J. Int. Environmental Application & Science, Vol. 11(1): 1-7 (2016)

Modelling of Atmospheric SO_2 Pollution in Seydişehir Town by Artificial Neural Networks[#]

Z. Cansu Ozturk[*], Sukru Dursun

Environmental Engineering Department, Selcuk University, Konya, Turkey

Received August 22, 2015; Accepted December 16, 2015

Abstract: Air pollution has become a major environmental problem since last century because of the effects of fast population growth and industrial developments. Sulphur dioxide is considered as one of the major and most common air pollutant with using fossil fuels causing severe health problems such as disrupting tissues and mucous membranes of the eyes, disturbing nose and throat because of the irritating toxic odour, and affecting badly to upper part of respiratory system and bronchi. Seydişehir town of Konya was selected as working area for this study because heavy industrial activities are very wide in many fields such as mining and manufacturing industry. Also, usage of fossil fuels for heating system in winter period is other important atmospheric pollutants source. Eti Aluminium facility is the biggest industrial unite for SO_2 pollution source in Seydişehir town. In this study, SO_2 pollution in Seydişehir town was modelled with Artificial Neural Networks (ANN) which uses characteristics of biological neurons and capable of solving highly complex problems constructing parallel computations. Meteorological factors and previous day's SO_2 concentrations were integrated to model as input parameters and next day's SO_2 concentration was tried to be predicted. Two seasons were selected for model development namely winter and summer. Prediction performances of develop models are 67% for winter season and 81% for summer season. These values are compatible compared with previous studies using ANN modelling and can be improved with larger data sets.

Keywords: *SO_2, ANN, Air Pollution, Modelling, Prediction*

Introduction

Air pollution and its effects have become global issues since middle of 19s. Air pollution is transported to long distances with air movements and has global effects. The major concerns should be cautious that greenhouse effect major cause of global warming and depletion of ozone layer with effect of many primary and secondary pollutants (Zannetti, 1990). Major primary air pollutants which are contaminants causing some adverse effects on environment are particulate matter (PM), sulphur compounds (e.g., SO_2, H_2S), nitrogen compounds *(e.g.,* NO, NH_3), carbon compounds (*e.g.*, HCs, CO, CO_2), halogen compounds (*e.g.*, fluorides, bromides, chlorides) (Zannetti, 1990). Moreover, chemical reactions happening in the atmosphere causes the transformation of primary pollutants to secondary ones, for example, SO_2 gas transforms to SO_4^{2-} (Zannetti 1990).

Sulphur dioxide emissions are considered as one of the major and most common air pollution problem in the world. SO_2 emissions are responsible from London winter-type smog event and other many lethal smog events. Because of acid deposition problem, water bodies and animal life is also affected from SO_2 gas. Main anthropogenic sources of SO_2 are fuel combustion, petroleum, mining, industrial processes like paper production (Zannetti, 1990).

In Turkey with the development of the industry and population growth air pollution has started to become a problem since beginning of 1980s. Seydişehir is a town of Konya which is one of the biggest cities of Turkey. In Seydişehir industrial activities are very wide in many fields such as mining and manufacturing industry. Especially mining has significant importance because of the minerals, bauxite, chromium and lignite, extracting from the area. Both extraction and manufacturing steps of minerals produce huge amount of emissions. Moreover, other manufacturing industries like chemical production, agricultural material production, and food production emits SO_2 gas to the atmosphere

[*]*Corresponding: E-Mail: cansoztrk@gmail.com; Tel: +903322232057; Fax: +903322410635*

[#]*This paper is presented from M.Sc. Thesis of Z. Cansu Ozturk*

(URL 1). Also, fossil fuel combustion from residential buildings and other institutions for heating purposes largely contributes to SO_2 pollution in Seydişehir.

Eti Aluminium Factory is the biggest metal production and the only aluminium manufacturing facility of Turkey. It was constructed in closed area of 12,000,000 m² (URL 2). During aluminium production Sulphur in the anodes is oxidized, releasing SO_2 from the potlines which isa row of electrolytic cells used in the production of aluminium as the anodes are consumed. SO_2 is also released from the anode bake furnace as pitch used to help form the anodes are oxidized during the baking process (URL 3). As a result of this anode process during aluminium production, Eti Aluminium facility contributes SO_2 pollution in Seydişehir town.

All of these sources of SO_2 in Seydişehir make it a problem for town. SO_2 is not lethal in all concentrations but it decreases life quality for human in many manners such as giving harm to tissues and mucous membranes of the eyes, disturbing nose and throat because of the irritating odour, and affecting badly to upper part of respiratory system and bronchi (Hussainn, 2011). Also, it affects the animal life, vegetation and environment badly. All these negative effects of SO_2 pollution are considerable.

Nowadays, prediction of air pollutants is increasing trend because with the help of future prediction of air pollutants establishing emission control legislations, evaluating the future emission control impacts, selecting possible locations in which it can be possible source of air pollutants, controlling air pollutant episodes and taking preventive precautions, and assessing present air pollutant sources and their responsibility in future events may be achieved (Zannetti, 1990). In this study a model was established in order to predict SO_2 pollution in Seydişehir town. There are many techniques used for prediction modelling such as statistical models, neural networks, and fuzzy logic. Artificial neural network (ANN) was used for this study and MATLAB (R2011a) software was used for ANN development.

ANNs are structures which consist of interconnected simple adaptive elements having ability to make parallel computations for processing and representation of knowledge (Basheer and Hajmeer 2000). It has many advantages when it's compared with traditional models because these models only require known input data without any assumption (Gardner & Dorling 1998). With the help of suitable connecting weights and transfer functions, approximation of a multilayer function can be measurable function between input and output (Hornik*et al*., 1989; Gardner & Dorling 1998). Most of the parallel structured ANN models include intense interconnected adaptive units. ANN models have very significant characteristic that highly nonlinear problems may be solved (Zurada, 1997).

Back propagation (BP) feedforward algorithm used for training the multilayer neural networks is the most famous learning algorithm used in ANN models (Basheer and Hajmeer 2000).In this study Multilayer Feed forward Neural Network including input layer, hidden layer and final layer consisting of neurons was used. Meteorological parameters which are hourly average temperature (^oC), hourly average pressure (bar), hourly relative humidity (%), hourly wind speed (m/sec), hourly rainfall (mm), hourly cloudiness, hourly sun duration, hourly sun radiation (kW/m^2) and previous day's hourly SO_2 pollution concentration (ppb) was integrated to model as input parameters and next day's SO_2 concentration was tried to be predicted. SO_2 data used in this model was gathered between 2012-2013 years just on particular months by M100E UV Fluorescence SO_2 Analyser device and meteorological data measured with automatic meteorological observation station (AMOS) was provided from General Directorate of Meteorology. Gathered data was separated to two datasets namely, winter and summer and one model was developed for each.

Material and Method

Data Gathering

Pollution data of the study was gathered by SO_2 analyser device which was positioned in Seydişehir town between 2012 and 2013 years. Detailed information related to this device is given in the next section.

M100E UV Fluorescence SO_2 Analyser was used for SO_2 measurement in this area. The Model 100E uses the proven UV fluorescence principle, coupled with state of the art microprocessor technology to provide accurate and dependable measurements of low level SO_2. Measurement range of the device is between 0-50 ppb to 0-20 ppm (M100E Operation Manual).

Meteorological data was provided by General Directorate of Meteorology. Data including average hourly temperature, average hourly pressure, hourly relative humidity, hourly wind speed, hourly rainfall, hourly cloudiness, hourly sun duration, hourly sun radiation values for 2012 and 2013 years was taken. Meteorological measurements of Seydişehir are carried with 17898 no of automatic meteorological observation station (AMOS).

Data Analysis

Before using data for modelling, it should be normalized in order to get better prediction performance. There are two main normalization methods such as z-score and max-min which can be used to prepare input data for modelling. In order to understand which one give better results, both of them was applied on winter and summer data separately and 10 trials of feedforward neural network with default settings were done to normalized data in MATLAB (R2011a). Default settings separate data into training, testing, and validation subsets as 70%, 15%, and 15%, respectively. Moreover, 10 neurons in hidden layer are used and as a training function Levenberg-Marquardt is applied in default settings of feed forward neural network. After these trials, average of mean square errors (MSE) was taken and the normalization method that gave lowest MSE was chosen as a best option because MSE is defined as the average squared difference between outputs and targets. Lower values are better. Zero means no error.

As a result of trials, z-scores normalization of input data for winter gave better results than max-min normalization while max-min normalization gave better results for summer input data than z-scores normalization. The summary of the results taken from proposed methods is given in Table 1.

Table 1. MSE values taken by z-scores and max-min normalization approaches for input data

	Winter (Avg. MSE)	**Summer (Avg. MSE)**
Z-scores normalization	**0.0509801**	0.001189285
Max-min normalization	0.0554530	**0.000885296**

Model Development

In order to develop ANN model, first of all architecture of model should be determined after data optimization process. Every ANN model consists of three layers namely input layer, hidden layer, and output layer. For this study there are 9 neurons in input layer and 1 neuron in the output layer. Hourly average temperature (^oC), hourly average pressure (bar), hourly relative humidity (%), hourly wind speed (m/sec), hourly rainfall (mm), hourly cloudiness, hourly sun duration, hourly sun radiation (kW/m^2) and previous day's hourly SO_2 pollution concentration (ppb) were integrated to the model as input parameters and SO_2 concentration (ppb) of next day was tried to be predicted.

The neuron number in hidden layer has significant importance for model development because it affects the model accuracy and performance. "The best network topology corresponds to a feedforward neural network which presents a minimum value of MSE for the validation data set." (Sousa, Martins et al. 2007) As a result, determination of neuron number in hidden layer was done according to (MSE).

Another important step of model development is determination of activation function which connects the corresponding parameters. Mostly used activation functions are linear, sigmoidal, hyperbolic tangent, squashing, and linear threshold. For this model, hyperbolic tangent activation function was chosen for connecting input layer and hidden layer and linear activation function was chosen for connecting hidden layer and output layer. In Figure 1 general architecture of developed ANN model is given.

Moreover, data is separated in three parts namely training, testing and validation sets. Majority of the data is used in the training data set for training purpose. The other data is used in the testing data set in which trained model performance is checked. When the model error is minimized on test data set, the training step is stopped. Lastly, predicted model is validated by validation data set. Thus, training function should be selected and this selection affects the performance of the model. There are many choices of training function in NN toolbox of MATLAB which uses backpropagation learning algorithm. In this study 70% of data was used in training set, 15% in testing and 15% in validation sets. Training functions were selected with respect to trials of several functions and comparing average *R* values.

Neuron number in the hidden layer was designated according to minimum value of average MSEs. By using training function (trainlm) which was selected inside from 25 trials of different training functions, neuron numbers 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50 was used in hidden layer separately and 10 trials for each were done. Consequently, 15 neurons in hidden layer gave the best average MSE result for winter data and 5 neurons for summer data.

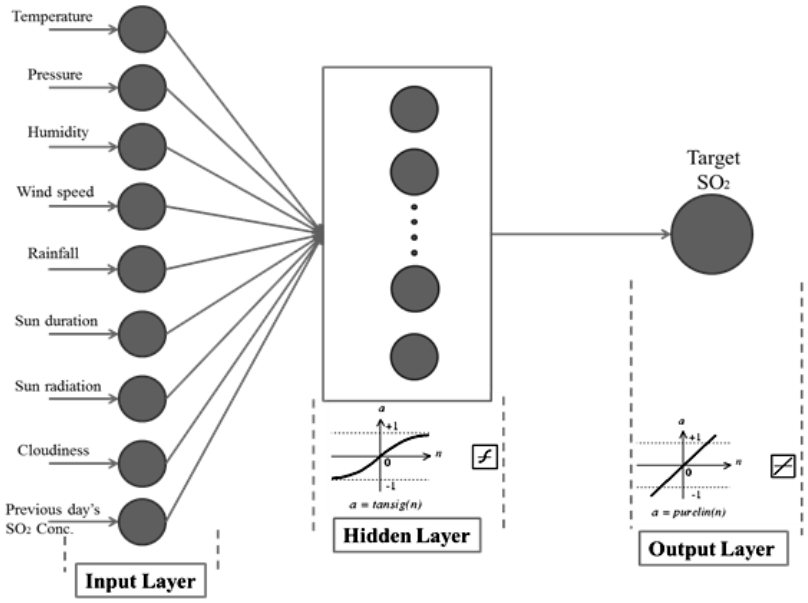

Figure 1 General architecture of developed ANN model

Results and Discussion

Winter Model

Developed model with winter data produced mean square error closer to zero which is important because it shows how close a fitted line is to data points and is used to evaluate the performance of a predictor or an estimator. MSE result of the model is 0021277 at best validation point. Other important criterion showing the prediction ability of the model is the *R* values which gives the correlation between output and target values. If R values close to 1, this means high correlation between target and predicted values. *R* values taken out by winter model are given in Figure 2 R value of training is 0.74569, testing is 0.44229, validation is 0.73291, and overall value is 0.67012.

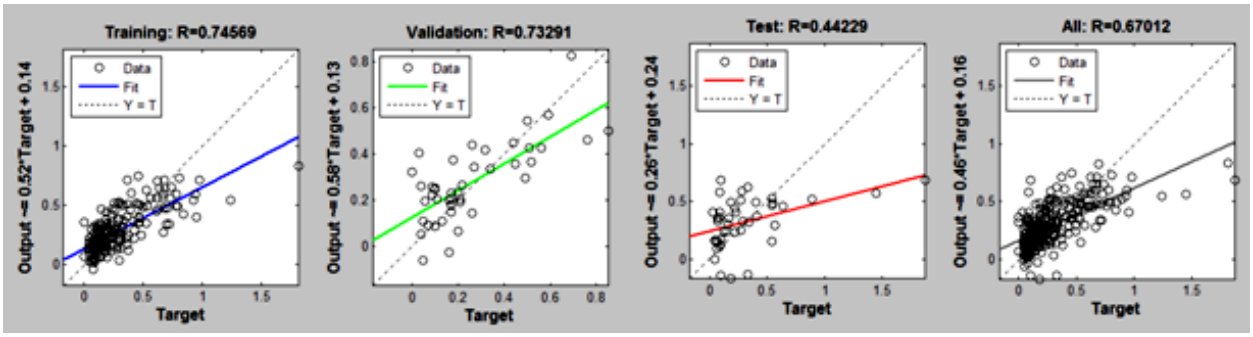

Figure2 R values taken out from winter model

Summer Model

MSE of summer model is 0.00019252. According the results of model developed with summer data, MSE value is closer to zero than winter model that means prediction performance of this model is better than winter. R values of the summer model are 0.92841 for training, 0.26353 for testing, 0.96069 for validation, and 0.81231 for overall given in Figure 3. These R values are better than winter for training, validation and overall but testing results show worse performance than winter. Moreover, in Figures 4 and 5 shows the linear equations taken out by actual and predicted values for

winter and summer. R^2 values taken from these graphs show that summer models' prediction performance is better than winter.

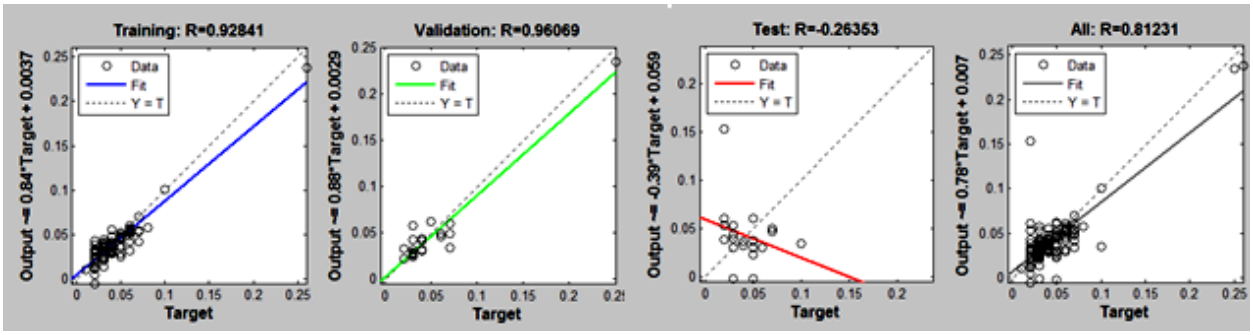

Figure 3. *R* values taken out from summer model

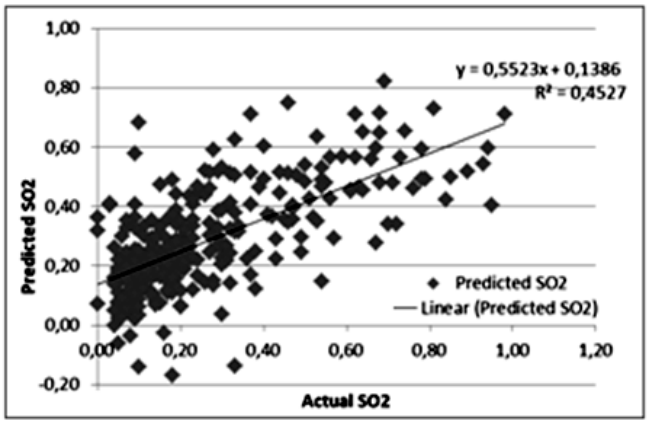

Figure 4 Comparison of actual and predicted values with linear equations for winter

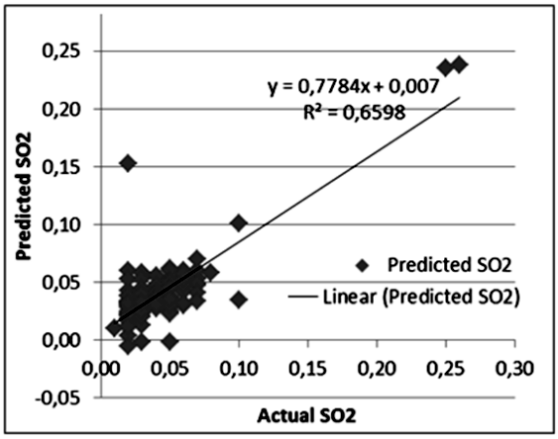

Figure 5 Comparison of actual and predicted values with linear equations for summer

Prediction performances of developed models for winter and summer seasons in Seydişehir and 67% for winter and 81% prediction for summer were achieved in overall. It may be affected from many reasons. Because of the problems in SO_2 analyser device, there are some missing values for some months so it was impossible to evaluate SO_2 data for whole year. Thus, model was developed just for winter and summer season with available data. Moreover, meteorological data was provided from General Directorate of Meteorology and integrated to model as input parameter with SO_2 concentrations. For those reason measurement locations of meteorological data and SO_2 data were not enough closer from each other.

Mean square error is important estimator for determining average squared difference between output and target values. Lower values of MSE means good estimation and 0 means no error. According to results of this study, MSE of winter model is 0.021277 and MSE of summer model is 0.00019252. Akkoyunlu *et al*. (2010) predicted SO_2 pollution conducted for İstanbul. Three models were developed for winter, summer and overall. MSE values taken out by these models were as 0.0042634, 0.33397419 and 0.125757. Furthermore, Yüksek *et al*. (2007) made SO_2 prediction with ANN approach for Sivas and they found the MSE value as 0.00204885. Therefore, when the MSEs of this study are compared with previous ones, it is clearly seen that these values are acceptable and good.

Another important estimator used for determining model accuracy is R values. In ANN approach data is divided in three subsets called training, testing and validation. After training, ANN model gives figures for these subsets. Evaluation of the performance of the model is done with R values which show prediction performance of the model. R values closer to 1 means that high performance. In this study R values are unstable between subsets for summer season. There are several reasons of that kind of variance. One of the reasons is memorizing the learning process during training. (Basheer & Hajmeer 2000) Moreover, size of the datasets is important because too many or too low data may produce bad results

Chelani*et al*. (2002) tried to predict SO_2 concentrations for Delhi city in India with artificial neural networks including wind speed, temperature, relative humidity, and the wind direction index as input parameters. Area was separated into three parts namely commercial, industrial and residential and three models were developed for each. Correlation between target and predicted values for each area were found as 0.68, 0.72, and 0.63. Moreover, Dursun *et al*. (2015) conducted a study for centre of Konya and SO_2 values predicted with indication of meteorological parameters by ANN and adaptive neuro-fuzzy inference system (ANFIS). Study results showed that total response of the ANN model is 0.778.

Developed models for winter and summer can be compared according to their R values taken by validation data set. *R* value of validation set is 0.73 for winter and 0.96 for summer. Validation set is used in most models for evaluate generalisation performance of ANN that is whether the approximation performance of model is efficient enough (Viotti *et al.* 2002).

Error in the validation set should be decreased during training. If data start to overfit which means generalisation ability of the model does not fit the new data, validation set error start to increase. In order to prevent this situation early stopping was applied to models. “When the validation error increased for a specified number of iterations, the training was stopped, and the weights and biases at the minimum of the validation error were returned.” (Akkoyunlu, 2011) The accuracy between test and validation sets shows model working unproblematic. R value of testing set for summer data did not show the same characteristics with validation data while for winter data test set and validation set performances showed similar characteristics. Although validation performance of summer model gave better results than winter season, test accuracy of winter season is better than summer season. Test set is defined as “the group of data, given to the network still in the learning phase, by which the error evaluation is verified in order to effectively update the best thresholds and the weights” by Viotti and Lituti (2002). This definition explains main purpose of test set in ANN modelling. In this study R values of test do not represent the efficiency of models. They just help weight and bias determination during training step. Test set is used for understanding the over fitting problem. In the case of summer model there might be over fitting problem occurred in validation set. However, when overall prediction performances of the models are compared, results of summer model are more successful.

Conclusion

In this study SO_2 pollution in Seydişehir town was modelled with artificial neural networks. Two models were developed for winter and summer seasons in Seydişehir and 67% prediction for winter and 81% prediction for summer were achieved in overall. Besides, *R* value of validation set is 0.73 for winter and 0.96 for summer. According to these results, generalisation performance of summer model gave better results than winter model. Besides, the prediction performance of this study is acceptable when it was compared with the studies in literature. It is possible to take more accurate results using ANN approach but many factors affect the prediction performance during modelling such as

abundance or lack of data and under or over training. Moreover, one of the problems for developing model is different measurement locations of meteorological data and SO_2 data. In order to overcome this problem, device which does meteorological measurements should be placed at the location of SO_2 pollution measured. By this means better model results may be taken out. Moreover, measurement duration should be increased and time intervals between each measurement should be decreased so more data may be gathered in order to reach better results. If more pollutants are measured such as NO_2, CO_2, CO, and PM, prediction capability of models may be increased. Thus, new measurement device which has capability of measuring more pollutants should be provided.

While there are many studies focusing on SO_2 modelling, it is a new subject in Turkey and most of the study subjecting to city air pollution modelling but there is no study focusing on town. In that manner this study is the first one focusing on town air pollution modelling and may be baseline of further researches.

Acknowledgement: *This study is financially supported by Selcuk University Coordinating Office of Scientific Research Projects (BAP) under grant no 15101016. This study has been prepared from Z. Cansu Öztürk's M.Sc. Thesis.*

References

Akkoyunlu A, Yetilmezsoy K, Erturk F,Oztemel E, (2010) A neural network-based approach for the prediction of urban SO_2 concentrations in the Istanbul metropolitan area, *Int. J. Environ. & Pollut.*, **40**, 301-321.

Basheer IA, Hajmeer M, (2000) Artificial neural networks: fundamentals, computing, design, and application, *J. Microb. Methods*, **43**, 3-31.

Chelani AB, Chalapati Rao CV, Phadke KM, Hasan MZ, (2001) Prediction of sulphur dioxide concentration using artificial neural networks, *Environ. Model. & Software* **17**, 161–168.

Dursun S, Kunt, F, Taylan O, (2015) Modelling sulphur dioxide levels of Konya city using artificial intelligent related to ozone, nitrogen dioxide and meteorological factors, International J. Environ. Sci.& Tech., **12**, 3915-3928.

Gardner MW, Dorling SR, (1998) Artificial neural networks (the multilayer perceptron) A review of applications in the atmospheric sciences, *Atmos. Environ*., **32**, 2627-2636.

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

Hussain ST, (2011) *Sulfur Dioxide*: *Properties, Applications And Hazards*, Nova Science Publishers, Inc., Chapter 3, 49-68.

M100E Operation Manual, (2011) UV Fluorescence SO_2Analyzer, Teledyne Advanced Pollution Instrumentation 9480, Carroll Park Drive San Diego, CA 92121-5201, USA, http://www.teledyne-api.com/manuals/04515F_100E.pdf, retrieval date: 02.08.2015.

Sousa SIV, Martins FG, Alvim-Ferraz MCM, Pereira MC, (2007) Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, *Environ. Model. & Software*,**22**, 97-103.

URL 1, 2014, Seydişehir İlçe Raporu, Mevlana Kalkınma Ajansı (MEVKA), Konya, http://www.mevka.org.tr/Download.aspx? Retrieval date: 15.07.2015.

URL 2, Seydişehir Eti Alüminyum Tesisi, http://www.etialuminyum.com/tr/Tesisler/Sayfalar/Seydisehir-Eti-Aluminyum-Tesisi.aspx, retrieval date: 15.07.2015.

URL 3, 2007, Bart Determination for Alcoa Intalco Works Ferndale, Washington http://www.ecy.wa.gov/programs/air/globalwarm_reghaze/BART/IntalcoBARTDeterminationFINAL.pdf, ENVIRON Corporation, retrieval date: 15.07.2015.

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

Yüksek GA, Bircan H, Zontul M, Kaynar O, (2007) Sivas İlinde Yapay Sinir Ağları ile Hava Kalitesi Modelinin Oluşturulması Üzerine Bir Uygulama, *C.Ü. İktisadi ve İdari Bil. Dergisi*, **8**, 97-112.

Zannetti P, (1990)*Air Pollution Modelling Theories, Computational Methods and Available Software*, Springer Science + Business Media, LLC, AeroVironment Inc. Monrovia, California, 3-20.

Zurada JM, (1997) *Introduction to Artificial Neural Systems*, WestPublish. Com., Mumbai, India.