

Prediction of Air Pollutants by Using an Artificial Neural Network

Sang Hyun Sohn, Seo Cheon Oh and Yeong-Koo Yeo[†]

Dept. of Chem. Eng., Hanyang University, Seoul 133-791, Korea

(Received 17 September 1998 • accepted 9 March 1999)

Abstract—The purpose of this study is to predict the amount of primary air pollution substances in Seoul, Korea. An artificial neural network (ANN) was used as a prediction method. The ANN with three layers is learned with past data, and the concentrations of air pollutants are predicted based on the pre-learned weights. The error back propagation method that has a powerful application to various fields was adopted as the learning rule. The concentrations of air pollutants from one to six hours in the future were predicted with the ANN. To verify the performance of the prediction method used in the present study, the predicted concentrations of air pollutants were compared with the measured data. From the comparison, it was found that the prediction method based on the ANN gives an acceptable accuracy for the limited prediction horizon.

Key words : Air Pollutants, Prediction, Artificial Neural Network, Error Back Propagation

INTRODUCTION

Air pollutants including nitrogen oxides (NO_x), sulfur oxides (SO_x) and hydrocarbons are puffing up from various sources, and their effects on human living conditions become a serious problem. For example, SO_x causes an eye, nose and throat irritant and has been correlated with respiratory illnesses [Koenig et al., 1982]. Especially, sulfur dioxide (SO_2) causes loss of chlorophyll in green plants. Most SO_x are generated during combustion of sulfur-containing fuels and emitted from industrial processes that consume sulfur-containing raw materials [Cooper and Alley, 1994]. Among various NO_x , the most important air pollutants are nitric oxide (NO) and nitrogen dioxide (NO_2) because they are emitted in large quantities. The mechanism of NO_x production has been studied extensively [Ammann and Timmins, 1966; Fenimore, 1971; Duterque et al., 1981]. Carbon monoxide (CO) is the most abundant air pollutant in the lower atmosphere. Since the principal source of CO in urban areas is motor vehicle exhaust, CO concentrations correlate closely with traffic volume.

Photochemical smog is the particular mixture of reactants and products generated when hydrocarbons and nitrogen oxides exist together in the presence of sunlight. Ozone, which is a very strong oxidant, is mainly formed by photochemical synthesis with NO_x and hydrocarbons. Ozone is not emitted directly by sources and is usually formed in the atmosphere by chemical reactions. For the analysis of the ozone formation in air pollution, an understanding of the various photochemical processes taking place in the atmosphere is very important. In order to characterize the ozone formation in air pollution much effort [Carter et al., 1979a, b; Sakamaki et al., 1982; Fan et al., 1996; Oh and Yeo, 1998] has been devoted to experimentally inves-

tigating and simulating air pollution chemistry.

Air pollutants released from various sources affect, directly or indirectly, the health of human beings and animals and do damage to plants. Therefore, an accurate estimation of the resulting ground level concentration patterns according to various air quality is very important for social planning and industrial growth. It is of great interest to determine whether computer simulation can produce reasonable accuracy in the prediction results of air pollutant formation in the atmosphere. The mass conservation equation represents the natural phenomenon of transport, source, and sinking terms such as emissions, chemistry and removal at the surface, so it has been used in most urban and regional photochemical models [Carmichael et al., 1986; Chang et al., 1987; Venkatram et al., 1992; Scheffe and Morris, 1993]. However, although the mass conservation equation can be used as a data analysis tool of air quality, it does not yet give a reasonable accuracy on the multi-period ahead prediction due to the deviations in the wind field, diffusion and chemistry by natural fluctuations. Furthermore, their natural fluctuations contribute to the observed variation in the frequency and the intensity of episodes in different geographic locations and at different times of the year. Real time parameter estimation has also been used to predict air pollutants in the atmosphere [Oh et al., 1999]. However, since most of these parameter estimations are based on linear models, they cannot handle nonlinear situations that arise in the atmosphere. An artificial neural network (ANN) has a prominent ability in the recognition of nonlinear patterns. Roadknight et al. [1997] used artificial neural networks to model the interactions that occur between ozone pollution and crop damage. In this study, the ANN method was used to predict the concentration of air pollutants. The error back propagation (EBP) method was adopted as the learning rule. To verify the performance of the prediction method used in this study, we compared the results of predictions of the concentration of air pollutants with the measured data.

[†]To whom correspondence should be addressed.

E-mail : ykyeo@email.hanyang.ac.kr

DESIGN OF ARTIFICIAL NEURAL NETWORK

Artificial neural networks have been widely used in modeling, control, pattern recognition, signal processing, prediction and so on [Zurada, 1995]. A neural network is a group of processing elements where one typical subgroup makes independent computations and passes the results to the next subgroup that may in turn perform independent computations and pass on the results to a subsequent subgroup. Finally, a subgroup of one or more processing elements determines the output from the network. Each processing element performs computation based upon a weighted sum of inputs. A subgroup of processing elements is called a layer in the network. The first layer is the input layer and the last one is the output layer, and the layers in between the first and the last layers are the hidden or intermediate layers. The general structure of an artificial neural network is well known and can be found in numerous literatures. Basically, the learning in the network is achieved through an iterative algorithm that minimizes the mean square errors between the desired and actual outputs. It has been shown that networks of this form can map any set of data [Funahashi, 1989; Sprecher, 1993]. Fig. 1 shows the basic structure of the three-layer network used in this study, which consists of an input layer, output layer and a hidden layer. In this structure there are i inputs and one threshold in the input layer, j neurons and one threshold in the hidden layer and k neurons in the output layer. Also, each of the neurons is associated with a bias weight. Inside each neuron a weighted sum of the inputs is calculated, a bias weight is added, and this value, called "net", is transformed by a bipolar sigmoid function. The transformed result is sent to neurons in the next layer. Usually the sigmoid function with range of $-1 \rightarrow +1$ is defined as

$$\text{Bipolar Sigmoid Function } F(\text{net}) \stackrel{\Delta}{=} \frac{2}{1+\exp(-\lambda \cdot \text{net})} - 1 \quad (1)$$

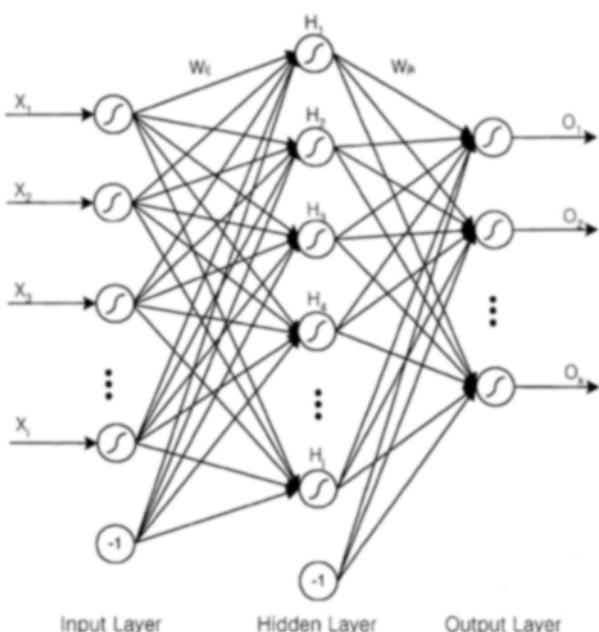


Fig. 1. Neural network with three-layer feedforward structure.

where λ is proportional to the neuron gain determining the steepness of the continuous function f (net) near net=0 [Zurada, 1995].

The neural network employed in this study has an input layer to which previous and current data such as time and concentrations of ozone, SO_2 , NO_2 , NO , total hydrocarbon (THC), CO and CH_4 are fed. The number of neurons in the hidden layer can be chosen to provide a sufficiently good fit. In this ANN, the number of hidden nodes affects the speed of learning and the convergence of errors. Too many hidden nodes make the system so slow and the insufficiency of that do it not converge the prefixed error. In this study we chose 30 hidden neurons. The number of neurons in the output layer is fixed by the number of outputs predicted by the network. The connection weights are computed during the training process. The error back propagation (EBP) method, which is the prevailing learning algorithm, was used to train the network [Jeong and Lee, 1993]. During the training process initial weights are assigned to the connections randomly. Inputs entered through the input layer are propagated forward through the hidden layer of neurons until they reach the output layer. The outputs generated by neural network are compared with the measured data. The errors between the predicted and the actual output values are reduced by changing the weights according to

new weight change

$$\eta \times \text{output error} \times \text{input} + \alpha \times \text{previous weight change} \quad (2)$$

where η is learning rate and α is momentum.

This process is repeated until some predefined stopping criteria are satisfied. When the learning is complete, the neural network is used for prediction.

A completely trained neural network can be thought of as an approximating function $P(V, W, X)$ to the actual function $P(X)$, relating the input-output mapping. X is the vector of input-output pairs and V and W are the weight matrices that give the best fit. These approximating functions can be written as

$$O_k = F(\sum_j W_{jk} H_j + \Phi_k) \quad (3)$$

and

$$H_j = F(\sum_i V_{ij} X_i + \Phi_j) \quad (4)$$

In Eqs. (3) and (4), F is the sigmoid function, X_i are the linearly scaled inputs, O_k are the linearly scaled outputs, H_j are the outputs from the neurons in the hidden layer, W_{jk} are the weights corresponding to the connections between the hidden neurons and the output neurons, V_{ij} are the weights corresponding to the connections between the input nodes and hidden neurons and Φ_j and Φ_k are the bias weights for the neurons i of the hidden and output layers respectively. In this ANN, the number of hidden nodes affects the speed of learning and the convergence of errors. Too many hidden nodes make the system very slow and the insufficiency of that do it not converge the prefixed error. The neural networks are trained in the scaled domain to get a uniform distribution of data

within the data space. To predict the outputs, the given inputs are scaled linearly with minimum and maximum values in the $-1 \sim +1$, and the approximating functions are evaluated to obtain the scaled outputs.

PREDICTIONS USING SLIDING WINDOW LEARNING

In this study, to increase the accuracy of the predictions of this system, the sliding window learning procedure was used. The procedure results in an adaptive neural model that is updated at each sampling instant. This procedure is inspired by the recursive parameter estimation techniques that are widely used in identification and control. The adaptation scheme restricts the memory of the neural network by adding the effects of new data and by progressively removing the influence of obsolete data.

Sliding window learning procedure used in this study is as follows :

Step 1. Obtain an initial neural model from previous experiments or simulated data.

Step 2. Choose the length L of the learning window.

For each sampling instant τ greater than or equal to L

Step 3. Form a new learning data set with the L successive pairs of input and output data vectors corresponding to the sampling instants $(\tau-L+1)$ to τ .

Step 4. Teach the network with the L newly formed learning data set to update the weights of the current neural model.

The learning procedure consists of two successive steps with an updated learning data set and the neural model updated by the learning algorithm. Step 4 goes on for a maximum number of iterations or until the desired convergence is achieved.

At each sampling instant τ , prediction is performed using the current concentrations of air pollutants, with the updated neural model if τ is greater than or equal to L , or with the initial neural model otherwise. Fig. 2 illustrates schematically the prediction procedure used in this study at sampling instant τ . The learning data set is constituted by the L successive pairs of input and output data obtained at sampling instants $(\tau-L+1)$ to τ and the on-line prediction is performed with the neural model updated by this data set until new real data are obtained. Real data are hourly obtained. In this study, learning sets are composed of each 24 and 30 hour set. The prediction results can be obtained by successive substitutions. This is the method that receives output data as next time input data, as time goes by, and predicts the next time value with the former value successively. For example, it predicts $(t+1)$ time value using (t) time value, and predicts $(t+2)$ time value using that of $(t+1)$ time. In general, it is necessary to upgrade the number of newly updated data set according to the prediction periods. Newly updated data are transformed into input and desired output data pairs. In this way we can make the prediction system easily solvable as a closed system. In the prediction, the input data set must be prepared within at least a 24 hour period because ozone pattern

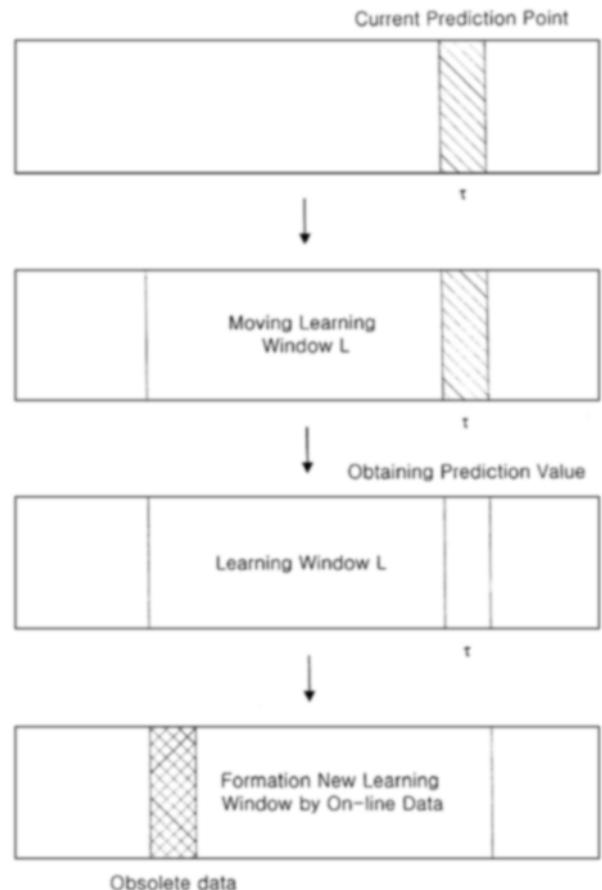


Fig. 2. Formation schemes of learning data set.

exhibits cyclic behavior of 24 hours period. Scaling is a very important technique that is a pretreated process, in that the prediction and input and output values are transformed into $-1 \sim +1$ in the present study. The way is simple. That is $[1 - (-1)] \cdot (\text{Input} - \text{Minimum value}) / (\text{Maximum value} - \text{Minimum value}) + (-1)$. Here, it is required to assume a large enough range of pollutant concentrations because the input values change as time varies. A false assumption of the range of the concentrations causes false results. The output data by ANN are transformed into the original range of each concentration following reverse scaling procedure.

RESULTS AND DISCUSSION

The measured data of the air pollutants concentrations were provided by the Korean Environmental Office for May 1-15, 1996. Comparisons between measured data and results of prediction were performed based on the data during May 11-13, 1996 (Fig. 3-Fig. 6) and May 5-8, 1996 (Fig. 7). Input data are composed of time and the concentration of air pollutants that consist of ozone, NO_2 , SO_2 , THC, CH_4 , CO and NO. For the measured data used in this study, ozone varies between 0 and 0.09, SO_2 between 0 and 0.03, NO, between 0 and 0.12, CO between 0 and 2, THC between 0 and 0.04, NO between 0 and 0.25 and CH_4 between 0 and 0.038. Overall, 358 patterns during 15 days are obtained. These data sets are used for training and testing the neural network models. In this study

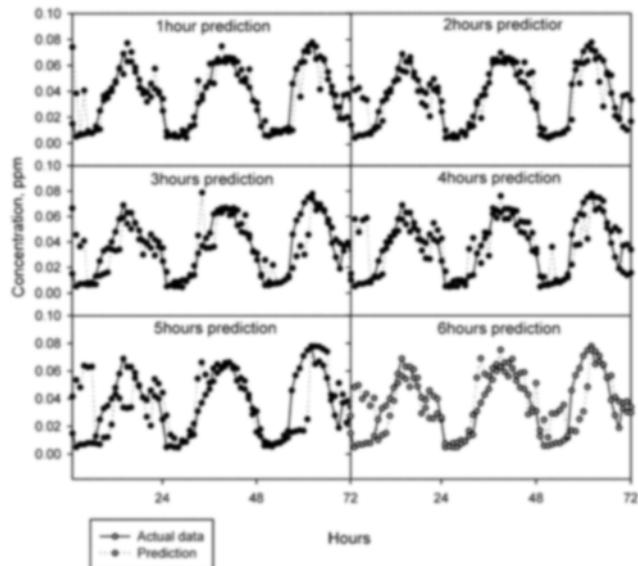


Fig. 3. Ozone prediction with 24 hours learning data set.
(error limit in learning=0.05)

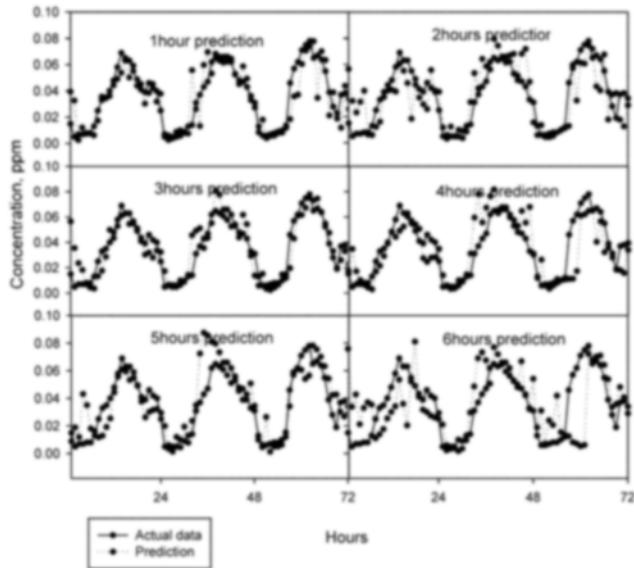


Fig. 4. Ozone prediction with 30 hours learning data set.
(error limit in learning=0.08)

A back-propagation algorithm is used for the training of the network. This algorithm takes 24 and 30 patterns as training sets with $\lambda=1$, $\eta=0.5$ and $\alpha=0.5$ in Eqs. (1) and (2). The weight matrix obtained by the training is stored to predict the air pollutants concentrations. Desired output data are then used as new input data and prediction value of next time can be obtained using ANN with these weights. Table 1 shows the index of 16 input and 7 output nodes used in this study. A neural network containing 30 neurons in the hidden layer is adopted. Fig. 3 shows ozone concentration according to prediction time, where an error limit of 0.05 in learning is employed. Fig. 4 shows the prediction results with various time horizons based on 30 hours learning sets. In this case an error limit of 0.08 in learning is employed. As can be seen, the prediction method proposed in this study gives a reason-

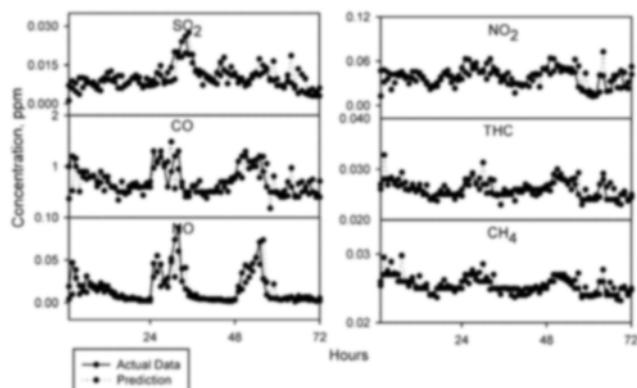


Fig. 5. One hour prediction of other pollutants concentration with 24 hours learning data set.
(error limit in learning=0.05)

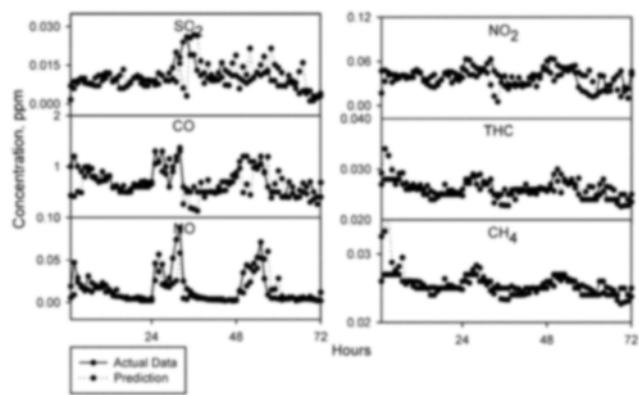


Fig. 6. 3 hours prediction of other pollutants concentration with 24 hour learning data set.
(error limit in learning=0.05)

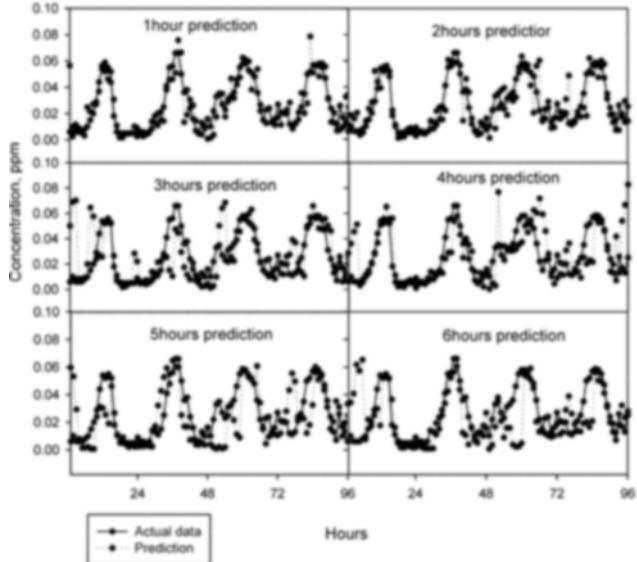


Fig. 7. Ozone prediction with 24 hours learning data set.
(error limit=0.05)

able accuracy on the limited prediction horizon. But as the prediction periods increase, the prediction errors in Fig. 3 and Fig. 4 also increase. For short-term prediction horizons, it can be seen that the prediction results based on a 24 hour learn-

Table 1. Input and output variables

Input variables		Output variable	
Feature	Description	Feature	Description
X ₁	Time (t)	O ₁	Ozone (t+1)
X ₂	Time (t-1)	O ₂	SO ₂ (t+1)
X ₃	Ozone (t)	O ₃	NO ₂ (t+1)
X ₄	Ozone (t-1)	O ₄	CO (t+1)
X ₅	SO ₂ (t)	O ₅	THC (t+1)
X ₆	SO ₂ (t-1)	O ₆	NO (t+1)
X ₇	NO ₂ (t)	O ₇	CH ₄ (t+1)
X ₈	NO ₂ (t-1)		
X ₉	CO (t)		
X ₁₀	CO (t-1)		
X ₁₁	THC (t)		
X ₁₂	THC (t-1)		
X ₁₃	NO (t)		
X ₁₄	NO (t-1)		
X ₁₅	CH ₄ (t)		
X ₁₆	CH ₄ (t-1)		

ing data set is better than those based on a 30 hour learning data set; thus this shows that use of large learning data sets does not always mean better results. Fig. 5 and Fig. 6 show the 1 hour and 3 hour prediction results of the concentration of other pollutants based on a 24 hour learning data set, respectively. From these results, we can see that 1 hour predictions show better results while 3 hours prediction results are inappropriate. This is due to the fact that the learning process of ANN is achieved by pattern recognition but the other pollutants excluding ozone are emitted directly by source and do not have strict patterns. This causes prediction errors of ozone concentrations for multi-step prediction horizons, because ozone is directly affected by NO₂ and hydrocarbons [Seinfeld, 1986]. Thus, a prediction method needs to be developed of other air pollutants that are related to ozone formation. Fig. 7 shows the prediction results of ozone concentration of other days to verify the performance of the prediction method used in this study. It was found that the prediction method based on ANN gives a reasonable accuracy for the limited prediction horizons, but for more precise predictions, it is necessary for the ANN to be obtained with input data related to weather and sunlight.

CONCLUSION

Predictions of air pollutant concentration by using ANN were performed. The prediction results were compared with the actually measured data. Usually, it is very difficult to express ozone because it is a strong oxidant that is very sensitive material according to various surrounding conditions like sunlight, wind direction and velocity, temperature, humidity and chemical composition of air. This study indicates that the use of ANN is useful to represent ozone as well as other air pollutants. The prediction errors sometimes showed stiff behavior, because the input data are composed of only the concentration of air pollutants. However, it was found that the ANN prediction meth-

od developed in the present study is useful to predict air pollutants. For the selection of learning window sets it is necessary to consider the convergence of learning window set in ANN because too long learning window sets prevent effective ANN learning. In further study, it would be required to add weather conditions so that the prediction of ozone concentrations may fit well at any time, and we will consider other ways of using spatial and meteorological conceptions. Also, in data preprocessing, it would be good to consider the method that learns differences of each training data.

ACKNOWLEDGEMENT

This work was supported in part by the Korea Science and Engineering Foundation [KOSEF] through the Automation Research Center at POSTECH.

REFERENCES

Ammann, P. R. and Timmins, R. S., "Chemical Reactions During Rapid Quenching of Oxygen Nitrogen Mixture from Very High Temperature," *AIChE J.*, **12**, 956 (1966).

Carmichael, G. R., Peters, L. K. and Kitada, T., "A Second Generation Model for Regional-Scale Transport/Chemistry/Deposition," *Atmos. Environ.*, **20**, 173 (1986).

Carter, W. P. L., Lloyd, A. C., Sprung, J. L. and Pitts, J. N. Jr., "An Experimental Investigation of Chamber-Dependent Radical Sources," *Int. J. Chem. Kinet.*, **11**, 45 (1979a).

Carter, W. P. L., Winer, A. M., Darnal, K. R. and Pitts, J. N. Jr., "Smog Chamber Studies of Temperature Effects in Photochemical Smog," *Environ. Sci. Technol.*, **13**, 1094 (1979b).

Chang, J. S., Brost, R. A., Isaksen, I., Madronich, S., Middleton, P., Stockwell, W. R. and Walcek, C. J., "A Three-Dimensional Eulerian Acid Deposition Model: Physical Concepts and Formulation," *J. Geophys. Res.*, **92**, 14681 (1987).

Cooper, C. D. and Alley, F. C., "Air Pollution Control: A Design Approach," Waveland Press, Inc., Prospect Heights, Illinois (1994).

Duterque, J., Avegard, N. and Borghi, R., "Further Results on Nitrogen Oxides Production in Combustion Zones," *Combustion Science and Technology*, **25**, 85 (1981).

Fan, Z., Kamens, R. M., Zhang, J. and Hu, J., "Ozone-Nitrogen Dioxide-NPAH Heterogeneous Soot Particle Reactions and Modeling NPAH in the Atmosphere," *Environ. Sci. Technol.*, **30**, 2821 (1996).

Fenimore, C. P., "Formation of Nitric Oxide in Premixed Hydrocarbon Flames," 13th Symposium (International) on Combustion, Pittsburgh, 373 (1971).

Funahashi, K., "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks*, **2**, 183 (1989).

Jeong, S. H. and Lee, K. S., "A Study on Interpolating Behavior of Neural Networks for Nonlinear Engineering Problems," *HWACHEON KONGHAK*, **31**, 54 (1993).

Koenig, P. Q., Pierson, W. E., Horike, M. and Frank, R., "Effects of Inhaled Sulfur Dioxide on Pulmonary Function in Healthy Adolescents: Exposure to SO₂ Alone and SO₂+Sodium Chloride Droplet During Rest and Exercises," *Arch. Environ. Health*, **37**, 5 (1982).

Oh, S. C., Sohn, S. H., Yeo, Y. K. and Chang, K. S., "A Study on the Prediction of Ozone Formation in Air Pollution," *Korean J. Chem. Eng.*, **16**, in press (1999).

Oh, S. C. and Yeo, Y. K., "Modeling and Simulation of Ozone Formation from A Propene-Nitrogen Oxide-Wet Air Mixture in A Smog-Chamber," *Korean J. Chem. Eng.*, **15**, 20 (1998).

Roadknight, C. M., Balls, G. R., Mills, G. E. and Palmer-Brown, D., "Modeling Complex Environmental Data," *IEEE Transactions on Neural Networks*, **8**, 852 (1997).

Sakamaki, F., Okuda, M., Akimoto, H. and Yamazaki, H., "Computer Modeling Study of Photochemical Ozone Formation in the Propene-Nitrogen Oxides-Dry Air System. Generalized Maximum Ozone Isopleth," *Environ. Sci. Technol.*, **16**, 45 (1982).

Scheffe, R. D. and Morris, R. E., "A Review of the Development and Application of the Urban Airshed Model," *Atmos. Environ.*, **27B**, 23 (1993).

Seinfeld, J. H., "Atmospheric Chemistry and Physics of Air Pollution," John Wiley & Sons Inc., New York (1986).

Sprecher, D. A., "A Universal Mapping for Kolmogorov's Superposition Theorem," *Neural Networks*, **6**, 1089 (1993).

Venkatram, A., Karamchandani, P., Kuntasal, G., Misra, P. K. and Davies, D. L., "The Development of the Acid Deposition and Oxidant Model (ADOM)," *Environ. Pollut.*, **75**, 189 (1992).

Zurada, J. M., "Introduction to Artificial Neural Systems," PWS Publishing Company, Boston, 33 (1995).