Ekoloji 19, 74, 32-37 (2009)

The Neural Network Modeling of Suspended Particulate Matter with Autoregressive Structure

Mehmet AKTAN[1], Hanefi BAYRAKTAR[2]*

[1] Atatürk University, Faculty of Engineering, Department of Industrial Engineering, 25240 Erzurum-TURKEY

[2] Atatürk University, Faculty of Engineering, Department of Environmental Engineering, 25240 Erzurum-TURKEY

*Corresponding author: hanefib@atauni.edu.tr

Abstract

Pollution sources and emissions, and their interactions with terrain and the atmosphere are necessary in developing appropriate air pollution management plans and action strategies. In this study, we will investigate the relationship between the total suspended particulate matter (TSP) concentration and meteorological parameters such as wind speed and direction, temperature, air pressure, precipitation and relative humidity. The TSP measurements of the past two days, sunshine duration, and sunshine amount in the winter season (November through March) in the city of Erzurum between the years 1990 and 2007 were investigated. The artificial neural network (ANN) models were constructed using a mixed autoregressive relationship to realize the stochastic nature of the TSP levels for each month of the winter season. The impact of wind direction on TSP concentration was introduced to the model by defining two variables. Linear regression models were also constructed to check the performance of the neural networks. The most significant factors affecting the TSP concentration are found to be the TSP level of the previous day (TSP_{t-1}), expected temperature (*temp*), wind speed (*w*), air pressure (*p*), and precipitation (*pc*).

Keywords: Autoregressive, Erzurum, meteorological parameters, neural network, particulate matter.

Yapay Sinir Ağı ile Otoregresif Yapıda Asılı Partiküler Madde Modellemesi

Özet

Kirlilik kaynakları ve emisyonunun, ve bölge ve atmosferle etkileşimlerinin anlaşılması, uygun hava kirliliği yönetimi planları ve aksiyon stratejilerinin geliştirilmesi için gereklidir. Bu çalışmada, Erzurum'da 1990 ve 2007 yılları arasında kış aylarında (Kasım'dan Mart'a kadar) toplam asılı partiküler madde (TSP) konsantrasyonu ile rüzgar hızı ve yönü, sıcaklık, hava basıncı, yağış, bağıl nem, geçmiş iki günün TSP ölçümleri, günışığı süresi ve miktarı gibi meteorolojik parametreler arasındaki ilişki araştırılmıştır. Kış mevsiminin her ayında TSP seviyesinin stokastik davranışını kavramak için karma otoregresif yapıda yapay sinir ağı modelleri geliştirilmiştir. İki değişken tanımlayarak, rüzgar yönünün TSP konsantrasyonu üzerindeki etkisi modele tanıtılmıştır. Yapay sinir ağının performansını ölçmek için doğrusal regresyon modelleri de geliştirilmiştir. TSP konsantrasyonunu en çok etkileyen faktörler, bir gün önceki TSP seviyesi (TSP_{t-1}), beklenen sıcaklık (*temp*), rüzgar hızı (*w*), hava basıncı (*p*), ve yağış (*pc*) olarak belirlenmiştir.

Anahtar kelimeler: Erzurum, meteorolojik parametreler, otoregresif, partiküler madde, sinir ağı,

Aktan M, Bayraktar H (2010) The Neural Network Modeling of Suspended Particulate Matter with Autoregressive Structure. Ekoloji 19, 74, 32-37.

INTRODUCTION

Rapid urbanization in developing countries has led to uncontrolled growth and severe environmental deterioration. In particular, urban air pollution is still on the rise in many cities worldwide, or has experienced only small improvements. Air pollution takes place under the combined effect of meteorological factors, earth surface topographic features and the release of air pollutants from various sources. The meteorological factors such as wind speed and direction, temperature, pressure, humidity and inversion height play an important role in determining pollutant levels for a given rate of pollutant emission (Bayraktar et al. 2005).

The emissions of criteria air pollutants such as sulphur dioxide (SO_2) and total suspended particulate matter (TSP) are health threats. Erzurum, as well as many cities in Turkey, suffers from high ambient concentrations of TSP, especially in the winter seasons. TSP is the general term used for a mixture of solid particles and liquid droplets found in the air. TSP has a wide range of sizes and originates from many different stationary and

Received: 31.03.2009 / Accepted: 07.07.2009

mobile sources. They may be emitted directly from a source or formed in the atmosphere by the transformation of gaseous precursor emissions such as SO_2. Scientific studies show a link between inhaled particles and a series of significant health effects (Liu et al. 2009). In addition, particles cause adverse impacts on the environment by changing the nutrient balance through deposition processes and reduced visibility (Hyslop 2009).

Several mathematical approaches are used for modeling air pollution. Given a set of air monitoring observations, discovering the relationship among variables is possible by using techniques such as neural networks and regression models. Such models have low development costs and resource requirements. There are several studies in literature for statistically determining the effects of meteorological parameters on TSP concentration. TSP concentrations have been related to meteorological factors and some policies have been proposed for the city of Shangai (Chao 1990). Particulate matter dry deposition fluxes were measured in the Uludağ University Campus by Taşdemir and Çağlar (2002). Multiple linear regression analysis has been used to assess the relation of pollutant concentrations with several meteorological factors in a study done by Çuhadaroğlu and Demirci (1997). The results showed that there is a moderate and weak level in the relationship between the pollutant level and meteorological factors in the city of Trabzon. Topçu et al. (1993) investigated the relationship between TSP and meteorological factors such as temperature, pressure, relative humidity, precipitation, and wind speed in Erzurum using linear regression analysis. Air pollution forecasting and control was studied in similar studies. Yıldırım and Bayramoğlu (2006) proposed an adaptive neuro-fuzzy logic method to estimate the SO_2 and TSP trends in the city of Zonguldak. Kurt et al. (2008) developed an online air pollution forecasting system for the Greater Istanbul Area for a user-friendly website (http://airpol.fatih.edu.tr).

This study focuses on examining the severity of the winter season TSP concentration in the city of Erzurum and investigates the relationship between the TSP concentration and meteorological parameters such as wind speed and direction, temperature, pressure, precipitation, sunshine amount, sunshine duration, and relative humidity in the winter season (November through March), which is the time period when severe air pollution episodes are most likely to occur. The TSP concentrations were measured during the period by the Atatürk University Environmental Pollution Research Center. In the Material and Methods section, information about the city of Erzurum is given, measurement of TSP and its evaluation is explained, and construction of the artificial neural networks for each month of the winter season between the years 1990 and 2007 is presented. Next, the results are presented and then concluding remarks are given in the last section.

MATERIAL AND METHODS

Features of the Region: Erzurum is located in the eastern region of Turkey and situated on a plateau surrounded by high mountains to the north, east and south. The altitude and longitude of the Erzurum urban center is 40.0°N and 41.2°E, respectively. The height of the plateau is 1857 meters above sea level and the population of the city was 485107 in 2008 (www.tuik.gov.tr).

Erzurum is one of the coldest cities in Turkey with annual average temperature of 6°C, and the average number of days below 0°C is 162. For that reason, the city needs heating for at least six months of the year. While the average wind speed is above 3.1 m/s in summer, this parameter becomes 2.4 m/s in the winter. Dominant wind direction is WSW-ENE in the city. The climate and topography of the city cause frequent episodes with high concentration of atmospheric pollutants during the winter. Since there are not many industrial facilities in the city, the major source of air pollution is domestic heating, a sugar and cement factory are the two main factories close to the city (NNW), which are 20 km and 50 km away from the city center, respectively.

Measurement of TSP and its evaluation: Daily TSP concentration was measured only in the winter seasons (November, December, January, February, and March) since 1980 at six different stations (by considering the topography of the city) by the Atatürk University Environmental Pollution Research Center. The TSP sampling stations are shown in Fig. 1.

Measurements were made with refractometric evaluations for 24 hours by integrated dust filter samples in accordance with WHO recommended measurement methods (Elbir et al. 2000). The daily average values of the TSP in the city were calculated by using arithmetic averages of the data obtained from the six stations. The daily meteorological data was provided by the Meteorology Department in Erzurum as 8-h average values. The daily average temperature, humidity and wind speed and direction were taken as the integral average of these 8-h values.

Neural Network Model: An artificial neural network (ANN) makes it possible to define the relation (linear or nonlinear) among a number of variables without knowledge of their cause-effect mechanisms. They can act as universal approximations of non-linear functions (Hornik et al. 1989) and consequently, can be used in assessing the dynamics of complex non-linear systems.

In this study, an artificial neural network model was developed for predicting the TSP concentrations in Erzurum during the winter seasons using ten parameters; temperature, precipitation, humidity, pressure, sunshine, wind speed, wind direction, and previous TSP measurements. A total of 2000 records between the years 1990 and 2007 were collected to build the model. Alyuda Neurointelligence software Version 2.2(577) was used to develop and train the ANN model (www.alyuda.com).

A multilayer perceptron (MLP) is composed of multiple neurons (or nodes) arranged in several different layers. The configuration of the best MLP model includes choosing the number of layers (typically, it requires three- input, hidden and output), the number of neurons in the hidden layer, the activation function, the error function and the learning algorithm. After the proper architecture of the MLP has been established, all the training cases are run through the network. In each neuron, a linear combination of the weighted inputs (including a bias) is computed, summed and transformed using a transfer function (linear or nonlinear). The value obtained is transferred as an input to the neurons in the subsequent layer until a value is computed in neurons of the output layer. The output values are compared with the target outputs. The difference between the output and target is calculated using a certain error function in order to give the prediction error made by the network. Then, the training algorithm is used to adjust the network's weights and thresholds in order to minimize this error (Ripley 1996, Haykin 1994).

In this study, the linear transfer function was used for the input neurons, the hyperbolic tangent function was used in the hidden layer neurons, and the logistic function was used in the output layers. Networks were trained using the back-propagation algorithm. Weighted sums of the input components are calculated by using Eq. (1) as

$$u_j = \sum_{i=1}^{n} w_{ij} x_i - b_j \tag{1}$$

where u_j is the weighted sum of the jth neuron for the input received from the preceding layer with n neurons, w_{ij} is the weight between the jth neuron and the ith neuron in the preceding layer, x_i is the output of the ith neuron in the preceding layer, and b_j is the bias of the jth neuron. The output of the jth neuron a_j in a hidden layer is calculated with a hyperbolic tangent function as

$$a_j = f(u_j) = \frac{\exp(u_j) - \exp(-u_j)}{\exp(u_j) + \exp(-u_j)} \tag{2}$$

and the single output $f(u)$ of the output layer is calculated with a logistic function as

$$a = f(u) = \frac{1}{1 + \exp(-u)} \tag{3}$$

Variables *east* and *north* were defined in order to include the wind direction factor in the ANN model. Variable *east* changes from -1 to 1 as the wind direction changes from west to *east*. Variable *north* changes from -1 to 1 as the wind direction changes from south to *north*. Fig. 2 lists the wind directions and corresponding *east* and *north* values in the ANN model. Values in parenthesis show *east* and *north* values, respectively.

Wind direction has been expressed with a single variable as radian or 360 degrees in previous literature (Berkowicz 2000a, Berkowicz 2000b, Henry et al. 2002). This approach is not appropriate since wind directions with small and large degrees are apart as numbers, but very close in reality (such as 5 and 360 degrees), and therefore such modeling is not reliable. The wind direction modeling approach as in Aktan (2008) is more appropriate since such modeling can distinguish the difference between all wind directions.

Independent variables temperature (*temp*), wind speed (*w*), pressure (*p*), relative humidity (*h*), precipitation (*pc*), east, north, sunshine amount (*sa*),

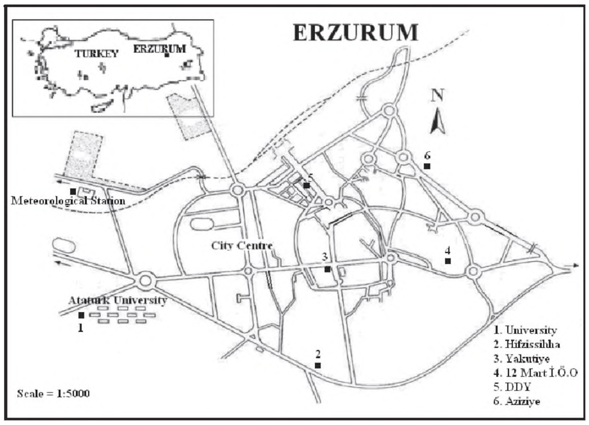

Fig 1. Map of the study area and air quality monitoring sites.

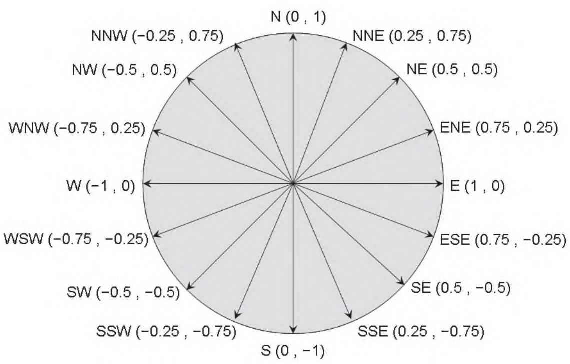

Fig 2. Defining the wind direction with numeric variables.

and sunshine duration (*sd*) were defined as numeric variables (Khare and Nagendra 2007). Input data was randomly partitioned into training (%68.1), validation (%15.95) and test (%15.95) sets. The training set is used for the neural network training, i.e. for adjustment of network weights. The validation set is used to tune the network topology or network parameters other than the weights. The test set is used to test how well the neural network will perform on new data.

RESULTS

The best network was searched using one hidden layer and two hidden layer network architectures. The network with minimum mean absolute error on the test set was selected. The best performance was obtained with one hidden layer network architectures. A network with too many hidden neurons tends to memorize all data instead of finding relations, which also leads to bigger network errors.

Statistical analysis of the data showed that the correlation coefficient of the TSP levels for a one day difference (between TSP_t and TSP_{t-1}) is 0.85, and for a two-day difference (between TSP_t and TSP_{t-2}) is 0.69. These values indicate that a TSP prediction is strongly related with the TSP levels of the past two days. Therefore, TSP is defined with an autoregressive model as a function of eleven variables is as follows:

$$TSP_t = f\,(TSP_{t-1}, TSP_{t-2}, temp, w, p, h, pc, sa, sd, east, north) \qquad (4)$$

Reducing the error more and more on the training set causes the network to learn also the noise contained in the data. Therefore, a stopping condition was used which terminates the iterations if the error improvement is less than 10^{-4}, or when 500 iterations are completed.

Five separate neural networks were constructed with the available data to model the TSP levels of the five winter season months, i.e., November, December, January, February, and March. Three error minimization algorithms were applied for the network training: conjugate gradient descent, quasi-Newton, and Levenberg-Marquardt (Bertsekas 1995). Fig. 3 presents the selected ANN architecture that provided the best fit for the month January. The ANN in Fig. 3 has eleven input neurons for the eleven factors, three neurons in the hidden layer, and one output neuron for the TSP estimate. Table 1 summarizes the prediction performance of the constructed neural networks and linear regression models containing the eleven factors for each month.

The R^2 value of a model is the proportion of the total variability in the dependent variable (TSP) that is accounted for by the model (Montgomery and Runger 2003). The R^2 values of the fitted ANN and linear regression models are shown in Table 1. Large R^2 values for all of the five neural networks show that the TSP values estimated by the neural networks are close to the observed TSP values.

Fig. 4 is a scatter plot where the observed and the estimated TSP values in February are shown. This figure also implies that TSP values estimated by the ANN are close to the actual TSP levels.

DISCUSSION

Apparent air pollution has been a problem in the city of Erzurum during the winter seasons. Maximum TSP concentrations occur in November, December, January, February, and March. An understanding of pollution sources and emissions, and their interactions with the terrain and the

Table 1. ANN and regression model information.

Month	Number of neurons in the hidden layer	R^2 value of the neural networks		R^2 value of the linear regression models
		all data	training data	
November	5	0.771	0.782	0.700
December	6	0.793	0.899	0.762
January	3	0.763	0.802	0.757
February	5	0.813	0.880	0.769
March	7	0.804	0.893	0.786

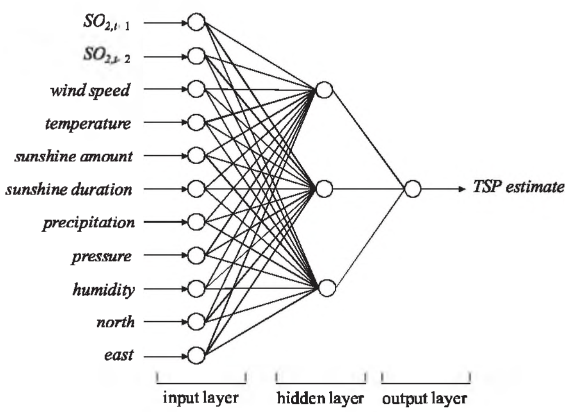

Fig 3. ANN structure for January TSP modelling.

atmosphere, is the most important step in developing appropriate air pollution management plans and action strategies. Without this type of knowledge, incorrect decision making related to air pollution management is possible, creating wasted resources and undesirable results (Bridgman et al. 2002).

In order to predict the TSP concentrations using meteorological parameters for each winter season month, ANN models were developed. The ANN

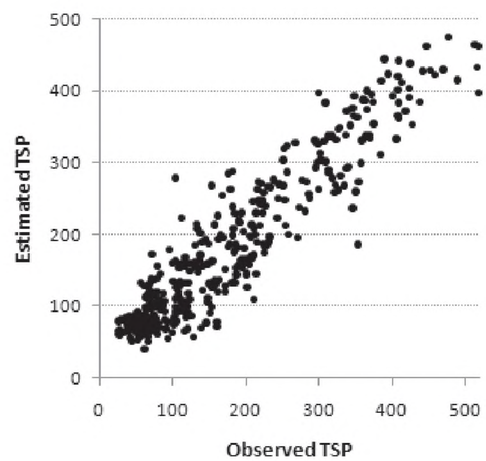

Fig 4. Scatter plot of observed and estimated TSP.

models provided R^2 values greater than 0.76.

The impact of wind direction was defined using the combination of two variables, which are the east and the north components of the wind direction. This definition of the wind direction is more appropriate than using a single variable as a radian or 360 degrees. Statistical data analysis showed that the TSP values were strongly related with the previous two days' TSP levels. Therefore, we defined an autoregressive structure for the model. With the help of the TSP predictions of the ANN model, appropriate warning signals can be obtained, and relevant actions can be taken by the government or the public to reduce the pollution to non-harmful levels.

REFERENCES

Aktan M (2008) Artificial neural network modeling of winter season sulphur dioxide concentrations in Erzurum city. Fresenius Environmental Bulletin 17, 2, 218-223.

Bayraktar H, Turalıoğlu FS (2005) A Kriging-based approach for locating a sampling site-in the assessment of air quality. Stochastic Environmental Research Risk Assessment 19, 4, 301-305.

Berkowicz R (2000) A simple model for urban background pollution. Environmental Monitoring and Assessment 65, 259-267.

Berkowicz R (2000) OSPM - a parameterised street pollution model. Environmental Monitoring and Assessment 65, 323-331.

Bertsekas DP (1995) Nonlinear Programming. Athena Scientific, Belmont.

Bridgman HA, Davies TD, Jickells T, Hunova I, Tovey K, Bridges K, Surapipith V (2002) Air Pollution in the Krusne Hory region, Czech Republic during the 1990s. Atmospheric Environment 36, 3375-3389.

Chao Z (1990) Urban climate and air pollution in Shangai. Energy Build 15, 647-656.

Çuhadaroğlu B, Demirci E (1997) Influence of some meteorological factors on air pollution in Trabzon city. Energy Build 25, 179-184.

Elbir T, Müezzinoğlu A, Bayram A (2000) Evaluation of some air pollution indicators in Turkey. Environment International 26, 5-10.

Haykin S (1994) Neural networks: a comprehensive foundation. Macmillan, New York.

Henry RC, Chang YS, Spiegelman CH (2002) Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction. Atmospheric Environment 36, 2237-2244.

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

Hyslop NP (2009) Impaired visibility: the air pollution people see. Atmospheric Environment 43, 182-195.

Khare M, Nagendra SSM (2007) Artificial Neural Networks in Vehicular Pollution Modelling. Springer-Verlag, Berlin, Heidelberg.

Kurt A, Gulbagci B, Karaca F, Alagha O (2008) An online air pollution forecasting system using neural networks. Environment International 34, 592-598.

Liu J, Mauzerall DL, Horowitz LW (2009) Evaluating Inter-Continental Transport of Fine Aerosols: (2) Global Health Impact. Atmospheric Environment DOI: 10.1016/j.atmosenv.2009.05.032.

Montgomery DC, Runger GC (2003) Applied Statistics and Probability for Engineers. John Wiley and Sons Inc., New York.

Ripley BD (1996) Pattern recognition and neural networks. Cambridge University Press, Cambridge.

Topçu N, Keskinler B, Bayramoğlu M, Akçay M (1993) Air pollution modeling in Erzurum city. Envimental Pollution 79, 9-13.

Taşdemir Y, Çağlar Ö (2002) Uludag Üniversitesi Kampüsü`nde Ölçülen Partikül Madde Kuru Çökelme Akıları. Ekoloji 11, 42, 8-11.

Yıldırım Y, Bayramoğlu M (2006) Adaptive neuro-fuzzy based modelling for prediction of air pollution daily levels in city of Zonguldak. Chemosphere 63, 1575-1582.