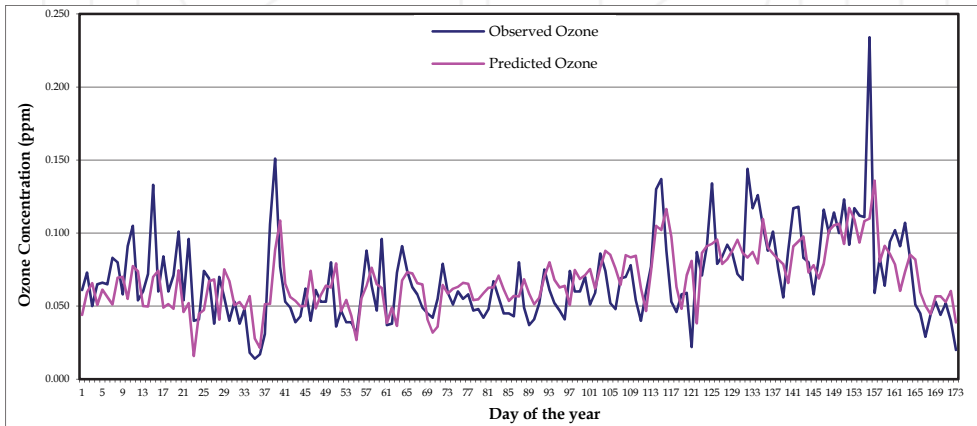

Fig. 7. Comparison of predicted concentrations Vs. observed concentrations of ozone from ANN model with 12 neurons in the hidden layer for 2005 without the scaling factor.

In Figure 8 a scaling factor of 1.216 has been applied. Although accuracy is lost in the low range of concentrations, performance in the midrange is improved from 43% to 74%, that is, by 31 percentage points.
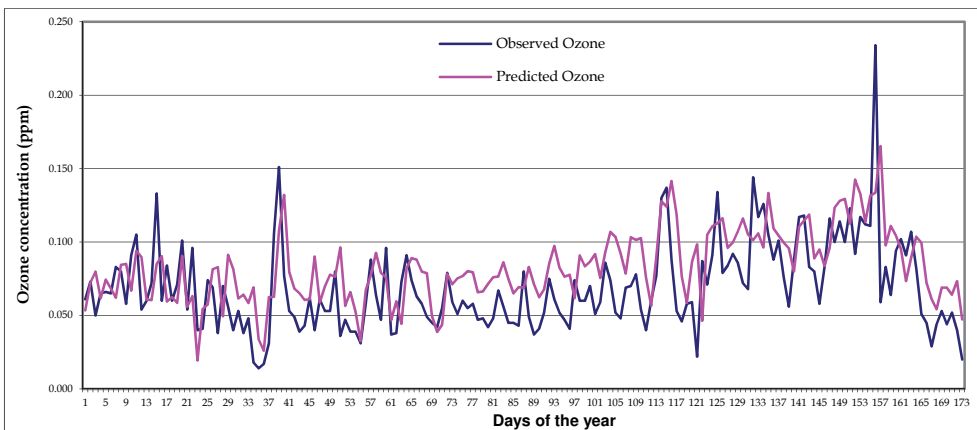

Fig. 8. As in Figure 7, but with the scaling factor.

In the high range, it is clear that after the unscaled model did not detect any concentrations, after applying the scaling factor, efficiency was improved by 65%, i.e., it predicted 11 of the 17 days that exceeded the concentration of 0.11 ppm.

6. Conclusions

With the results obtained by the model and selected variables, it was concluded that in the GMA, the most important meteorological variables for significantly reducing the tropospheric O_3 concentration are wind speed, which can disperse ozone precursors or ozone itself, decreasing the concentration of the pollutant; and rainfall, as this will wash out the atmosphere thereby lowering the concentration of ozone, as well as the concentrations of precursors and other pollutants (it can be observed that the morning after a day with rainfall tends to be clear in the early hours).

The chemical variables that are important in increasing ozone concentration are maximum temperature, maximum solar radiation, O_3 on the previous day, and oxides of nitrogen (NO_X and NO_2) (Gomez et al., 2006) because they are involved directly in the photochemical cycle of ozone formation (Wark & Warner, 2005).

The important variables related to nitrogen oxides (NO_2 and NO_X) are maximum values, since these elements are directly involved in the photolytic cycle and influence the formation of ozone. They are measured from 3 p.m. to 11 p.m. because ozone formation takes place between 11 a.m. and 3 p.m., so there is no consumption of these two pollutants and accumulations of these oxides in the afternoon will serve as raw material for the next day.

Along with the physical and chemical meteorological variables involved in increasing or decreasing O_3 concentrations, the characteristics of the basin of the Jalisco Basin are also important. It is located 1583 meters above sea level, surrounded by the Sierra Madre Occidental, plateaus and the Neovolcanic Belt, and with industrial parks to the NW, SE, and SW. The Sierra Madre Occidental is formed by the Los Huicholes, Los Guajolotes and San Isidro mountains, the Gordo hill and the Tequila volcano. The Neo-Volcanic or Transversal Volcanic Belt includes the Cacoma, Manantlán Tapalpa and Lalo mountains, among others. Other notable peaks are El Tigre and Garcia, Cerro Viejo, the Tequila Volcano and to the south, the Nevado de Colima mountain and Colima Volcana, which create a particular basin structure and formation and dispersal patterns of specific pollutants that directly affect the GMA in terms of the model performance.

The ANN models perform acceptably for predicting ozone in the lower and intermediate range, however the aim of this study was to predict high levels of ozone, so it was necessary to use the scaling factor so that the models would be able to predict concentrations in the high range. This scaling factor was obtained in training the models by matching observed and predicted days that exceeded the standard. Using the scaling factor, the model obtained can predict maximum O_3 concentrations in ppm with an overall efficiency of approximately 50%, and 65% for the detection of high concentrations.

These models were trained on data from the period between 1999 and 2004 in the Guadalajara Metropolitan Area (GMA) and their performance was evaluated using data from the period from January to June 2005 with data from the SMN and INE. The performance of the two models was assessed and compared by comparing forecast and actual ozone concentrations in three ranges (called percent of correct answers) related to the NOM-020-SSA1-1993 standard, low ($C_{O_3} < 0.060$ ppm), intermediate ($0.060 \leq C_{O_3} \leq 0.110$

ppm) and high (C_{O_3} > 0.110 ppm), with high concentrations considered to be values that exceeded the standard.

The general characteristics of the selected ANN model for forecasting of O_3 are: 9 independent variables (3 chemical and 6 weather). The structural arrangement of the network was 9×12×1 (input × hidden × output); transfer functions were sigmoidal in the hidden layer and linear in the output layer, the training function was TRAINLM; the performance criterion was mean square error (MSE) and the scaling factor was 1.21.

This model is able to predict 22% of concentrations lower than 0.060 ppm, i.e. it predicted 17 of 79 days for this range; a 74% success rate in the intermediate range of concentrations from 0.060 to 0.110 ppm, i.e. 57 days of the 77 days recorded; and 65% success for concentrations greater than 0.110 ppm, i.e. 11 of the 17 days recorded for the 2005 period.

The overall efficiency of the model for the period January to June 2005 was 49.13% with the scaling factor and 54.34% without the factor.

The models obtained employed the meteorological variables maximum temperature and solar radiation, and average values of wind speed, barometric pressure, rainfall, relative humidity for each day of interest, and maximum values of the chemical variables ozone, NOx and NO_2 on the previous day (measured from 3 p.m. to 11 p.m.).

Finally, the models generated are easy to implement, have only moderate technological requirements and simple, easily understood structures, giving them minimal operating costs. These models can be used to help alert the community at times when the air quality is undesirable, so that precautionary measures can be taken to safeguard the health of the population.

7. References

Ayllón, M. T. (Second edition). (2003). Elementos *de meteorología y climatología*. Trillas, ISBN 968-24-6725-X, México.

Beale, M. H.; Hagan, M. T. & Demuth, H. B. (Version 7). (2010). *Neural Network Toolbox 7, User's Guide,* The MathWorks, Inc., ISBN 0-9717321-0-8, Natick, Massachussetts, USA.

Brunelli, U.; Piazza, U. & Pignato, L. (2007). Two-days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo, Italy. *Atmospheric Environment,* Vol. 41, No. 14 (May 2007), pp. 2967–2995, doi: 10.1016/j.atmosenv.2006.12.013

Chen, J.-L.; Islam, S. & Biswas, P. (1998). Nonlinear dynamics of hourly ozone concentrations: nonparametric short term prediction, *Atmospheric Environment*, Vol. 32, No. 11 (June 1998), pp. 1839–1848, doi: 10.1016/S1352-2310(97)00399-3

Cobourn, W. G. & Hubbard, M.C. (1999). An enhanced ozone forecasting model using air mass trajectory analysis, *Atmospheric Environment*, Vol. 33, No. 28 (December 1999), pp. 4663–4674, doi: 10.1016/S1352-2310(99)00240-X

Comrie, A. (1997). Comparing neural networks and regression models for ozone forecasting, *Air & Waste Manage. Assoc.,* Vol. 47, (June 1997), pp. 653 – 663, ISSN 1047– 3289.

Comrie, A.C. & Diem, J.E. (1999). Climatology and forecast modeling of ambient carbon monoxide in Phoenix, Arizona, *Atmospheric Environment,* Vol. 33, No. 30, (October 1999), pp. 5023–5036, doi: 10.1016/S1352-2310(99)00314-3

Davis, J.M. & Speckman, P. (1999). A model for predicting maximum and 8 h average ozone in Houston, Atmospheric Environment, Vol. 33, No. 16, (July 1999), pp. 2487–2500 begin_of_the_skype_highlighting, doi: 10.1016/S1352-2310(98)00320-

3end_of_the_skype_highlighting

Del Brío, B. M. & Sanz, A. (3ª Edición). (2006). *Redes Neuronales y Sistemas Borrosos*, Ra–Ma Editorial, ISBN 978-84-7897-743-7, Spain.

Draxler, R. R. (2000). Meteorological factors of ozone predictability at Houston, Texas, *Journal of the Air & Waste Management Association,* Vol. 50, No. 2, (February 2000), pp. 259–271, PMID: 10680356

Freeman, J. A. & Skapura, D. M. (1993). *Redes Neuronales: algoritmos, aplicaciones y técnica de programación*. Díaz de Santos, ISBN 9780201601152, Spain.

García, I. (2003). *Aplicación de modelos semi-empíricos para el análisis y pronóstico de la calidad del aire en el Área Metropolitana de Monterrey, N.L.*, Master's thesis, ITESM, Monterrey, México.

García, I.; Marbán, A.; Tenorio, Y. M. & Rodríguez, J. G. (2008). Pronóstico de la Concentración de Ozono en Guadalajara-México usando Redes Neuronales Artificiales, *Información Tecnológica*, Vol. 9, No. 3, (Junio 2010), pp. 89–96, doi: 10.1612/inf.tecnol.3925it.07

García, L. A. & Shigidi, A. (2006). Using neural networks for parameter estimation in ground water, *Journal of Hydrology,* Vol. 318, No. 1–4, (March 2006), pp. 215–231. doi: 10.1016/j.jhydrol.2005.05.028

Gardner, M. W. & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) – a review of applications in the Atmospheric Sciences, *Atmospheric Environment,* Vol. 32, No. 14–15, (August 1998), pp. 2627–2636, doi: 10.1016/S1352-2310(97)00447-0

Gardner, M.W. & Dorling, S. R. (2000). Statistical surface ozone models: an improved methodology to account for nonlinear behaviour, *Atmospheric Environment*, Vol. 34, No. 1, (January 2000), pp. 21–34, doi: 10.1016/S1352-2310(99)00359-3

Goethals, P. L. M.; Dedecker, A. P.; Gabriels, W.; Lek, S. & De Pauw, N. (2007). Applications of artificial neural networks predicting macroinvertebrates in freshwaters, *Aquatic Ecology,* Vol. 41, No. 3 (May 2007), pp. 41, 491–508, doi: 10.1007/s10452-007-9093-3

Gómez, J; Martín, J.D.; Soria, E.; Vila, J.; Carrasco, J. & Valle, S. (2006). Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration, *Atmospheric Environment,* Vol. 40, No. 32, (October 2006), pp. 6173–6180, doi: 10.1016/j.atmosenv.2006.04.067

Grivas, G. & Chaloulakou, A. (2006). Artificial neural network models for prediction of PM_{10} hourly concentrations, in the Greater Area of Athens, Greece, *Atmospheric Environment*, Vol. 40, No. 7 (March 2006), pp. 12161229. doi: 10.1016/j.atmosenv.2005.10.036

Guardani, R.; Nascimento C.; Guardani, M.; Martins, M. & Romano, J. (1999). Study of atmospheric ozone formation by means of a neural network-based model, *Air & Waste Manage. Assoc.,* Vol. 49, (March 1999), pp. 316323, ISSN 1047 – 3289.

Gurney, K. (1997). *An Introduction to Neural Networks*, University College London Press, ISBN 1-85728-503-4, London.

Helle, H. B.; Bhaatt, A. & Ursin, B. (2001). Porosity and permeability prediction from wireline logs using artificial neural networks: a North Sea case study, *Geophysical Prospecting*, Vol. 49, No. 4 (December 2001), pp. 431–444. doi: 10.1046/j.1365-2478.2001.00271.x

Hooyberghs, J; Mensink, C.; Dumont, G.; Fierens, F. & Brasseur, O. (2005). A neural network forecast for daily average PM10 concentrations in Belgium, *Atmospheric Environment*, Vol. 39, (January 2005), pp. 3279–3289, doi: 10-1016/j.atmosenv.2005.01.050.

Hubbard, M. & Cobourn, W. G. (1998). Development of a regression model to forecast ground-level ozone concentration in Louisville, KY, *Atmospheric Environment*, Vol. 32, No. 14-15, (August 1998), pp. 2637–2647, doi: 10.1016/S1352-2310(07)00444-5

Jiang, D.; Zhang, Y.; Hu, X.; Zeng, Y.; Tan, J. & Shao, D. (2004). Progress in developing an ANN model for air pollution index forecast, *Atmospheric Environment*, Vol. 38, (October 2003), pp. 7055–7064, doi: 10-1016/j.atmosenv.2003.10.066.

Kuo, Y.; Chen-Wuing, L. & Lin, K. (2004). Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan, *Water Research,* Vol. 38, No. 1 (January 2004), pp. 148–158. doi: 10.1016/j.watres.2003.09.026

Lek, S. & Guégan, J. F. (2000). *Artificial Neural Networks: application to ecology and evolution*, Springer, ISBN 3540669213, Michigan.

McKendry, I. (2002). Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2.5) forecasting, *Air & Waste Manage. Assoc.*, Vol. 52, (September 2002), pp. 1096–1101, ISSN 1047–3289.

Melas, D.; Kioutsioukis I. & Ziomas I. (2000). Neural network model for predicting peak photochemical pollutant levels, *Air & Waste Manage. Assoc.*, Vol. 50, (April 2000), pp. 495–501, ISSN 1047-3289.

Mehrotra, K.; Mohan, C. K. & Ranka, S. (Second printing) (2000). *Elements of Artificial Neural Networks*, MIT Press, Cambridge MA.

Milanchus, M.L.; Rao, T. & Zurbenko, I. G. (1998). Evaluating the effectiveness of ozone management efforts in the presence of meteorological variability, *Journal of the Air & Waste Management Association,* Vol. 48, No. 3, (1998), pp. 201-215, ISSN 1096-2247.

Miller, I.R.; Freund, J. E. & Jonson. R. (4th edition). (2004). *Probabilidad y Estadística para Ingenieros*, Reverté Ediciones, Spain.

Secretaria de Salud, (2000). Norma Oficial Mexicana NOM-020-SSA1-1993, Salud ambiental. Criterio para evaluar el valor límite permisible para la concentración de ozono (O$_3$) de la calidad del aire ambiental, Mexico

Pérez P.; Trier, A. & Reyes, J. (2000). Prediction of PM$_{2.5}$ concentrations several hours in advance using neural networks in Santiago, Chile, *Atmospheric Environment*, Vol. 34, No. 8 (February 2000), pp. 1189–1196, doi: 10.1016/S1352-2310(99)00316-7.

Perez P. & Reyes, J. (2006); An integrated neural network model for PM10 forecasting, *Atmospheric Environment*. Vol. 40, (January 2006), pp. 2845–2851, doi: 10.1016/j.atmosenv.2006.01.010.

Pun, B. K.; Louis, J. F.; Pai, P; Seigneur, C.; Altshuler, S. & Franco, G. (2000). Ozone formation in California's San Joaquin Valley: A critical assessment of modeling and data needs, *Journal of the Air & Waste Management Association,* Vol. 50, No. 6, (2000), pp. 961 – 971, ISSN 1096-2247.

Raga, G. B. & Le Moyne, L. (1999). On the nature of air pollution dynamics in Mexico City – I. Nonlinear analysis, *Atmospheric Environment,* Vol. 30, No. 23, (February 1999), pp. 3987–3993, doi: 10.1016/1352-2310(96)00122-7

Rodríguez, J. G. & Tenorio, Y. M. (2006). *Desarrollo de modelos pronóstico para la calidad del aire en la Zona Metropolitana de Guadalajara, Jalisco*, Bachelor's thesis, ESIQIE-IPN, Mexico.

Roth, P. M. (1999). A qualitative approach to evaluating the anticipated reliability of a photochemical air quality simulation model for a selected application, *Journal of the Air & Waste Management Association,* Vol. 49, No. 9, (1999), pp. 1050–1059, ISSN 1096-2247.

Russell, A. & Dennis, R. (2000). NARSTO critical review of photochemical models and modeling, *Atmospheric Environment,* Vol. 34, No. 12-14, (March 2000), pp. 2283–2324, doi: 10.1016/S1352-2310(99)00468-9

Salcedo, R. L. R.; Alvim, M. C. M.; Alves, C. A. & Martins, F. G. (1999). Time-series analysis of air pollution data, *Atmospheric Environment,* Vol. 33, No. 15, (July 1999), pp. 2361–2372, doi: 10.1016/S1352-2310(99)80001-6

Thomas, S. & Jacko, R. B. (2007). Model for forecasting expressway fine particulate matter and carbon monoxide concentration: Application of regression and neural network model, *Air & Waste Management Association*, Vol. 57, No. 4 (April 2007), pp. 480–488, ISSN 1096-2247.

Sebald, L.; Treffeisen, R; Reimery, E. & Hies, T. (2000). Spectral analysis of air pollutants. Part 2: Ozone time series, *Atmospheric Environment,* Vol. 34, No. 21, (June 2000), pp. 3503–3509, doi: 10.1016/S1352-2310(00)00147-3

Seinfeld, J. (1978). Contaminación stmosférica. Fundamentos físicos y químicos, *Instituto de Estudios de Administración Local*, ISBN 8470882139, Madrid.

Secretaria de Medio Ambiente para el Desarrollo Sustentable Jalisco. (2006). *Informe de calidad del aire, evaluación: 2001–2005*, Semades, Retrieved from <http://www.jalisco.gob.mx/wps/wcm/connect/813336004dbe3344a756 ef5160bedb77/ReporteAire2006.pdf?MOD=AJPERES&CACHEID=813336004dbe33 44a756ef5160bedb77>

Secretaria de Medio Ambiente y Recursos Naturales & Instituto Nacional de Ecología (2003). *Segundo almanaque de datos y tendencias de la calidad del aire en seis ciudades mexicanas*, Semarnat & INE, ISBN 968 817 614 177, Mexico.

Thomas, S. & Robert, B. J. (2007). Model for forecasting expressway fine particulate matter and carbon monoxide concentration: Application of regression and neural network model, *Air & Waste Management Association*. Vol. 57, No. 4, (2007), pp. 480–488, ISSN 1096-2247.

Thompson, M. L.; Reynolds, J.; Cox, L. H.M; Guttorp, P. & Sampson, P. D. (2001). A review of statistical methods for the meteorological adjustment of tropospheric ozone, *Atmospheric Environment*, Vol. 35, No. 3, (November 2000), pp. 617–630, doi: 10.1016/S1352-2310(00)00261-2

Wang, M. X.; Liu, G. D.; Wu, W. L.; Bao, Y. H. & Liu, W. N. (2006). Prediction of agriculture derived groundwater nitrate distribution in North China Plain with GIS-based BPNN, *Environment Geology*, Vol. 50, No. 5, (April 2006), pp. 637–644, doi: 10.1007/s00254-006-0237-x

Wark, K. & Warner, C. F. (2004). *Contaminación del aire: Origen y control*, Limusa – Wiley, ISBN 9789681819545, Mexico.

Yesilnacar, M. I.; Sahinkaya, E.; Naz, M & Ozkaya, B.; Naz M. & Bestamin O. (2007). Neural network prediction of nitrate in groundwater of Harran Plain, Turkey, *Environmental Geology,* Vol. 56, No. 1 (November 2007), pp. 19–25 doi: 10.1007/s00254-007-1136-5

Yetilmezsoy, K. & Demirel, S. (2008). *A*rtificial neural networks (ANN) approach for modeling of Pb (II) adsorption from aqueous solution by Antep pistachio (*Pistacia Vera L.*) Shells, *Journal of Hazardous Materials*. Vol. 153, No. 3, (May 2008), pp. 1288–1300. Doi: 10.1016/j.jhazmat.2007.09.092
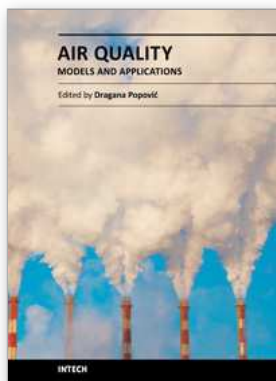
Air Quality-Models and Applications

Edited by Prof. Dragana Popovic

Air pollution has been a major transboundary problem and a matter of global concern for decades. High concentrations of different air pollutants are particularly harmful to large cities residents, where numerous anthropogenic activities strongly influence the quality of air. Although there are many books on the subject, the one in front of you will hopefully fulfill some of the gaps in the area of air quality monitoring and modeling, and be of help to graduate students, professionals and researchers. The book is divided in five sections, dealing with mathematical models and computing techniques used in air pollution monitoring and forecasting; air pollution models and application; measuring methodologies in air pollution monitoring and control; experimental data on urban air pollution in China, Egypt, Northeastern U.S, Brazil and Romania; and finally, the health effects due to exposure to benzene, and on the influence of air pollutants on the acute respiratory diseases in children in Mexico.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds