Energy Sources, Part B, 1:127–136, 2006

ISSN: 1556-7249 print/1556-7257 online
DOI: 10.1080/009083190881526

Forecasting Ambient Air SO_2 Concentrations Using Artificial Neural Networks

SAIT C. SOFUOGLU
AYSUN SOFUOGLU

Izmir Institute of Technology
Department of Chemical Engineering
Gülbahçe Köyü, Urla, Izmir, Turkey

SAVAS BIRGILI

Izmir Institute of Technology
Environmental Engineering MS Program
Gülbahçe Köyü, Urla, Izmir, Turkey

GOKMEN TAYFUR

Izmir Institute of Technology
Department of Civil Engineering
Gülbahçe Köyü, Urla, Izmir, Turkey

An Artificial Neural Networks (ANNs) model is constructed to forecast SO_2 concentrations in Izmir air. The model uses meteorological variables (wind speed and temperature) and measured particulate matter concentrations as input variables. The correlation coefficient between observed and forecasted concentrations is 0.94 for the network that uses all three variables as input parameters. The root mean square error value of the model is 3.60 $\mu g/m^3$. Considering the limited number of available input variables, model performances show that ANNs are a promising method of modeling to forecast ambient air SO_2 concentrations in Izmir.

Keywords air pollution, SO_2, forecasting, artificial neural networks

A number of methods have been applied to model air pollutant concentrations, which include theoretical and empirical models. Empirical models have found more use due to a high number of variables that have a determining effect on the concentrations. Among those, linear and multiple linear regression analyses are the most used methods. However, regression methods are thought to have limited computational efficiency and

The authors thank Dr. Fikret Inal for his valuable comments. The Greater City Municipality of Izmir is gratefully acknowledged for supplying the air quality data.
Address correspondence to Sait C. Sofuoglu, Izmir Institute of Technology, Department of Chemical Engineering, Gülbahçe Köyü, Urla, 35430 Izmir, Turkey. E-mail: cemilsofuoglu@iyte.edu.tr

generalization ability. A relatively new method, Artificial Neural Networks (ANNs), has been compared with and found to be preferable to common statistical methods (Comrie, 1997; Gardner and Dorling, 1998) and is used considerably in modeling atmospheric concentrations of NO_x (Gardner and Dorling, 1999; Perez and Trier, 2001), ozone (Abdul-Wahab and Al-Alawi, 2002; Gardner and Dorling, 2000; Jorquera et al., 1998; Soja and Soja, 1999; Spellman, 1999; Wang et al., 2003), and benzene (Viotti et al., 2002). Sulfur dioxide (SO_2) (Boznar et al., 1993; Chelani et al., 2002; Hecq et al., 1994; Lu et al., 2003; Mok and Tam, 1998; Perez, 2001) and Particulate Matter (PM) (Perez et al., 2000; Perez and Reyes, 2002) are frequently studied constituents.

Competent models are needed for short-term forecasting of air pollutant concentrations to take preventive and evasive action during pollution episodes in populated and industrialized areas to protect people's health (Kolehmainen et al., 2001). Izmir, Turkey is such a metropolitan city where the population reaches about four million, and a number of industries are located just outside the metropolitan area. The fraction of emissions that come from industrial activities in the city center is estimated to be 91% (Elbir, 2003). Elbir (2003) recently modeled SO_2 concentrations in Izmir air with the CALMET-CALPUFF dispersion model with an accuracy of about 68%. This model requires extensive surface meteorological data and upper air meteorological data along with source characteristics and geophysical data. ANN has not been used for modeling ambient air pollutant concentrations in the city of Izmir. This study aims to construct an ANN model to forecast ambient air SO_2 concentrations in Izmir.

Artificial Neural Networks Modeling

ANNs are interconnected parallel systems consisting of simple processing elements, neurons. Mathematically, an ANN is often viewed as a universal approximator. In contrast to the traditional model-based methods, ANNs are data driven self-adaptive methods. They learn from examples and capture suitable functional relationships among the data even if the underlying relationships are unknown or hard to describe. The ability to identify a relationship from given patterns makes it possible for ANNs to solve large-scale complex problems such as pattern recognition, nonlinear modeling, and control. Consequently, ANNs have found wide variety of applications in many different fields of business, industry, and science.

In this study, a multilayer feed-forward type of ANN was considered to forecast ambient air SO_2 concentrations in Izmir. In a feed-forward network, the input quantities are fed into input layer neurons, which in turn pass them onto hidden layer neurons after multiplying by a weight. The weights are adaptive coefficients within the network that determine the intensity of the input signal (Nelson and Illingworth, 1994). A hidden layer neuron adds up the weighted input received from each input neuron, associates it with a bias, and then passes the result on through a nonlinear transfer function. The output neurons do the same operation as that of a hidden neuron. The bias neurons are connected to the all neurons in the next hidden and output layer neurons to improve the convergence properties of the network. Each bias neuron is assigned a constant random number.

The network is first trained before its application to any problem. In the training process, the target output at each output neuron is compared with the network output, and the difference or error is minimized by adjusting the weights and biases through a training algorithm. Training in ANNs consists of three elements: (1) weights between neurons that define the relative importance of the inputs, (2) a transfer function that controls the generation of the output from a neuron, and (3) learning laws that describe how the adjustments of the weights are made during training. During training, a neuron receives inputs from a previous layer, weighs each input with a prearranged value, and combines these weighted inputs. The combination of the weighted inputs is represented as

$$net_j = \sum x_i v_{ij} \tag{1}$$

where net_j = the summation of the weighted input for the jth neuron; x_i = the input from the ith neuron to the jth neuron; and v_{ij} = the weight from the ith neuron in the previous layer to the jth neuron in the current layer.

The net_j is passed through a transfer function to determine the level of activation. If the activation of a neuron is strong enough, it produces an output that is sent as an input to other neurons in the successive layer. In this study, sigmoid and hyperbolic tangent functions were employed as activation functions in the training of the network. The sigmoid function is a bounded, monotonic, nondecreasing function that provides a graded, nonlinear response between 0 and 1. This function enables a network to map any nonlinear process. One of the main reasons that the sigmoid function was employed is because of its simplicity of its derivative that is required during the training process. The sigmoid function is expressed as

$$f(net_j) = \frac{1}{1 + e^{-net_j}} \tag{2}$$

Alternatively, a hyperbolic tangent transfer function can be used in multilayer networks. The upper and lower limits of this function are [−1, 1], and its derivative is also continuous (Fu, 1994). The mathematical expression of the hyperbolic tangent function is

$$f(net_j) = \frac{2}{1 + e^{-2net_j}} - 1 \tag{3}$$

Due to the upper and lower limits of the activation functions, the linear normalization method was performed for the input-output data so they are in the same range required by the corresponding activation function used. The following expressions were used for the data normalization (Zurada, 1992):

$$\text{Normalization to } [0, 1] \qquad x_n = \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right) \tag{4}$$

$$\text{Normalization to } [-1, 1] \qquad x_n = 2^* \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right) - 1 \tag{5}$$

where x_n is the normalized value and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of x, respectively.

The networks' ability to change weights is called learning (Engel, 2001). The learning of ANNs was accomplished by a back-propagation algorithm. Back-propagation is the most commonly used supervised training algorithm in the multilayer feed-forward networks. In back-propagation networks, information is processed in the forward direction from the input layer to the hidden layer and then to the output layer. The objective of a back propagation network is, by minimizing a predetermined error function, to find the optimal weights that would generate an output vector $\boldsymbol{Y} = (y_1, y_2, \ldots, y_p)$ as close as possible to target values of output vector $\boldsymbol{T} = (t_1, t_2, \ldots, t_p)$ with a selected accuracy. A predetermined error function has the following form:

$$E = \sum_{P} \sum_{p} (y_i - t_i)^2 \tag{6}$$

where y_i = the component of an ANN output vector $\boldsymbol{Y}$; t_i = the component of a target output vector $\boldsymbol{T}$; p = the number of output neurons; and P = the number of training patterns.

The least square error method, along with a generalized delta rule, was used to optimize the network weights. The gradient descent method, along with the chain rule of derivatives, was employed to modify network weights as

$$v_{ij}^{\text{new}} = v_{ij}^{\text{old}} - \delta \frac{\partial E}{\partial v_{ij}} \tag{7}$$

where δ = the learning rate that is used to increase the chance of avoiding the training process being trapped in local minima instead of global minima. All the neural network calculations in this study were performed using MATLAB 6.0 with Neural Network Toolbox. The learning rate was constant and equal to 0.02 in these calculations.

In this study, the data were separated into two groups: a training set (the first 197 daily values) and a testing set (the last 76 daily values). The training data set was used for the ANN model development; the accuracy of the modeled concentrations was examined by using the test data set. Two topologies, one-hidden layer and two-hidden layer, were experimented. The performance of the networks was evaluated using two criteria. The first one is the Coefficient of Determination, R^2 value (correlation coefficient); it denotes the level of correlation between the observed and forecasted concentrations, and it is preferred due to comparability with conventional studies. The second performance measure is accuracy of forecasts. Root Mean Square Error (RMSE), one of the most employed measures, is used to denote the forecast accuracy in this study. The best performing network was sought by experimenting combinations of input variables, topologies, and transfer functions, which were evaluated using the two measures.

In addition to the main modeling efforts, a next-day forecast study was performed. The constructed model, for this purpose, uses daily average concentrations of consecutive seven days to forecast the average concentration of the eighth day. Several networks were constructed using different periods of learning data set, e.g., 2 to 10 days. The network that provided the minimum error on the forecasted value was determined as the network that used 7 days' data to forecast day-8's concentration value. Detailed information regarding the application of the ANN to model ambient air pollutant concentrations can be found elsewhere (Birgili, 2002).

SO_2 and PM Concentrations

SO_2 and PM concentrations are monitored by the city of Izmir at four metropolitan sampling locations. Izmir air is affected by both residential and industrial use of fossil fuels. The city center is mostly under the influence of industries located outside the metropolitan area where the sampling network is located. The fraction of emissions that come from industrial activities in the city center is estimated to be 91% of which 93% is from industries located outside the metropolitan area (Elbir, 2003).

SO_2 concentrations were measured with the acidimetric titration method using 8-hour samples. PM concentrations were measured with refractometric technique using 24-hour integrated filter samples. A mean of average daily concentrations of four sampling sites from January 1, 2001, to September 30, 2001, were used in model construction (training set). The use of mean concentration is preferred because meteorological variables were not measured at the pollutant sampling sites. The Department of Meteorological Forecasting measures these variables at a site, the International Airport of Izmir, just outside the metropolitan area. This issue directed the authors to use the mean concentration of the four stations in the sampling network, assuming the measured values of meteorological variables are representative of average values of the city of Izmir. Training set of SO_2 and PM concentrations and temperature and wind speed values are shown in Figure 1. Concentrations of both of the pollutants cover a wide range, of 10–120 $\mu g/m^3$ for SO_2 and 20–200 $\mu g/m^3$ for PM, representing all weather conditions including winter, spring, and summer days, as in the cases of temperature and wind speed values. Observed temperatures, in the training set, are in the range of 1–30°C; wind speeds vary between 15 and 55 km/h.

Results and Discussion

Two different network topologies were constructed to forecast pollutant concentrations. The first network has one hidden layer with three nodes (Figure 2a); the second network has two hidden layers with three nodes in each layer (Figure 2b). The networks shown in Figure 2 employ all three available input parameters. The results of different combinations of input variables together with different transfer functions are presented in Table 1. R^2 values are >0.85 when two or more input variables are used in the constructed networks. Values of R^2 are $>90\%$ in most of the cases (see Table 1).

The best performing network uses temperature, wind speed, and PM concentration as input variables, a hyperbolic tangent function as the transfer function; with two hidden layers (Network 1, $R^2 = 0.94$ and MSE $= 3.60$ $\mu g/m^3$, see Figure 3) and one hidden layer in the network (Network 3, $R^2 = 0.92$ and MSE $= 3.34$ $\mu g/m^3$, see Figure 4). The ANN model forecasts the trend in SO_2 concentration well, and the overall agreement between the forecasted and measured SO_2 levels was satisfactory. Increasing the topology of the networks provides higher correlation coefficient values and lower RMSE values when hyperbolic tangent function is used. Using an increased number of variables leads to better results; this is because the system can generalize data better, and possible errors are distributed through the system, therefore minimizing to produce the least amount of effects. It may also be due to the property that the network with more than one hidden layer is trained faster and more likely to avoid inefficient solutions (Gardner and Dorling, 1999). The sigmoid function yields better performances of up to two input variables based on the R^2-value comparisons. The correlation between observed and forecasted SO_2 concentrations is higher when a hyperbolic tangent function is used with three input variables. Nevertheless, all RMSEs are lower for networks that employ a hyperbolic tangent function compared to a sigmoid function except for the case of one input variable. A disadvantage of the ANN method could be that an increased number of input variables may result in reduced model performances due to system instability. Considerable system instability was not observed in this study because a limited number of variables were used.

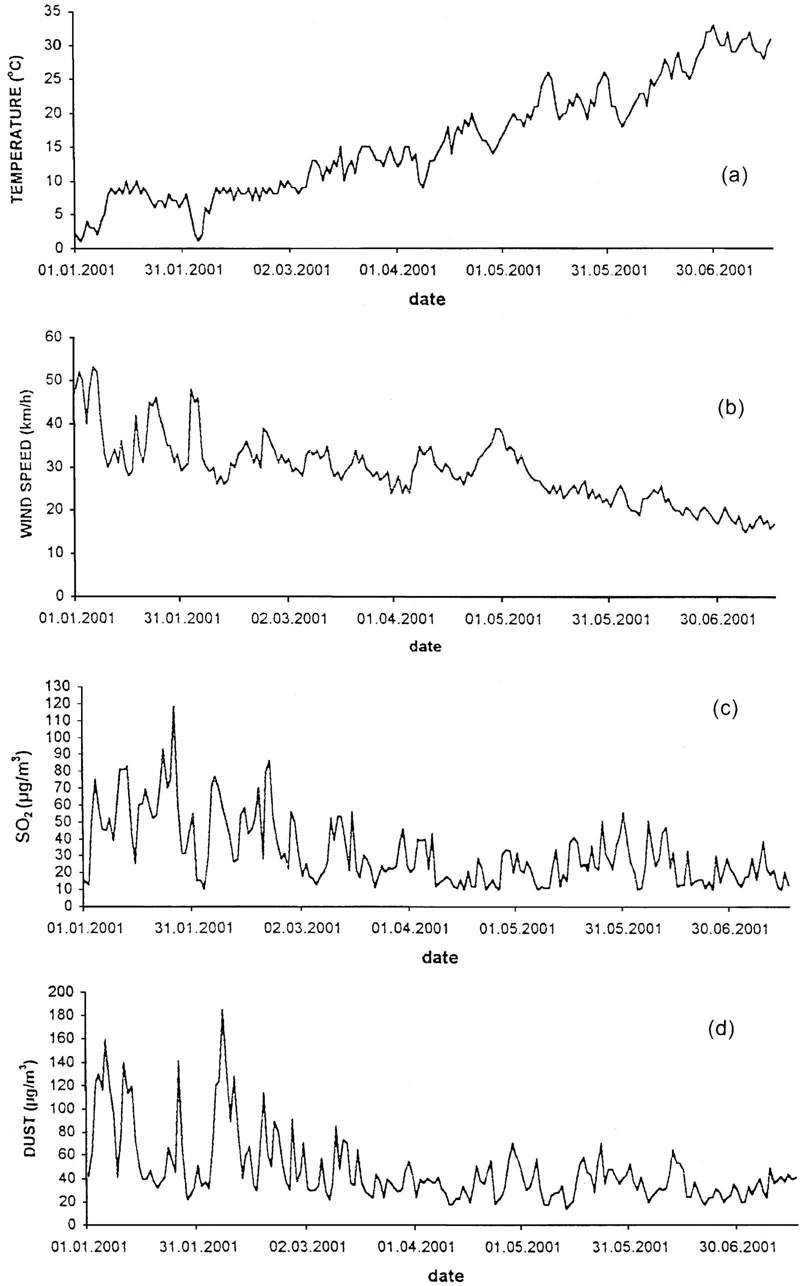

Figure 1. Training set daily average input variable values (a) temperature, (b) wind speed, (c) SO_2 concentration, (d) PM concentration.

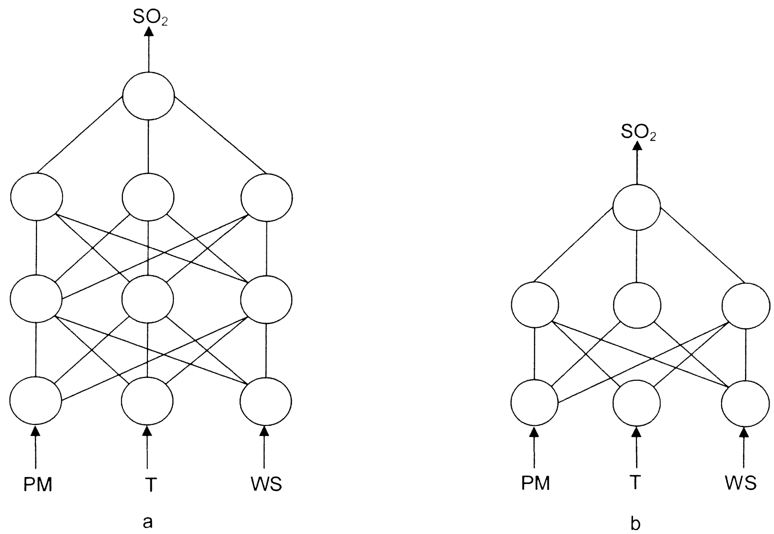

Figure 2. Topology of the networks with (a) two, (b) one hidden layer(s).

Table 1
Performance of the ANN model

Input variables	No. of hidden layers and nodes in the layers	Transfer function	Network no.	R^2	RMSE (μg/m^3)
PM	2 (3-3-3-1)	HTF	1	0.94	3.60
Temperature		SF	2	0.93	3.98
Wind speed	1 (3-3-1)	HTF	3	0.92	3.34
		SF	4	0.90	4.08
Temperature	2 (2-3-3-1)	HTF	5	0.92	3.84
Wind speed		SF	6	0.88	4.80
	1 (2-3-1)	HTF	7	0.90	4.19
		SF	8	0.91	4.77
PM	2 (2-3-3-1)	HTF	9	0.89	4.36
Wind speed		SF	10	0.91	4.17
	1 (2-3-1)	HTF	11	0.86	4.70
		SF	12	0.89	5.01
PM	2 (2-3-3-1)	HTF	13	0.90	4.51
Temperature		SF	14	0.87	4.75
	1 (2-3-1)	HTF	15	0.89	4.79
		SF	16	0.88	4.99
PM	2 (3-3-1)	HTF	17	0.80	5.21
		SF	18	0.83	4.77
	1 (3-1)	HTF	19	0.76	5.57
		SF	20	0.82	5.26

HTF: Hyperbolic Tangent Function, SF: Sigmoid Function, PM: Particulate Matter, RMSE: Root Mean Square Error, R^2: Correlation coefficient between observed and forecasted values.

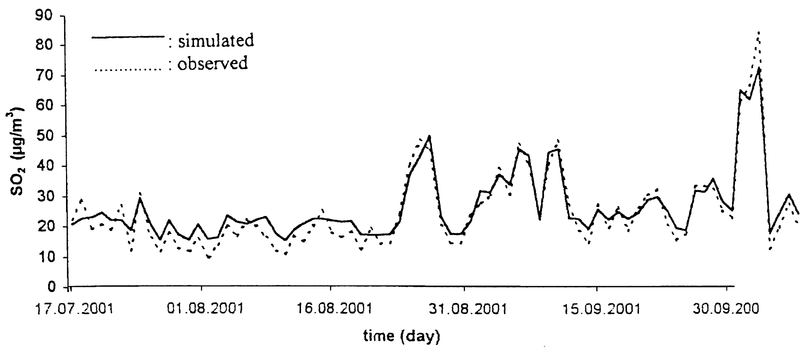

Figure 3. Forecasted SO_2 concentrations with three input variables and two hidden layers.

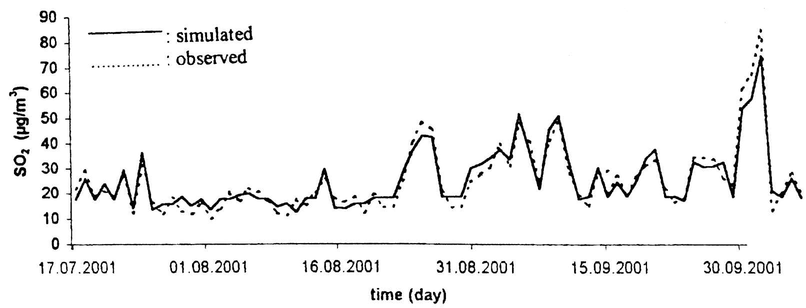

Figure 4. Forecasted SO_2 concentrations with three input variables and one hidden layer.

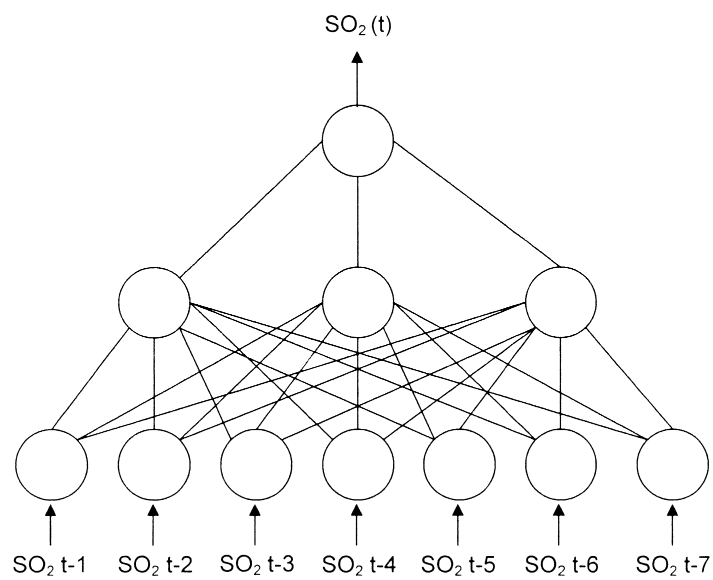

Figure 5. Topology of the network for forecasting the eighth-day concentrations.

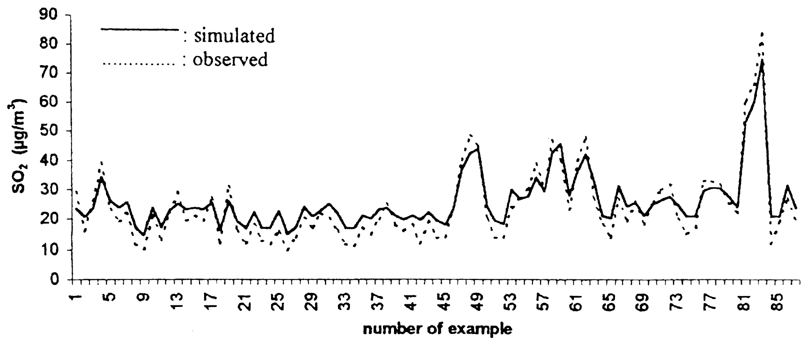

Figure 6. The eighth-day SO_2 forecasts.

Next-Day Forecasts

The performance of the ANNs was also investigated in forecasting the next-day ambient air SO_2 concentrations using previously measured concentration data. The best performing topology-transfer function combination was determined by applying a trial-and-error experimentation. Combinations of one to ten days and the two transfer functions were tried. Results showed that concentration data of a consecutive seven days (the network topology shown in Figure 5) produces the closest next-day, the eighth-day, forecasts with the use of a sigmoid function as the transfer function ($R^2 = 0.94$ and RMSE = 4.03 μg/m^3, see Figure 6).

Conclusions

ANN modeling is shown to be a successful method to forecast ambient air SO_2 concentrations in the city of Izmir. Correlation coefficients between forecasted and observed concentrations of both the main and the next-day networks are >0.90, and the RMSE values are <4.0 μg/m^3. Better model performances were obtained with the hyperbolic tangent function and two hidden layered topologies. Back propagation is found to be a suitable method for network training in modeling ambient air pollutant concentrations. The performance of the ANN model in forecasting the SO_2 concentrations in Izmir is superior to the conventional models. Future study needs include the incorporation of a number of variables such as wind direction, emission sources, etc. (because wind speed and temperature are not the sole determinants of ambient air SO_2 concentrations) and the incorporation of a sensitivity analysis.

References

Abdul-Wahab, S. A., and Al-Alawi, S. M. 2002. Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. *Environ. Model. Software* 17:219–228.

Birgili, S. 2002. Artificial Neural Networks model for air quality in the region of Izmir. MSc Thesis, Izmir Institute of Technology.

Boznar, M., Lesjak, M., and Mlakar, P. 1993. A neural network based method for short term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. *Atmos. Environ.* 27B:221–230.

Chelani, A. B., Chalapati Rao, C. V., Phadke, M. Z., and Hasan, M. Z. 2002. Prediction of sulphur dioxide concentration using artificial neural networks. *Environ. Model. Software* 17:161–168.

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

Elbir, T. 2003. Comparison of model predictions with the data of an urban air quality monitoring network in Izmir, Turkey. *Atmos. Environ.* 37:2149–2157.

Engel, A. 2001. Complexity of learning in artificial neural networks. *Theor. Comp. Sci.* 265:285–306.

Fu, L. 1994. *Neural networks in computer intelligence.* New York: McGraw-Hill.

Gardner, M. W., and Dorling, S. R. 1998. Artificial neural networks (the multi-layer perceptron)—A review of applications in atmospheric sciences. *Atmos. Environ.* 32:2627–2636.

Gardner, M. W., and Dorling, S. R. 1999. Neural network modeling and prediction of hourly NO_x and NO_2 concentrations in urban air in London. *Atmos. Environ.* 33:709–719.

Gardner, M. W., and Dorling, S. R. 2000. Statistical surface ozone models: An improved methodology to account for nonlinear behaviour. *Atmos. Environ.* 34:21–34.

Hecq, W., Borisov, Y., Debever, C., and Dupierreux, J.-M. 1994. An empirical hybrid model for the assessment of daily SO_2 concentrations in an urban environment. *J. Environ. Manage.* 42: 181–198.

Jorquera, H., Pérez, R., Cipriano, A., Espejo, A., Letelier, M. V., and Acuña, G. 1998. Forecasting ozone daily maximum levels at Santiago, Chile. *Atmos. Environ.* 32:3415–3424.

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

Lu, W. Z., Fan, H. Y., and Lo, S. M. 2003. Application of revolutionary neural network method in predicting pollutant levels in downtown area of Hong Kong. *Neurocomputing* 51:387–400.

Mok, K. M., and Tam, S. C. 1998. Short-term prediction of SO_2 concentration in Macau with artificial neural networks. *Energy Build.* 28:279–286.

Nelson, M. M., and Illingworth, W. T. 1994. *A practical guide to neural nets.* Reading, MA: Addison Wesley.

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

Perez, P., and Reyes, J. 2002. Prediction of maximum of 24-h average of PM10 concentrations 30h in advance in Santiago, Chile. *Atmos. Environ.* 34:4555–4561.

Perez, P., and Trier, A. 2001. Prediction of NO and NO_2 concentrations near a street with heavy traffic in Santiago, Chile. *Atmos. Environ.* 35:1783–1789.

Perez, P., Trier, A., and Reyes, J. 2000. Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago, Chile. *Atmos. Environ.* 34:1189–1196.

Soja, G., and Soja, A.-M. 1999. Ozone indices based on simple meteorological parameters: Potentials and limitations of regression and neural network models. *Atmos. Environ.* 33:4299–4307.

Spellman, G. 1999. An application of artificial neural networks to the prediction of surface ozone concentrations in the United Kingdom. *Appl. Geogr.* 19:123–136.

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

Wang, W., Lu, W., Wang, X., and Leung, A. Y. T. 2003. Prediction of maximum daily ozone level using combined neural network and statistical characteristics. *Environ. Int.* 29:555–562.

Zurada, J. M. 1992. *Introduction to artificial neural systems.* New York: West Publishing.