Global NEST Journal, Vol 20, No 1, pp 103-108

Investigation of performance of tropospheric ozone estimations in the industrial region using differential Artificial Neural Networks methods

Akdemir[1,*] A., Filiz B.[1] and Özel Akdemir Ü.[2]
[1]Department of Environmental Engineering, Ondokuz Mayis University, Samsun, 55200, Turkey
[2]Department of Civil Engineering, Giresun University, Giresun, 28200, Turkey
Received: 24/04/2017, Accepted: 12/10/2017, Available online: 24/01/2018
*to whom all correspondence should be addressed: e-mail: aakdemir@omu.edu.tr

Abstract

The method of Levenberg-Marquardt learning algorithm was investigated for estimating tropospheric ozone concentration. The Levenberg-Marquardt learning algorithm has 12 input neurons (6 pollutants and 6 meteorological variables), 28 neurons in the hidden layer, and 1 output neuron for the Ozone (O_3) estimate. The Multilayer Perceptron Model (MLP) performance was found to make good predictions with the mean square error (MSE) less than 1 μg/m^3 (0.002 μg/m^3). In addition, the correlation coefficient ranges from 0.74 to 0.95 in The Levenberg-Marquardt learning. The Levenberg-Marquardt learning algorithm that a multilayer perception method of Artificial Neural Network (ANN) has performed well and an effective approach for predicting tropospheric ozone. Ozone concentration was influenced predominantly by the nitrogen oxide (NO_x, NO_2, NO), SO_2 and temperature. The model did not predict solar radiation to ozone with sufficient accuracy.

Keywords: MLP, Levenberg-Marquardt, Ozone, ANN

1. Introduction

Tropospheric ozone is formed in the troposphere when the photochemical equilibrium between ozone and nitrogen dioxide is disturbed by the presence of nitrogen oxide and volatile organic compounds under the influence of sunlight and high temperatures (Lissens *et al.*, 2000; Wang *et al.*, 2003). The daily maximum temperature ranges and other temperature related variables are crucial parameters for identifying episodes of ozone pollution. Exposure to ozone also induces damage to human health, agricultural cops, forest, and ecosystems, as well as to materials such as rubber and paints (Lissens *et al.*, 2000; WHO,2006).

Statistical models were used to explain the relationships between ozone concentrations, other pollutants and meteorological variables, then to adjust the ozone time trends (Wang *et al.*, 2003; Lu *et al.*, 2005; Spelmann, 1999; Abdul-Wahab *et al.*, 2002; Baur *et al.,* 2004; Khatibi, 2013). Multilayer perceptron (MLP) neural network model received a large attention and was applied in predicting air pollutants or estimating meteorologically adjusted ozone trends

This study aims to present two correction methods to improve the predictions of ozone concentrations achieved by the statistical models (MLR and MLP). MLR and MLP methods were used to estimate the meteorologically-adjusted trends of daily average ozone concentration in Samsun, Turkey.

1.1. Tropospheric Ozone Forecasting Using Artificial Neural Network

An ANN is defined as a massively parallel distributed information processing system made up of simple processing units, having a natural propensity for storing experiential knowledge and making it available for use. The purpose of an ANN is to build a model for problems such as pattern recognition and classification. Once an ANN has been trained on samples of datasets, it can predict outputs from inputs (Bui, 2012).

A list of the training algorithms that are available in the Neural Network Toolbox in Matlab v8.1 software and that use gradient or Jacobian-based methods, is shown in Table.1.

Table 1. Training algorithms in the Neural Network Toolbox in Matlab v8.1

trainlm	Levenberg-Marquardt
trainbr	Bayesian Regularizarization
traincgp	Polak-Ribiére Conjugate Gradient
traincgf	Fletcher-Powell Conjugate Gradient
trainscg	Scaled Conjugate Gradient
trainrp	Resilient Backpropagation
trainbfg	BFGS Quasi-Newton
traincgb	Conjugate gradient with Powell/Beale Restarts
traingdx	Variable Learning Rate Gradient Descent
trainoss	One Step Secant
traingdm	Gradient Descent with Momentum
traingd	Gradient Descent

Akdemir A., Filiz B. and Özel Akdemir Ü (2018), Investigation of performance of tropospheric ozone estimations in the industrial region using differential Artificial Neural Networks methods, *Global NEST Journal*, **20**(1), 103-108.

2. Experimental section

2.1. Site Description

Samsun (is a one of the largest city in northern Turkey on the Black Sea coast. The study area is the Organized Industrial District (OID) in Tekkeköy, Samsun, where there are small and medium scale unit different sectors of industries. OID, whose altitude, longitude and latitude are 8 m above the ground, 36°26'02'' W and 41°14'20'' N respectively, is located 12 km east of downtown Samsun.

2.2. Measured Data

The Artificial Neural Networks (ANNs) models employed the meteorological data affecting surface ozone behavior comprising: temperature (T), global radiation (GR), wind speed (WS), wind direction (WD), relative humidity (NN), atmospheric pressure (P) and nitrogen oxides (NO_x, NO_2, and NO), sulphur dioxide (SO_2), particulate matter (PM), carbon monoxide (CO) and ozone (O_3). Surface meteorological data recorded at intervals of the fifteen minute. Both meteorological data and pollutant data were collected started from 6 October 2011 to 6 May 2013. Meteorological sensors, which are produced by Lastem Company, were placed on the top of air quality measurement vehicle so the height of these sensors is 4.5 m from the ground.

The ozone concentration was measured by using an analyzer O_3 41 M of Environnement S.A based on absorption of UV radiation. The concentration of ozone was related to the magnitude of the attenuation. The Error of the measurement was ±1.0 ppb. Meteorological variables (daily mean wind speed ($m.s^{-1}$), daily mean wind direction (°), daily mean temperature (°C), daily mean relative humidity (%), solar radiation (W/m^2), and atmospheric pressure (hPa) and the average concentrations of air pollutants (NO_x, SO_2, PM_{10}, ve CO) were used as ozone predictors. These meteorological and pollutant parameters were measured at the interval of hourly. The statistical information on the data is presented in Table 2.

Table 2. Statistics for meteorological and air quality variables in Samsun, period: 2011-2013.

	Parameters	Mean	Min.	Max.	Std. Deviation	Skew.
Output	O_3 ($\mu g.m^{-3}$)	26.04	0.1	272.3	35.27	2.35
Input	NO ($\mu g.m^{-3}$)	1.98	0.088	130	7.58	3.88
	NO_2 ($\mu g.m^{-3}$)	7.44	0.1	149.4	10.98	3.88
	NO_x ($\mu g.m^{-3}$)	7.59	0.1	238.2	16.94	4.63
	SO_2 ($\mu g.m^{-3}$)	10.85	0.1	500.3	23.40	9.00
	PM_{10} ($\mu g.m^{-3}$)	47.12	0.3	597.5	47.95	3.55
	CO ($\mu g.m^{-3}$)	0.45	0.001	4.02	0.57	1.59
	Solar Radiation ($W.m^{-2}$)	104.86	0.0	893.6	181.25	2.01
	Temperature (°C)	11.26	-8.4	31.9	7.87	0.36
	Pressure (hPa)	1011.0	889	7.2	1032.1	7.16	-0.07
	Reletive Humidity (%)	74.19	3.9	99.4	27.19	-0.76
	Wind Speed ($m.s^{-1}$)	2.2	0.1	37.3	4.18	3.56
	Wind Direction (°)	50.31	0.0	360	92.78	1.87

The regression model associated ozone concentration, meteorological variables and other air pollutants can be expressed as;

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+.....+\beta_nX_{n1}+\varepsilon_i \quad (1)$$

In the regression model above, while $O_3(t)$ refers to the daily maximum ozone concentration at different time, $\beta_0, \beta_1, \text{and } \beta_n$, which are determined by the linear regression, are the coefficients of the independent variables ($X_{i1}, X_{i2}, ...X_{n1}$). On the other hand, the ε_i corresponds to the residual error of regression model (Özbay *et al.*, 2011).

The hypotheses required to apply MLP are the predictor variables must be independent, the residual error must be independent, and they must be normally distributed, with 0 mean and σ^2 constant variance. The least square method is the technique used to estimate the parameters (Basurka *et al.*, 2006).

Meteorological, ozone, and other pollutants data were standardized before application of MLR procedure, which normalized data were calculated in regard to following equation;

$$N_{i,j}=\frac{I_{(i,j)}-\min_{(j)}}{\max_{(j)}-\min_{(j)}} \quad (2)$$

where, I is the input value, NI is the standardized value, i is the number of data, and j is the measured value of the variable (Özbay *et al.*, 2011).

The MLP neural network was used since it can represent the complex nonlinear relationships between ozone concentration, meteorological variables, and air pollutants. The MLP network is a feed forward network and contains the input layer, the hidden layer and the output layer.

The raw data obtained from measurement sensors contains a large number of missing data due to maintenance and repair. The hourly raw datum is removed from the dataset if the values of seven or more parameters are not recorded. Thus, the number of hourly recorded values in dataset is reduced from 13872 to 12863. Additionally, these variables are normalized by using the equation below to minimize the impact of instantaneous changes in recorded values on the prediction performance;

$$X_{norm}=\frac{X-X_{min}}{X_{max}-X_{min}} \quad (3)$$

Via the normalization equation, the values of those parameters are ranged from 0 to 1. After the prediction of the ozone concentration through the model, the predicted values are converted to original values via the reverse

normalization to find the correlation between the measured ozone concentration and the predicted ozone concentration.

Table 3 below shows a list of the training algorithms using gradient- or Jacobian-based methods in the Neural Network Toolbox software.

All training algorithms of NNT software are used for both the prediction of the tropospheric Ozone concentration and the effects of the parameters on the prediction of the tropospheric Ozone concentration.

Table 3. Training algorithms that are available in the Neural Network Toolbox software

Acronym	Function	Neural Network	Algoritm
LM	trainlm	LMNN	Levenberg-Marquardt
BR	trainbr	BRNN	Bayesian Regularization
CGP	traincgp	CGPNN	Polak-Ribiere Conjugate Gradient
CGF	traincgf	CGFNN	Fletcher-Powell Conjugate Gradient
SCG	trainscg	SCGNN	Scaled Conjugate Gradient
RP	trainrp	RPNN	Resilient Backpropagation
BFG	trainfg	BFGNN	BFGS Quasi-Newton
CGB	traingb	CGBNN	Conjugate Gradient with Powell/Beale Restarts
GDX	traingdx	GDXNN	Variable Learning Rate Gradient Descent
OSS	trainoss	OSSNN	One Step Secant
GDM	traingdm	GDMNN	Gradient Descent with Momentum
GD	traingd	GDNN	Gradient Descent

3. Results and discussion

The mean square error (MSE) is used to test the performance of each algorithm of the NNT software. The mean square error is equal to the variance of the estimator plus the squared bias (Ross, 2009). The mean square error (MSE) was calculated in regard to following equation;

$$MSE=\frac{1}{N}\sum_{i=1}^{N} e_i^2 = \frac{1}{N}\sum_{i=1}^{N}(\alpha_i - t_i)^2 \quad (4)$$

where, (a) input data and (t) output data

After eliminating missing data, a total of 12863 hours long data is left. Whereas 8899 hours of it used for training of the algorithms, 1907 hours of it used for testing purpose. Additionally, 1907 hours of it is used for validation purpose whilst 150 hours of it is used for finding prediction performance.

The number of neurons in the hidden layer is strongly related to the performance of the model, so it is one of the most important factors for building (training ya da developing) the model. The number of neurons in the hidden layer of each model is found by calculating the performance of the model based on from 5 neurons to 30 neurons. The number of neurons having the least mean square error (MSE) and the highest R^2 is used for the identifying model. For instance, the graph below shows the mean square error (MSE) of LMNN algorithm relying on the number of neurons. As seen, the mean square error (MSE) is the highest provided that the number of neurons is 5, since the mean square error (MSE) is the lowest for 28 neurons. Therefore, the LMNN is identified as 12-28-1. In addition, the number of neurons in the hidden layer of Gradient Descent with Momentum (traindgm) and Gradient Descent (traingd) could not be found based on the given number of neurons.

The model networks are created based on the appropriate number of neurons in their hidden layer. These created models are trained with 12713 hour long data to update the biases and weights of each model. Finally, the ozone concentration based on 6 pollutant parameters and 6 meteorological parameters is predicted through the updated models. The correlation (R^2) between the predicted ozone concentration and the measured concentration is calculated to find the prediction performance of the trained networks.

The table below illustrates the model of the networks consisting of the number of neurons in their hidden layers, the number of loops, The mean square error (MSE), training performance and prediction performance.

Table 4 illustrates that LMNN algorithm is the best predictor of Ozone concentration because in addition to being the fastest; it has the lowest measurement error, the best training and prediction performance as compared to the others. Considering the measurement error, training and prediction performance, BRNN algorithm performs as well as LMNN algorithm does, yet the training period of BRNN algorithm is extremely longer than the training period of LMNN algorithm. Thus, even though either LMNN algorithm or BRNN algorithm could be used for smaller data sets, LMNN algorithm should be used for relatively larger data sets. Additionally, although not only is the training period of CGPN algorithm is shorter than many algorithms, but also the prediction performance of it better than the majority of the algorithms, its prediction performance is poorer than LMNN algorithm and BRNN algorithm. Since CGPN algorithm is trained in the relatively shorter period, it might be chosen for predicting larger data sets instead of LMNN algorithm. Furthermore, the training performance of CGBNN algorithm is extremely better than the prediction performance of it as compared to the other algorithms. In addition, the prediction performance of

CGFNN algorithm, SCGNN algorithm, OSSNN algorithm, RPNN algorithm and BFGNN algorithm are found almost the same, but the training period of RPNN algorithm takes longer than the others. Finally, GDMNN algorithm and GDNN algorithm are unable to be trained, so their prediction performance could not be found.

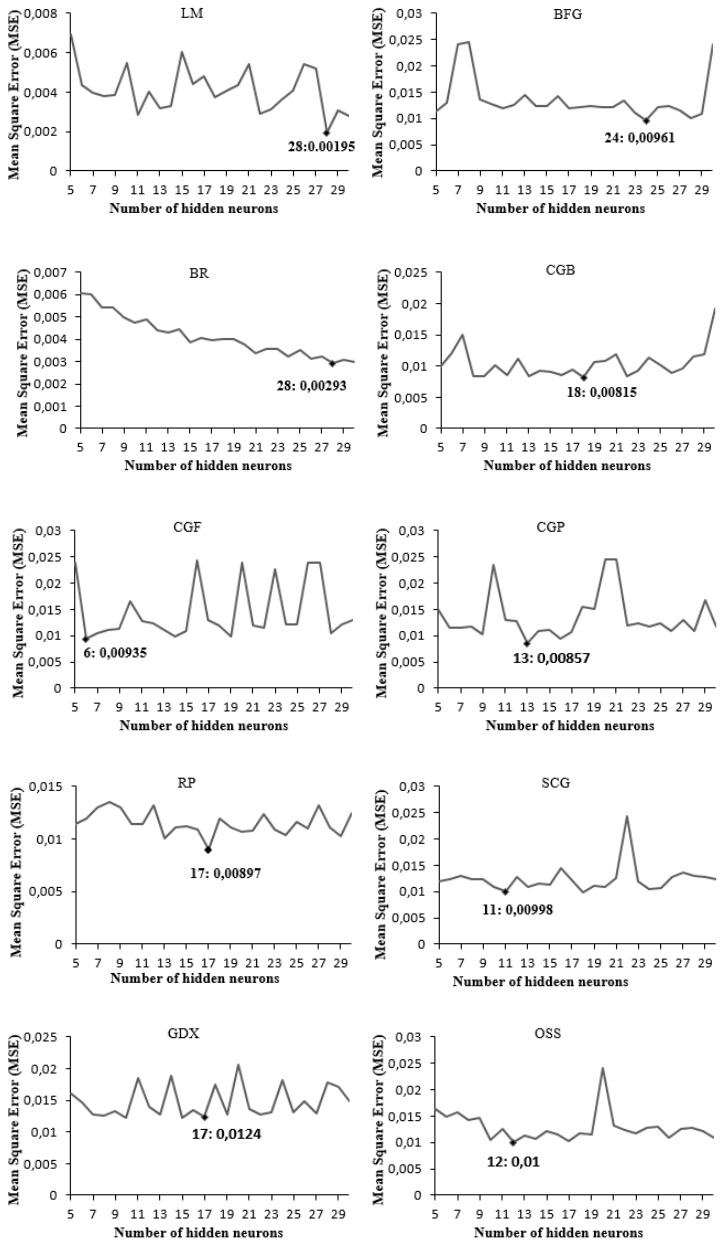

Figure 1. Results of the mean square error (MSE) values

Table 4. Predict performance values of training algorithms in the Neural Network Toolbox.

Neural network	Model	Epoch	MSE	Training (R^2)	Predict (R^2)
LMNN	12-28-1	27	0.00195	0.95	0.846
BRNN	12-28-1	829	0.00293	0.91	0.764
CGPNN	12-13-1	197	0.00857	0.68	0.637
CGFNN	12-6-1	159	0.00935	0.65	0.573
SCGNN	12-11-1	135	0.00938	0.62	0.571
OSSNN	12-12-1	199	0.01000	0.62	0.550
RPNN	12-17-1	500	0.00897	0.66	0.547
BFGNN	12-24-1	188	0.00961	0.65	0.494
CGBNN	12-18-1	210	0.00815	0.70	0.483
GDXNN	12-17-1	182	0.0124	0.49	0.321
GDMNN	--	0	0.0	0.0	0.0
GDNN	--	0	0.0	0.0	0.0

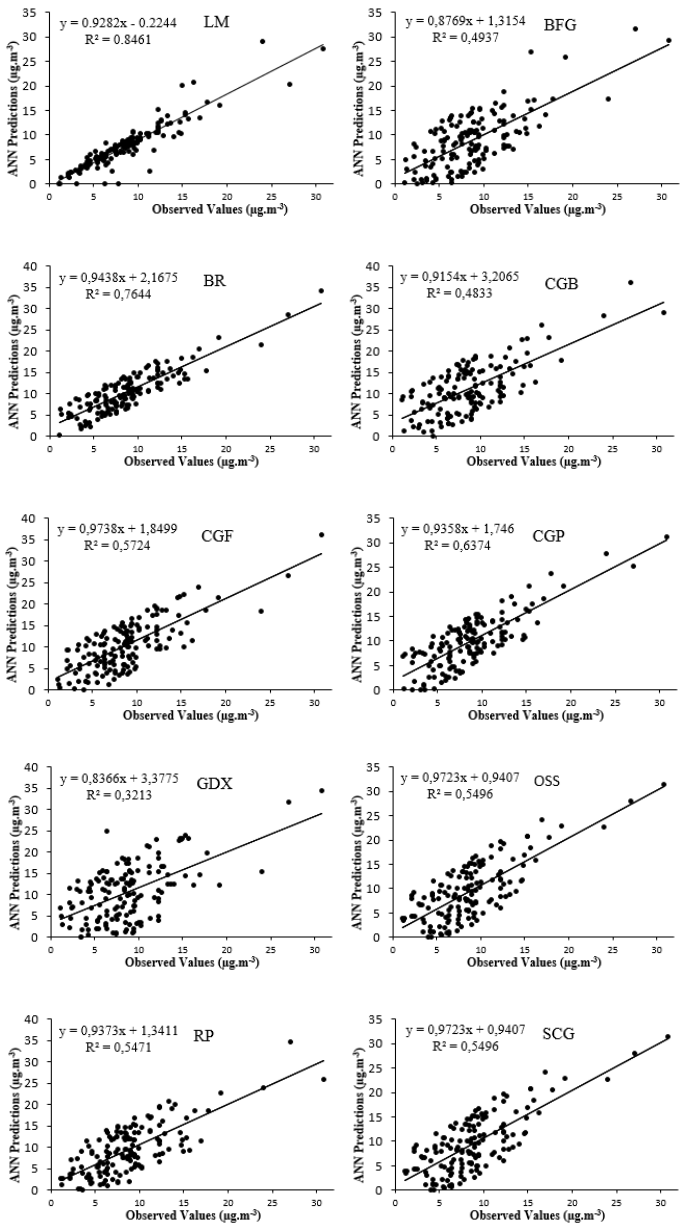

Figure 2. Comparison of the observed with the predictions of ANN with time delay of 1 h.

It is found that the prediction performance of LMNN algorithm is the best so the effect of the parameters on Ozone concentration is predicted by using LMNN algorithm. The figure below shows the significance level of these parameters for estimating ozone concentration.

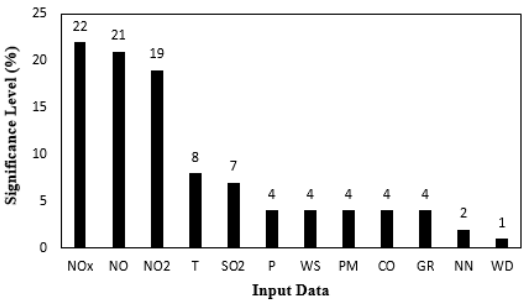

Figure 3. The importance levels of the parameter estimates correlation with ozone

In this study, the high importance level was observed totally 62 % with nitrogen oxides (NO_x, NO_2, and NO) as this issue supports in literatures (Arsic *et al.*, 2011; Karaca and Öztürk, 2012). Although the effect of solar radiation was expected to be higher, the importance level for solar radiation was observed 4 % due to the measurement errors for solar radiation.

In similar studies in the literature, three parameters (NO, NO_2, NO_x) were evaluated separately and have been associated separately. They cannot be considered apart from the high correlation between them (Gardner and Dorling, 1999; Vlachogianni *et al.*, 2011; Özbay *et al.*, 2011; Khedairia and Khadir, 2012).

4. Conclusions

The Levenberg-Marquardt learning algorithm that a multilayer perception method of ANN has performed well and is found that an effective approach for predicting

tropospheric ozone due to having the highest R^2 (0.74 to 0.95) and the lowest mean square error (MSE) (0.002 µg/m^3). Likewise, the prediction performance of BRNN algorithm is found acceptable as compared with the other algorithms, but its training period takes relatively longer than the others. Considering this issue, BRNN algorithm is more proper for predicting the ozone concentration based on the smaller data sets. Since the training of GDMNN algorithm and GDNN algorithm could not be succeeded, the prediction performance of these algorithms is unable to be calculated. Back propagation is an achieve method for network training in ambient air quality ozone concentrations but proper training algorithm and model configuring.

There are previous studies in which ozone meteorological parameters and regression, ARIMA, and ANN studies related to nitrogen oxides in particular (Yi and Prybutok, 1996). Ozon habercisi olan temel iki grup azot oksitler ve uçucu organik bileşiklerdir (Comrie, 1997; Xing *et al.*, 2017). The two main precursors of ozone are the nitrogen oxides (NO_x) and volatile organic compounds (VOC_s) (Comrie, 1997; Xing *et al.*, 2017). But, when the pollutant sources of the region to be measured are examined, there are plants that are predominantly sources of nitrogen oxides. In addition, when the emission reports of the installations in the region are examined, volatile organic compound release from the stacks is not realized. Another reason why VOC is not taken into consideration is the absence of continuous volatile organic measuring instrument according to research possibilities. It is considered that the estimation error of one of the reasons for the weakness of the ozone relation with solar radiation (Zhang *et al.*, 2017) is increased and the ANN does not show the desired performance according to the empirical models.

For further studies, the application of the multilayer perceptron (MLP) model should be extended to forecast of O_3 hourly levels in the industrial area. In addition, the incorporation of number variables such as volatile organic compounds and hydrocarbons ought to be included in predictors. But, the effect of VOCs could not explain for ozone because VOCs was not measured in this study.

Acknowledgments

This work was supported by the Research Fund of the Ondokuzmayıs University (Project number: PYO.MUH.1906.11.005).

References

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

Arsic M., Nikolic, D., Djordjevic P., Mihajlovic I. and Zivkovic Z. (2011), Epidoses of extremely high concentration of troposheric ozone in the urban environment in Bor-Serbia, *Atmospheric Environment*, **45**, 5716-5724.

Basurko E.A., Berastegi G.I. and Madariaga I. (2006), Regression and multilayer perceptron-based models to forecast hourly O_3 and NO_2 levels in the Bilbao Area, *Environmental Modelling and Software*, **21**, 430-446.

Baur D., Saisana M. and Schulze N. (2004), Modeling the effects of meteorological variables on ozone concentration-a quantile regression approach, *Atmospheric Environment*, **38**, 4689–4699.

Bui D.T., Pradhan B., Lofman O., Revhaug I. and Dick O.B. (2012), Landslide susceptibility assessment in the Hoa Binh province of Vietnam: A comparison of the Levenberg–Marquardt and Bayesian regularized neural networks, *Geomorphology*, **171-172**, 12-29.

Comrie A.C. (1997), Comparing Neural Networks and Regression Models for Ozone Forecasting, *Air and Waste Management Association*, **47**, 653-663.

Gardner M.W. and Dorling S.R. (1999), Neural Network Modelling and Prediction of Hourly NO_x and NO_2 Concentrations in Urban Air in London, *Atmospheric Environment*, **33**, 709-719.

Karaca F. and Öztürk Z. (2012), Statistical Evaluation of the Weekend Ozone Effect (WOE) in İstanbul, *Ekoloji,* **21**(82), 26-33.

Khatibi R., Naghipour L., Ghorbani A.M., Smith M.S., Karimi V., Farhoudi R., Delafrouz H. and Arvanaghi H. (2013), Developing a predictive tropospheric ozone model for Tabriz, *Atmospheric Environment*, **68**, 286-294.

Khedairia S. and Khadir M.T. (2012), Impact of clustered meteorological parameters on air pollutants concentrations in the region of Annaba, Algeria, *Atmospheric Research*, **11**, 89-101.

Lissens G., Debruyn W., Mensink C. and Dumont G. (2000), SMOGSTOP, a model for forecasting maximum daily ozone concentration in Belgium, *Environmetrics*, **11**, 511-521.

Lu H.S. and Chang T.S. (2005), Meteorologically adjusted trends of daily maximum ozone concentrations in Taipei, Taiwan, *Atmospheric Environment*, **39**, 6491-6501.

Özbay B., Keskin G.A., Doğruparmak Ş.Ç. and Ayberk S. (2011), Multivariate methods for ground-level ozone modeling, *Atmospheric Research,* **102**, 57-65.

Ross S.M. (2009), Introduction to probability and statistics for engineers and scientists: Elsevier, London, UK, pp.269.

Spelmann G. (1999), An application of artificial neural networks to the prediction of surface ozone concentration in the United Kingdom, *Applied Geography,* **19**, 123–136.

Vlachogianni A., Kassomenos P., Karppinen A. and Kukkonen J. (2011), *Science of the Total Environment*, **409**, 1559-1571.

Yi J. and Prybutok V.R. (1996), A Neural Network Model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area, *Environmental Pollution*, **92**,3,349-357.

Xing J., Wang J., Mathur R., Wang S., Sarwar G., Pleim J., Hogrefe C., Zhang Y., Jiang J., Wong D.C. and Hao J. (2017), Impact of Aerosol Direct On Troposheric Ozone Through Changes ın Atmospheric Dynamics and Photoysis Rate, *Atmospheric Chemistry and Physics*, **17**,9869-9883.

Wang W., Lu W., Wang X. and Leung A. (2003), Prediction of maximum daily ozone level using combined neural network and statistical characteristics, *Environment International,* **29**, 555-562.

World Health Organization, (2006) WHO Air Quality Guidelines for Europe, WHO Regional Publications, European Series, No.91, 181-182.

Zhang J., Zhao L., Deng S., Xu W. and Zhang Y. (2017), A critical review of the models used to estimate solar radiation, *Renewable and Sustainable Energy Reviews,* **70**, 314-329.