J. Int. Environmental Application & Science, Vol. 12(3): 256-263 (2017)

Modelling of PM_{10} Pollution in Karatay District of Konya with Artificial Neural Networks

Yasin Akın Ayturan[1, *], Ali Öztürk[2], Zeynep Cansu Ayturan[3]

[1]Electronic and Computer Engineering Department, Graduate School of Natural and Applied Science, Karatay University, Konya, Turkey; [2]Computer Engineering Department, Engineering Faculty, Karatay University, Konya, Turkey; [3]Environmental Engineering Department, Engineering Faculty, Selcuk University, Konya, Turkey

Received August 19, 2017; Accepted September 29, 2017

Abstract Air pollution is one of the most significant issues of human being faced nowadays because it can create adverse effects on both health of human and other livings. There are several air pollutants which are considered as dangerous such as sulphur dioxide (SO_2), nitrous oxide (NO_x), carbon monoxide (CO), volatile organic compounds (VOC) and particulate matter (PM). Particulate matter is one the most significant air pollutants because it may create respiratory, cardiological and pulmonary problems by inhalation by nose on humans. Also, heavy metals and hydrocarbons may be adsorbed on PM surface, so it is considered as carcinogenic by World Health Organization (WHO). When all these negative effects of PM are taken into consideration, it is important that PM future concentration should be determined for taking precautions. PM is classified according to the diameter of the particles and PM_{10} is described as particulates which has diameter smaller than 10 micrometres. In this study, PM_{10} pollution was predicted with artificial neural network (ANN) for Karatay district of Konya. ANN includes interconnected structures that can make parallel computations. Several meteorological factors and air pollutant concentrations was provided by database of Ministry of Environment and Urbanisation belonging to autumn period of 2016 such as SO_2 concentration, NO concentration, NO_x concentration, NO_2 concentration, CO concentration, O_3 concentration, wind speed, temperature, relative humidity, air pressure, wind direction and previous day's PM_{10} concentration. These parameters were used in the model as input parameters and PM_{10} concentration for one day later was used as an output parameter. Prediction performance of the obtained model was very promising when the similar studies are examined.

Keywords: *Artificial neural network, modelling, air pollution, PM_{10}, factor*

Introduction

Parallel with the population growth, air pollution is increasing day by day because of the energy consumption, rapid industrialization and urbanization. Air pollution mainly defined as increasing of the pollutants which can be found in the atmosphere as smoke, gas, dust and odour creating material damage and health problems on human and other livings (Güngör, 2013). Main air pollutants found in the atmosphere includes sulphur compounds (*e.g*. SO_2, H_2S), nitrogen compounds (e.g. NO_x, NO), volatile organic compounds (e.g. Benzene, Formaldehyde), carbon compounds (e.g. CO, HCs) and particulate matters (PM). PM is mainly emitted from natural resources such as volcanoes and sea, and anthropogenic activities like thermal power production, industrial processes, transportation. It is classified according to aerodynamic diameter such as particulates with diameter higher than 0.1 micrometer (μm) called as $PM_{0.1}$, particulates with diameter higher than 2.5 μm called as $PM_{2.5}$ and particulates with diameter higher than 0.1 μm called as PM_{10}.

In recent future activities causing PM pollution has been increased significantly. Because of this, negative effects on human health, decrease in the visibility range, acid deposition and deterioration of radiation balance of the world has become outstanding issues. There is a relationship has been found out between air pollution level originating from PM and death cases due to respiratory diseases, lung, cardio and respiratory problems (Karakaş, 2015). Moreover, in another research asthma disease has been investigated for 8530 households in 15 different cities of Turkey. According to the results of this research, at least one asthma patient was identified in each 495 (5.8%) households from 8530 households

**Corresponding: E-Mail: akinayturan@gmail.com; Tel: +903322410635; Fax:* +905548675660

(Şekerel *et al.*, 2006). Due to potential health and environmental impacts, the limit values for PM concentration in many regions of the world, including the European Union countries, have been determined by regulations. The European Parliament and the Council has determined outdoor air quality limit values as 40 µg/m^3 and 50µg/m^3 for PM_{10} and 17µg/m^3 for $PM_{2.5}$ with respect to directive 2008/50/EC. In order to decrease the negative health effects of pollution, air quality standards recommended by WHO (25 µg/m^3 and 50 µg/m^3 for $PM_{2.5}$ and PM_{10}, respectively) and European Council should be provided everywhere (Özdemir *et al.*, 2010). In Turkey, the air quality regulations that came into force on 6 June 2008, the standards for PM_{10} daily and annual basis are 50 µg/m^3 and 40 µg/m^3, respectively (Resmî Gazete, 2008).

Main emission sources in Konya are industrial activities and fossil fuel combustion. In Karatay district of Konya, there are huge industrial facilities, many residentials and excessive traffic density. Also, two big Organized Industrial Zones are present in the area. For this study, the measurement station located in Karatay was selected for providing the data because of the abundance of emission sources and lack of deficient values for that station.

There are several advantages of making future prediction such as constituting control regulations for emissions, assessing the future impacts of emissions, determining possible sources and locations of pollutants and taking precautions for air pollution episodes which will be come out (Zanneti, 1990). Therefore, when all these advantages are considered, making future prediction for Karatay district may be helpful. It can also create basis for the other studies which will be made about area although the prediction period is short. For this purpose, the model was created for PM_{10} emissions in the area with the capability of prediction of one-day period. MATLAB R2014a software was used in this study and artificial neural network (ANN) was determined as modelling method.

Neural network mechanisms use the computation features of human brain. High parallelism, nonlinearity, handling ability of fuzzy information, leaning, fault and failure tolerance and generalization ability of biological systems are used in neural networks (Basheer and Hajmeer, 2000). Like in human brain, there are neurons used for making computations with parallel structures formed by simple interconnected units (Pham and Pham, 1999). For solving complex problems about many different issues, ANN models have been developed (Rozlach, 2015). ANN uses back-propagation learning algorithm used in multilayer perceptron determines the weight modifications in the hidden layer of network and spread the error from output vectors (Basheer and Hajmeer, 2000).

In this study, SO_2 concentration (µg/m^3), NO concentration (µg/m^3), NO_x concentration (µg/m^3), NO_2 concentration (µg/m^3), CO concentration (µg/m^3), O_3 concentration (µg/m^3), wind speed (m/sec), hourly average temperature (ºC), hourly relative humidity (%), hourly average air pressure (bar), hourly wind direction and previous day's PM_{10} concentration (µg/m^3) were used as in the model as input parameters and PM_{10} concentration of one day later was predicted. Data of autumn period of 2016 (September, October, November) was used for developing the model.

Material and Method

Data Colletion

Data used for this study was provided from the database of Turkey air quality monitoring stations belongs to Ministry of Environment and Urbanisation. For this purpose, monitoring station found in Karatay district was selected because of the availability of sufficient data. In Figure 1 the location of Konya-Karatay air quality monitoring station is given in the map which includes other monitoring stations found in Konya.

Data Analysis

Data belong to autumn period of 2016 was arranged in Microsoft Excel and it was normalized before used in modelling for increasing the prediction performance of the model. There are several normalization techniques. In this study max-min normalization was used for reaching optimum values for the data. The equation 1 describes the max-min normalization method.

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \qquad \text{(Equation 1)}$$

Where:
zi : normalized value

min(x) : Min. value of x data
max(x) : Max. value of x data
xi : Value of ith in x data

Normalization was conducted in MATLAB R2014a software according to the code given below:

```
Function normNum = normalizeExcelFile(input)

    myPath    = fileparts(mfilename('fullpath'));  % Same folder as the M-file
    ExcelFile = fullfile(myPath, input);
    Num       = xlsread(ExcelFile);
    colMax = max(Num, [], 1);
    colMin = min(Num, [], 1);
    normNum = bsxfun(@rdivide, bsxfun(@minus, Num, colMin), colMax - colMin);
end
```

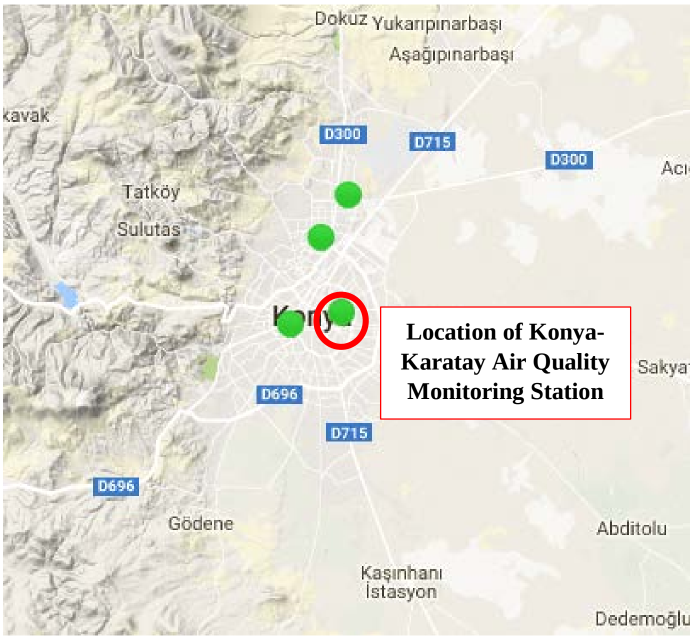

Figure 1. Location of Konya-Karatay Air Quality Monitoring Station

Model Development

In order to develop ANN model, there are several steps which should be followed. First of all, architecture of the model should be determined. ANN models generally include three layer such as input layer, output layer and hidden layer that enclose neurons. These layers are connected to each other and every connection in neural network is related to a weight that is a numeric number (Wang, 2003). The architecture of the model developed for this study is given in the Figure 2. As seen in Figure 2, input layer consists of 12 neurons, output layer consists of 1 neuron. SO_2 concentration, NO concentration, NO_x concentration, NO_2 concentration, CO concentration, O_3 concentration, wind speed, temperature, relative humidity, air pressure, wind direction and previous day's PM_{10} concentrations were entered to the model as input parameters and PM_{10} concentration of one day later was tried to be predicted as an output parameter.

The second step of developing ANN model is determination of activation function. The main role of activation function is connecting the neuron values for preventing the collapse of the model and come up with a nonlinear connection between input and output layers (Sarle, 1997). Determination of the activation function was made according to the previous studies conducted about this subject (Ozturk, 2015). As an optimum activation function hyperbolic tangent sigmoid transfer function (tansig) was selected for connection of the neurons in between input and hidden layer, and linear activation function (purelin) was selected for connecting neurons in between hidden and output layer as seen in Figure 2.

Third step is the determination of data division to training, testing and validation sets. For this work, back propagation (BP) learning algorithm which is the most well-known learning algorithm used in

ANN models was used. In this learning algorithm, gradient descent points are used for searching of the error. Iterations in BP include forward step in which solution is provided and backward step in which error is computed for modifying the weights (Basheer and Hajmeer, 2000). Also, training of the model is made according to this learning algorithm by adjusting the weights of the network and minimizing mean squared error (Gardner and Dorling, 1998). Three sets are used in BP algorithm which are training set, testing set and evaluation set. Major part of the data should be used for training, small portions are enough for testing and validation sets because training requires more data for learning process. Training is ended after errors are minimized in test data set. Lastly, performance of the model is validated in validation data set (Gardner and Dorling, 1998).

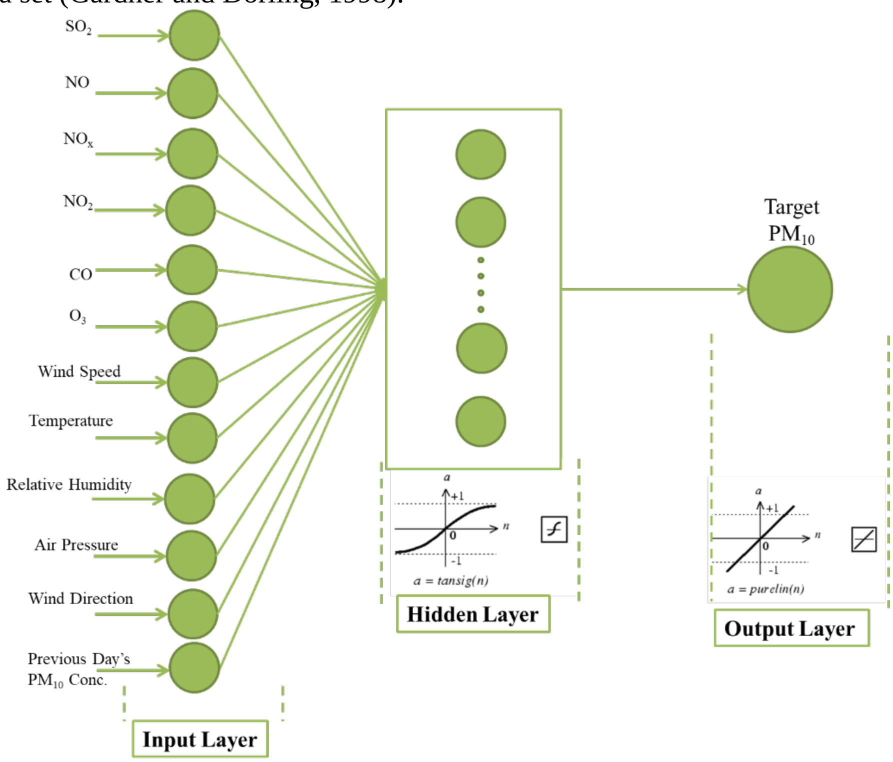

Figure 2. General Architecture of Developed ANN model

Forth step of developing the ANN model is determination of training function. MATLAB R2014a software offers three mostly used training function for ANN such as Levenberg-Marquardt backpropagation (trainlm), Bayesian regularization backpropagation (trainbr), and Scaled conjugate gradient backpropagation (trainscg). They have different characteristics and the results obtained by using each change one by one. Levenberg-Marquardt backpropagation algorithm is a standard technique used for nonlinear problems. Minimization of the sum of the squares of the errors between the data points and the function is the problem-solving technique of this algorithm (Gavin, 2017). Furthermore, Bayesian regularization backpropagation algorithm one of the most effective method for reducing the requirement of cross validation. It uses mathematical operation to convert nonlinear problem to statistical problem (Burden and Wrinkler, 2008). Moreover, scaled conjugate gradient backpropagation depends on well-known optimization techniques. It applies second order approximation for solving of the problem (Møller, 1990). For the present study, training function was determined according to several trials. Twenty trials were made for each function by keeping the determined data division and neuron number in the hidden layer was adjusted as 10. The function giving the smallest average value of mean square error (MSE) was selected as an optimum training function for this study.

The last step of the model development is determination of the neuron number in hidden layer. This number is very important for model performance because hidden layer affects the model performance and precision (Alsugair and Al-Qudrah, 1998; Sarle, 1997). If the neurons are fewer in hidden layer, results are obtained with high training and testing errors because of the statistical bias and under-fitting. Neuron number in hidden layer is decided according to the trials which is made until reaching the minimum value of average MSE. In this study, the selected training function was used for trials by keeping the determined data division for neuron numbers 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55 and 60. After 5 trials for each neuron number in hidden layer, optimum number was determined.

Results and Discussion

Results of Model Development Steps

In this study, 70% of the data was used in the training data set, 15% of the data was used in the test data set and 15% of the data was used in the validation data set.

Table 1 shows the average errors reached after trials for determination of training function. According to Table 1, Bayesian regularization backpropagation training function was determined as an optimum. Table 2 shows the results of MSE reached after 5 trials for determination of neuron number in hidden layer. 45 neurons were selected as suitable neuron number in hidden layer.

Table 1. Average MSE Reached After 20 Trials for Determination of Training Function

Training function	Criteria	Average Error
Levenberg-Marquardt backpropagation (trainlm)	MSE	0.00750633
Bayesian regularization backpropagation (trainbr)	MSE	**0.002564053**
Scaled conjugate gradient backpropagation (trainscg)	MSE	0.6039063

Table 2. Average MSE Reached After 5 Trials for Determination of Neuron Number in Hidden Layer

Neuron number	Criteria	Avg. Values
5	MSE	$4.56*10^{-3}$
10	MSE	$2.57*10^{-3}$
15	MSE	$1.52*10^{-3}$
20	MSE	$8.48*10^{-4}$
25	MSE	$5.32*10^{-4}$
30	MSE	$2.45*10^{-4}$
35	MSE	$8.39*10^{-5}$
40	MSE	$2.62*10^{-5}$
45	MSE	$\mathbf{1.70*10^{-5}}$
50	MSE	$1.00*10^{-4}$
55	MSE	$3.23*10^{-5}$
60	MSE	$7.93*10^{-5}$

Results of ANN Model

Results obtained from modelling by using MATLAB R2014a software are given in Figure 3. For this model one-day prediction was made for computing PM_{10} concentration of Karatay district. The prediction performance of the model can be understood with interpreting R values obtained after modelling. R values which are closer to one mean that the prediction performance is high in other words there is a positive correlation between model output and target values. Moreover, R value of validation set represents the overall performance of the model. According to Figure 3, R value of training set is 0.99996, testing set is 0.51528 and all is 0.86141. Therefore, overall performance of the model was found as almost 86%.

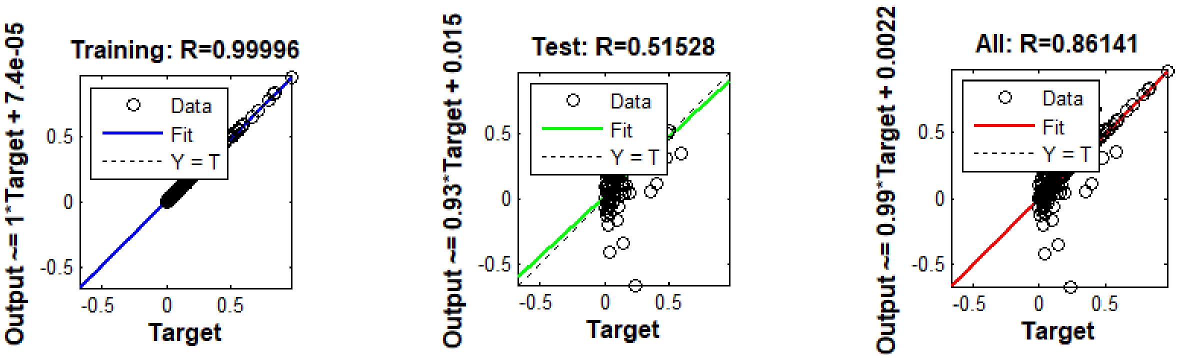

Figure 3. Regression Performance of Model

Another important estimator of model performance is MSE which is designated by calculating the average squared difference between output and target values. MSE values closer to zero mean that average squared difference between outputs and targets are better. Besides, MSE values are mainly used during model development steps. Figure 4 shows the MSE change during iterations. At 1000[th] epoch, error of the model reached to its minimum value. In this study, MSE of the best training performance was found as 1.5476×10^{-6}. This value may be considered as acceptable compared with the previous studies. According to study of Akkoyunlu SO_2 pollution for winter, summer and overall was predicted with ANN for İstanbul. MSE values taken out by these models were found as 0.0042634, 0.33397419 and 0.125757 (Akkoyunlu *et al.*, 2010). Besides, for Sivas city SO_2 prediction was made with ANN approach and they found the MSE value as 0.00204885 (Yüksek *et al.*, 2007).

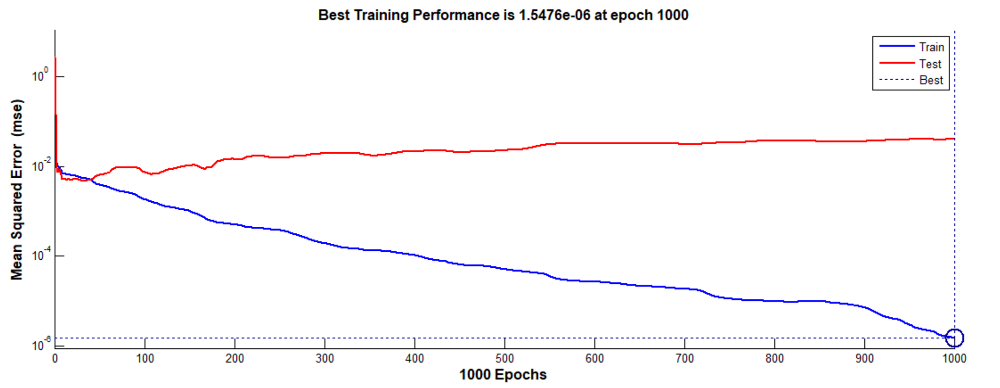

Figure 4. Change in MSE During Run of the Model

Moreover, performance of the model may be understood with respect to R^2 values obtained from the graph developed with predicted and target values in Microsoft Excel. In Figure 5 comparison of predicted and target (actual) values is given in linear graph and in Figure 6 comparison of predicted and actual values is given in scatter graph. R^2 value obtained in scatter graph given in Figure 6 is 0.8411. This value is compatible with result obtained by ANN modelling.

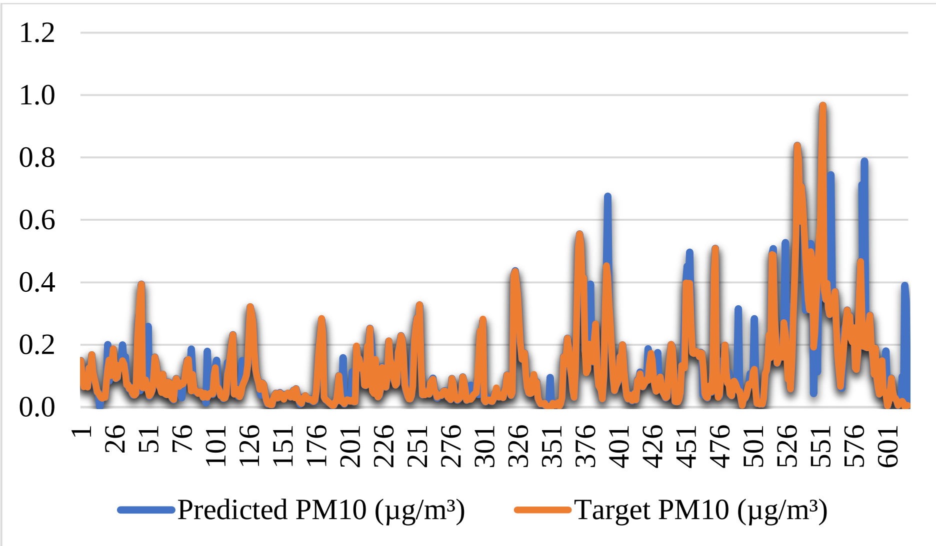

Figure 5. Comparison of Predicted and Target Values (Linear Graph)

For Çorum city of Turkey ANN models were developed for SO_2, CO, CO_2, NO_X, O_2 concentrations emitted from stacks of 16 brick and tile factories and the prediction result for SO_2 concentrations was given R values 0.9915 and 0.9807 respectively for training and testing (Arabacı *et al.*, 2010). According to another study, ANN models developed for Delhi city India and SO_2 concentrations were predicted three different areas such as commercial, industrial and residential. R values of these models were found as 0.68, 0.72, and 0.63, respectively (Chelani *et al.*, 2001). In study conducted in Zonguldak city PM and SO_2 concentrations were predicted for two measurement stations by using ANN. As a result of the study, model developed with ANN gave R values around 0.90 (Tecer, 2007). Moreover, in the study made for city centre of Konya SO_2 concentration was predicted and total response of the ANN model was found as 0.778 (Dursun *et al.*, 2015). With respect to the results of previous works, the result obtained from this study may be considered as good and acceptable.

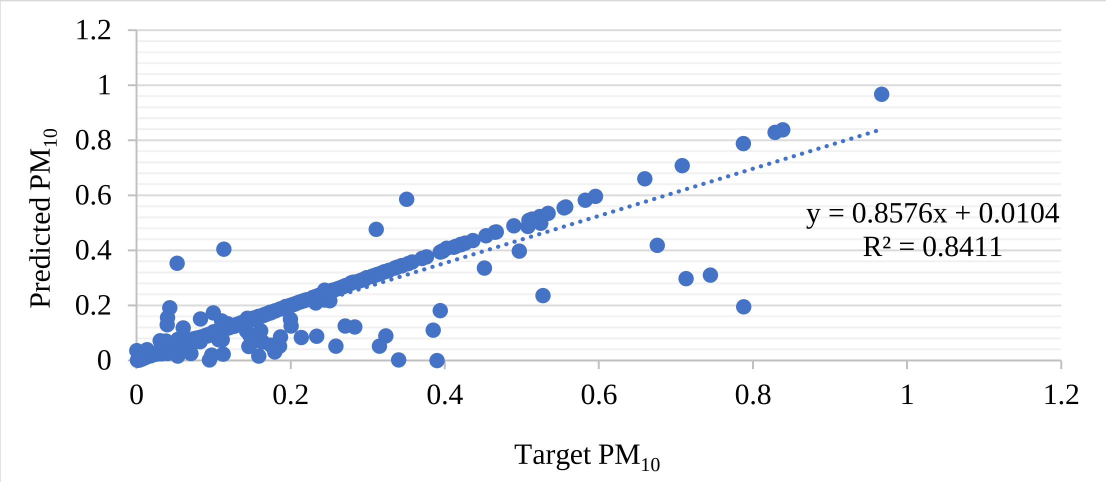

Figure 6. Comparison of Predicted and Target Values (Scatter Graph)

Conclusions

PM_{10} pollution in Karatay district was modelled with ANN in this study. With the model PM_{10} concentration for one-day period was predicted. According to R value obtained from the model, prediction performance was found as 86%. Moreover, R^2 value (0.8411) provided between predicted and actual PM_{10} values supports this result. Besides, when the prediction performance of this study compared with previous studies, it may be considered as acceptable. There are several factors affecting the prediction performance such as abundance or lack of data and under or over training. Thus, for more accurate results such factors may be arranged.

Particulate matter contamination can lead to serious respiratory system damage as it can pass through the bronchi when it reaches high levels. In addition, toxic effects of different pollutants such as heavy metals and hydrocarbons that are trapped in particulate matter are also create problems. For this reason, it is very important to monitor PM pollution and making future predictions for taking necessary precautions before serious air pollution events happen.

Furthermore, many models developed with ANN are able to make short term predictions such as one day and two days, but it may be possible to develop models which can predict longer periods with the help of several efforts such as decreasing time interval between each measurement, increasing measurement duration and integrating more parameter to the model.

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. Environment and Pollution*, **40**, No. 4, 301-321.

Alsugair AM, Al-Qudrah AA, (1998) Artificial neural network approach for pavement maintenance, J. Comput. Civil Eng. ASCE, **2** (4), 249–255.

Arabacı M., Bayram M., Yüceer M. and Karadurmuş E., (2010) Tuğla Ve Kiremit Fabrikalarının Hava Kirliliğine Katkılarının Yapay Sinir Ağı Modellemesi İle Araştırılması, UKMK- 9 – Gazi Üniversitesi, 22-25 Haziran.

Basheer IA, Hajmeer M, (2000) Artificial neural networks: fundamentals, computing, design, and application, *Journal of Microbiological Methods*, **43**(1), 3-31.

Burden F, Winkler D, (2008) Bayesian regularization of neural networks, *Methods Mol Biol.*, **458**, 25-44.

Chelani AB, Chalapati Rao CV, Phadke KM, Hasan MZ, (2001) Prediction of sulphur dioxide concentration using artificial neural networks, *Environmental Modelling & 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 Journal of Environmental Science and Technology*, **12**(12), 3915-3928.

Gardner MW, Dorling SR, (1998) Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences, *Atmospheric Environment*, **32**(14–15), 2627-2636.

Gavin HP, (2017) The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems, Department of Civil and Environmental Engineering Duke University, web page: http://people.duke.edu/~hpgavin/ce281/lm.pdf, retrieval date: 08.09.2017.

Güngör A, (2013) Isparta İlindeki Atmosferde Bulunan Kükürtdioksit (SO_2) ve Partikül Madde (PM) Konsantrasyonunun Çoklu Doğrusal Regresyon Yöntemi İle Modellenmesi, Yüksek Lisans Tezi, Isparta/Türkiye.

Karakaş B, (2015) İç ve Dış Hava Ortamlarında Partiküler Madde (PM10, PM2,5 ve PM1) Konsantrasyonlarının Değerlendirilmesi, Hacettepe Üniversitesi, Fen Bilimleri enstitüsü, Yüksek Lisans Tezi, Ankara.

Møller MF, (1990) A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Computer Science Department University of Aarhus Denmark, web page: ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia353_1s07/papers/moller_90.pdf, retrieval date: 15.09.2017.

Özdemir H, Borucu G, Demir G, Yiğit S, Ak N, (2010) İstanbul'daki Çocuk Oyun Parklarında Partikül Madde (PM2,5 ve PM10) Kirliliğinin İncelenmesi, *Ekoloji* 20, **77**, 72-79.

Ozturk ZC, (2015) Modelling of Atmospheric SO_2 Pollution in Seydişehir Town by Artificial Neural Networks, Master of Science Thesis, Selçuk University, The Graduate School of Natural and Applied Sciences.

Pham DT, Pham PTN, (1999) Artificial intelligence in engineering, International *Journal of Machine Tools and Manufacture*, **39**(6), 937-949.

Resmî Gazete, (2008) Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği, Çevre ve Orman Bakanlığı, Number: 26898, 06 June 2008, Ankara.

Rozlach Z, (2015) Data-driven Modelling in River Channel Evolution Research: Review of Artificial Neural Networks, *J. Int. Environmental Application & Science*, **10**(4), 384-398.

Sarle W, (1997) Neural network frequently asked questions, web page: ftp://ftp.sas.com/pub/neural/FAQ.html, retrieval date: 12.09.2017.

Şekerel BE, Gemicioğlu B., Soriano J. B., (2006) Asthma insights and reality in Turkey(AIRET) study, *Respiratory Medicine*, **100**(10), 1850-1854.

Tecer LH, (2007) Prediction of SO_2 and PM Concentrations in a Coastal Mining Area (Zonguldak, Turkey) Using an Artificial Neural Network, *Polish J. of Environ. Stud.*, **16**(4), 633-638.

Wang SC, (2003) Artificial Neural Network, *The Springer International Series in Engineering and Computer Science*, Volume **743**, 81-100.

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 Bilimler Dergisi*, **8**, 97-112.

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