BENZENE AIR QUALITY FORECASTING IN ROMANIA USING ARTIFICIAL NEURAL NETWORKS

Viviana-Mihaela BOSTAN[1], Brândusa PANTELIMON[2]

The European Environmental Protection Agency has classified benzene as known human carcinogen for all routes of exposure. Since benzene is used as a constituent in motor fuels; as solvent for fats, waxes, resins, oils, inks, paints, plastics, and rubber; in the extraction of oils from seeds and nuts; printing and is also used in the manufacture of detergents, explosives, pharmaceuticals it has been selected in this paper. MLP and Radial Basis Function are the two methods used for predicting the benzene concentration in air. The input data is represented by other very important parameters, like: temperature, humidity and wind speed.

Keywords: air quality, forecasting, ANN

1. Introduction

In Romania like in any other fast developing countries in the world air quality is now one of the main concerns due to present pollution. Due to the fast growing number of cars summed with the fact that part of the vehicles are outdated the health risks of air pollution are extremely serious. Poor air quality increases respiratory ailments like asthma and bronchitis, heightens the risk of life-threatening conditions like cancer, and burdens our health care system with substantial medical costs. [1]

Chronic (long-term) inhalation exposure has caused various disorders in the blood, including reduced numbers of red blood cells and aplastic anemia, in occupational settings. Reproductive effects have been reported for women exposed by inhalation to high levels, and adverse effects on the developing fetus have been observed in animal tests. Increased incidence of leukemia (cancer of the tissues that form white blood cells) have been observed in humans occupationally exposed to benzene. [2]

Air pollution is monitored in our country through the national network for air quality monitorisation. Several urban areas in south of Romania are extremely polluted and therefor air pollution parameters are carefully measured. One of

[1] Eng. PhD Student, University POLITEHNICA of Bucharest, Romania, e-mail: vivianabostan@gmail.com
[2] Prof. PhD, University POLITEHNICA of Bucharest, Romania, Romania, e-mail: brandusa.pantelimon@upb.ro

these parameters is benzene that is found in the air from emissions from burning coal and oil, gasoline service stations, and motor vehicle exhaust.

Acute (short-term) inhalation exposure of humans to benzene may cause drowsiness, dizziness, headaches, as well as eye, skin, and respiratory tract irritation, and, at high levels, unconsciousness.

According to the European Environment Agency air pollution harms human health and the environment. In Europe, emissions of many air pollutants have decreased substantially in recent decades, resulting in improved air quality across the region. However, air pollutant concentrations are still too high, and air quality problems persist. This presents compiled information based on the latest official air pollution data reported by the European Environment Agency's (EEA) member countries.[3] A comprehensive overview of information about Europe's air quality is also published each year by the EEA in the report 'Air quality in Europe'. A number of other publications addressing air pollution are also published by the EEA each year. [4]Air pollutants are emitted from a range of both manmade and natural sources including: burning of fossil fuels in electricity generation, transport, industry, and households; industrial processes and solvent use, for example in the chemical and mining industries; agriculture; waste treatment; natural sources, including volcanic eruptions, windblown dust, sea-salt spray and emissions of volatile organic compounds from plants.t.[5] monoxide and organic compounds coming from incomplete combustion of fossil fuels such as coal and sulfur dioxide to fuel impurities. This smog is caused by combustion in vehicle engines and aircraft fuel which produces nitrogen oxides and unburned fuels releases hydrocarbons. Sunlight makes nitrogen oxides and hydrocarbons to combine and transform oxygen into ozone, a chemical agent that attacks the rubber, injures plants and irritate the lungs. Hydrocarbons are oxidized into substances that condense and form a visible and pervasive fog. Most pollutants are eventually "washed" by rain, snow or fog , but after having traveled long distances, sometimes even continents. As the pollutants are collected in the atmosphere, sulfur and nitrogen oxides are combined and converted to acid rain that falls on lakes and forests leading to death of fish or plants and can affect the entire ecosystems. One of the biggest problems caused by air pollution is global warming, the Earth's temperature increases caused by the accumulation of atmospheric gases such as carbon dioxide. Because of the greenhouse effect is expected that global temperature rise of 1.4 °C to 5.8 ° C by the year 2100. [6]

In our country Ministry of Environment, Water and Forests and the National Environmental Protection Agency are the two competent authorities in the field that are in charge with air quality monitoring and reporting. The national network for Air Quality Monitoring (RNMCA) is composed from 142 stations automatic air quality monitoring stations and 17 mobile stations: 24 traffic stations; 57 stations of industrial type; 37 stations type urban background; 15 suburban stations type; 6 regional background stations type; 3 EMEP stations. Several air pollutants are monitored in our country, like: Sulphur dioxide SO2, Nitrogen oxides NOx (NO/NO2), Ozone O3, Carbon monoxide CO, benzene C6H6, PM10 and PM2.5 particulate matter, Lead and other toxic metals Pb, Cd, As and Hg, Polycyclic aromatic hydrocarbons PAHs. One of the most dangerous air pollutor in our country is benzene. The general features of benzene describe him as an important organic chemical compound with the chemical formula C_6H_6. It smell is aromatic, it's very easy, volatile and water soluble.

The reports in our country show that 90 % of the amount of benzene in the ambient air comes from the road and the remaining 10 % comes from the evaporation of fuel storage and distribution.[7]

The world health organisation reports that exposure to benzeneis a major public health concern.Human exposure to benzene has been associated with a range of acute and long-term adverse health effects and diseases, including cancer and aplastic anaemia. Exposure can occur occupationally and domestically as a result of the ubiquitous use of benzene-containing petroleum products, including motor fuels and solvents. Active and passive exposure to tobacco smoke is also a significant source of exposure. Benzene is highly volatile, and exposure occurs mostly through inhalation.1–3 Public health actions are needed to reduce the exposure of both workers and the general population to benzene. Benzene is found in A1 toxicity Class, known as carcinogenic to humans and has harmful effects on the central nervous system.

The reference method for the measurement of benzene is that of the SR EN 14662 Ambient air quality. Standard method for measurement of benzene in Romania are given by the norms cited in the Law no. 104 from 2011 June the 15^{th} and the limit value for benzen is 5 ug / m3 -that represents the annual limit value for human health protection. In order to predict the benzene concentration it was mandatory to have dayly or weekly data in advance. After that both methods multilayer perceptron and radial basis function can be used to forecast future values for benzene. The main purpose of this paper is to use both neural networks to forecast the concetration level of benzene in the air in south Muntenia (Area 3 on the map).

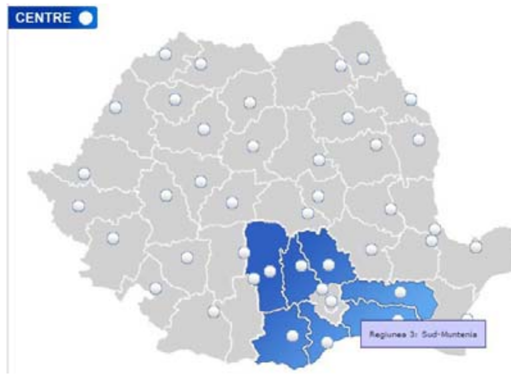

Fig. 1. Area 3- South Muntenia

Another goal of the study was to find out which of the three parameters: temperature, humidity and wind speed have the greatest influence on benzene forecasting. South Muntenia is a part of Romania and it is represented by nine important countys: Arges, Prahova, Ialomita, Calarasi, Teleorman, Dâmbovi□a, Giurgiu, Ilfov and Bucarest. Total area of south Muntenia is 34,453 km2 (14.54 % of the total area of Romania). About 90 % of the region has a continental climate, the rest of the region has a temperate continental climate, with temperatures that decrees along with the altitude from south to nord. Annual average maximum temperatures range between 24 ° C and 27 ° C in summer and between -5 ° C and -10 ° C in winter. Average annual rainfall (calculated across Romania) is 637 mm annually, significantly higher values in mountainous areas 1000-1400 mm / year , and progressively lower east Bărăgan being below 500 mm / year. Summer is a hot season, which lasts from early June to mid-September in the Southern Muntenia. In this region there are over 40 days "tropical" (with temperatures over 30 degrees Celsius) and over 90 summer days (with temperatures above 25 degrees Celsius).

2. Method of investigation

In the present study several air parameters were used, like: the hourly temperature rate, benzene concentration, the relative humidity and the wind speed. The data was collected from the national air quality monitoring network in Romania in the period June-September 2014, since in this period due to the increasing weather temperature, benzene levels have the tendency to grow. Almost eight hundred pieces of data were used for training the network.

Another one hundred and fifty pieces of data served for comparing the observed data with the one resulted from the simulation. Temperature plays the first role in benzene production in real life, the sensitivity analysis of the feed forward neural network MLP has also shown that temperature has a direct effect on predicting benzene level. The goal of this paper is to use both MLP and RBF architectures to predict the concentration of benzene in the air and to find out which of the three parameters mentioned before has the most significant influence on benzene forecasting.

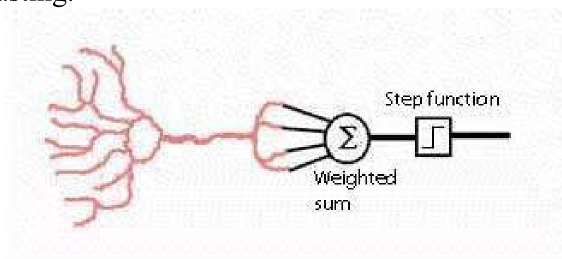

Fig. 2. Network Architecture

The network parameters (weight: w, bias: b) were adjusted via learning rule. „The net input: n „ and the transfer function, f (design choice).

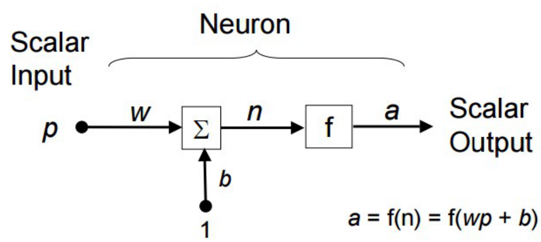

Fig. 3. Network Architecture and Notation-„ Single-Input Neuron”

In the Figure above we can observe the fundamental building block for neural networks: single-input neuron. For transfer we used Tan-Sigmoid Function because it produces an output in the range (-1, 1).

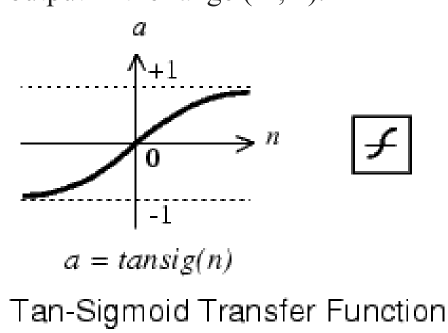

Fig. 4. The transfer function

Therefor the networks tend to be nonlinear. The interface between the data input variable and the model is given by the input layers. It's possible to have models that have not only one but two ore more hidden layers.

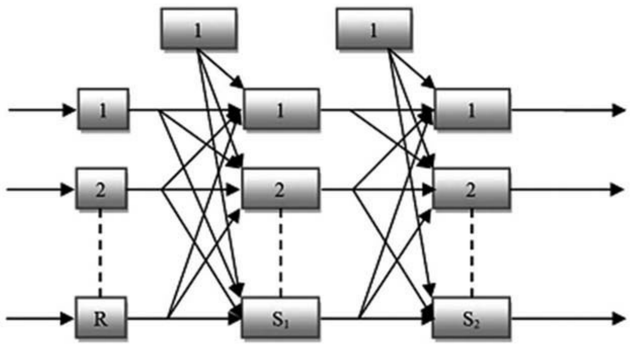

Fig. 5. The MLP with a hidden layer

The number of input vector components is R, P is the input vector. . w1, w2 are weight- ing matrix in hidden and output layers, respectively. b1, b2 are bias vectors. The MLP is a type of static neural network. The number of neurons from the hidden layers is obtained from trial and error. The MLP with a hidden layer, tangent sigmoid transfer function and linear layer outputs can be modeled by next equations:

$$a_j^1(t) = F\left(\sum_{i=1}^{R} w_{i,j}^1 p_i(t) + b_j^i\right) 1 \le j \le S_1 \quad (1)$$

$$a_k^2(t) = G\left(\sum_{j=1}^{S} {}_1 w_{k,j}^2 a_j^1(t) + b_k^2\right) 1 \le k \le S_2 \quad (2)$$

The best performance of the network is depending directly on the number of hidden layers. In this case of study we used an ANN with one and two hidden layers. As transfer function of the hidden layer we have sigmoid tangent and the function of the output layer was considered a linear tangent.

The program for selecting the number of neurons from the hidden layers and to calculate the error number was developed with the help of Matlab software. The cases with a minimum number of errors were selected. Applying the back propagation algorithm is the most sensitive process in order to obtain the perfect learning rate. We aimed to decrease the sum square of outputs. The learning rate is α and determines the velocity of convergence in this algorithm. The performance of the steepest descent algorithm is enhanced if the learning rate is allowed to change during the training process. An adaptive learning rate attempts to make the learning step as big as possible to keep the learning stable and requires some changes in the training procedure. [8]

The sigmoid functions are applied in the MLP and characterized by the fact that their slopes must approach zero as the input gets larger. This causes a problem when steepest descent is applied to train a multilayer network with sigmoid functions, since the gradient can have a very small magnitude; consequently, it can potentially cause small changes in the W and B, although the W and B are far from their optimal values. In order to avoid the saturation of the network all data applied, both input and output were transformed in the range (-1, 1). At the end of the process, all predicted values were transformed back to the real data. For changing the scale of data sequent equation was applied:

$$A_S = \frac{Q_t - A}{B - A} * 2 - 1 \tag{3}$$

Where A_s and O_t are scaled and the observed value of the benzene, temperature, humidity and wind speed at time t, respectively. A and B is the lowest and highest amount of a series of the parameters.

3. Model evaluation

For obtaining the amount of error in predicting of benzene and performance evaluation of the models, we applied Volume Error, Mean Absolute Error, a Root Mean Squared Error and Mean Bias Error which are indicated in the next equations:

$$VE = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{O_t - F_t}{O_t}\right) * 100 \tag{4}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}(O_t - F_t) \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(O_t - F_t)^2} \tag{6}$$

$$MBE = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{F_t - A_t}{A_t}\right) \tag{7}$$

Even if the structure of the RBF is identical to the MLP, the RBF simulates the unknown air quality through a network of Gaussian basis functions in the hidden layer and linear activation functions in the output layer

$$f(x) = e^{x^2/2\sigma^2} \tag{8}$$

Where x is the weighted sum of inputs to the neuron, σ is the width of the basis function, and f(x) is the matching output of the neuron [9]. The structure of RBF neural network has an input layer, a single hidden layer and an output layer. The network has an easy architecture which, at each output node, makes available a linear combination of the outputs of the hidden layer nodes. [10]

The training of RBF is done includes two steps. In the first part basic functions are computed using an algorithm to cluster data in the training set. The self organizing maps (SOMs) or a k means clustering algorithm is most often used. Kohohen SOMs are a form of 'self organizing' neural network that learn to differentiate patterns within input data. Therefore, a SOM will, consequently, cluster input data according to perceived patterns without containing a corresponding output response.[11]

4. Experimental results

The data can be observed in the next table:

Table 1

DATA	BENZENE	TEMP	WIND SPEED	HUMIDITY
AVERAGE	5.52	31	2.42	23
MAX	33.46	38	10,1	29
MIN	1.22	23	0.3	12
COUNT	948	948	948	948

The artificial neural network was trained as shown in next figure is using input parameters which were temperature, wind speed and humidity.

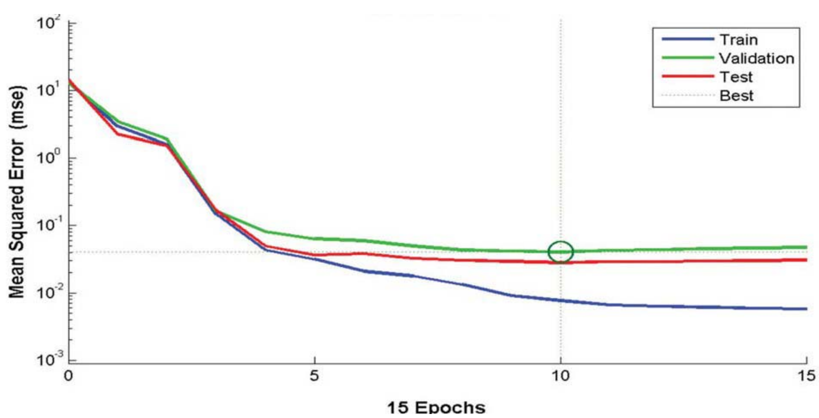

Fig. 6. MLP- Training, validation and testing errors.

The Root Mean Squared Error decreased as iteration was increasing. Since after 10 iteration numbers the numbers of errors was almost constant we stopped the training of the network at 10 iteration numbers. In each epoch (10), input parameter data were introduced to the network to create an output which was benzene.

Next step was to calculate the error of the network. At the end the artificial network parameters were adjusted according to the number of errors.

The figure above shows training, validation and testing errors of the MLP neural network for different iterations.

In the next tables: 2 and 3 a part of the neuron results together with their errors in one and two hidden layers in the MLP neural network can be seen. Since the number of errors in testing of two hidden layers was less than a hidden layer in the MLP, we selected the neural network with two hidden layers, containing 50 and 60 neurons in the first and second layer.

It contained a minimum error for the MLP neural network, Mean Bias Error was 0.209 and Mean Square Error was 0.007.

Table 2

Neurons NO.	VE error	MAE error	RMSE error	Neurons NO.	VE error	MAE error	RMSE error
1	88.32	2.76	3.05	110	38.88	1.83	2.01
5	77.1	2.42	2.68	120	32.3	1.72	1.08
10	84.9	2.62	2.9	130	37.8	1.2	1.31
40	42.3	1.33	1.47	140	35.1	1.1	1.22
50	62.4	1.95	2.15	155	25.4	0.80	0.89
60	45.8	1.43	1.58	160	26.3	0.82	9.93
70	41.06	1.28	1.41	170	25.5	0.84	0.98
80	35.49	1.11	1.23	180	28.7	1.05	1.11
90	34.4	1.05	1.19	190	26.32	0.83	1.02
100	57.6	1.22	1.35	200	27.2	0.87	1.15

Table 3

Neuron number	VE error	MAE error	RMSE error
50-60	20.81	0.66	0.73

The observed and predicted Benzene value data of the MLP network during testing can be seen in figure 7. In the figure above the observed value is given by the horizontal, and the predicted value is the vertical. The coefficient of determination between observed and predicted data was 0.868 and the Index of Agreement was 0.889, which shows the accuracy of the model.

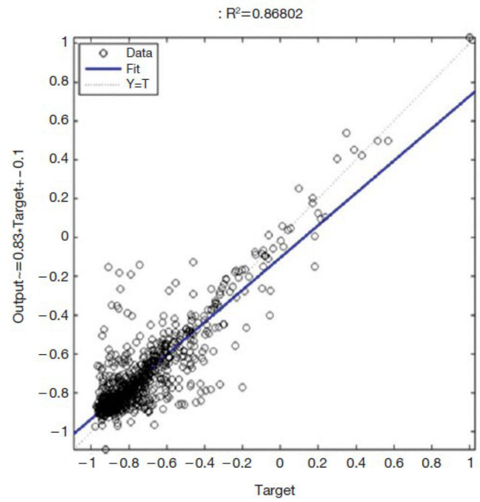

Fig. 7. Normal plot between observed and predicted benzene

Fig. 7 also illustrates a comparison between the observed and the predicted benzene data. As indicated in the figure between the types of data it exists a relatively good agreement. The training point of the network was 800 neurons in the hidden layers. During the training process of the RBF network it's been observed that when iteration numbers (epoch) increase, the errors of the network decline.

The training must be stopped eighther, when the number of errors reaches zero or when the increasing the numbers of epochs brought a constant number of errors. During all the process the amount of error was about 0.1. It has been observed that with error decreasing the performance of the model raises. The most important goal during the training process was to reach zero errors and then to stop it. The coefficient of determination between observed and predicted benzene was 0.907 and the Index of Agreement (IA) was 0.937, which indicates the accuracy of the model, which in turn confirms the suitability of the RBF network.

For sensitivity analysis, we increased and decreased one of the input parameters and the rest of the parameter data remained unchanged. After that, the changed data and two other parameter data were used as a new input to the MLP neural network. Afterward, we determined the new coefficient of determination between new predictions and the observed data for benzene. We continued this procedure for the rest of the parameter data.

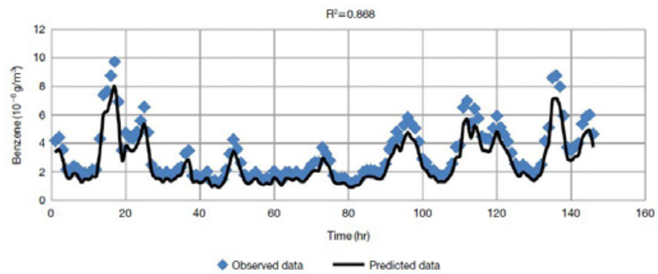

Fig. 8. Observed and preicted benzene value using MLP network

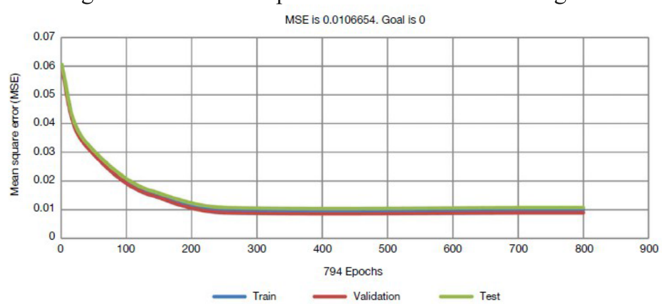

Fig. 9. RBF network performance

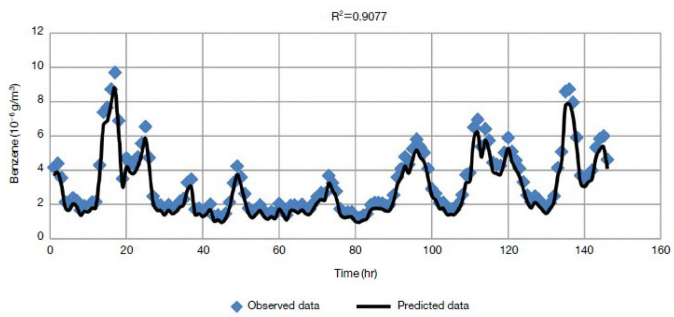

Fig. 10. Observed and preicted benzene value using RBF ntwork

As presented in sensitivity analysis, temperature was the first factor affecting the predicting of benzene since it is volatile compost. Wind speed and humidity were the second and third factor in the prediction of the benzene, respectively. It can be observed that decreasing wind speed indicators affect the forecasting ability of the developed model in comparison to the increasing wind speed.

5. Conclusions

The relative humidity demonstrates that it's not an important role in benzene prediction. The best neural network was declared the MLP neural network with two hidden layers including 50 neurons in the first layer and 60 neurons in the second layer since it returned a minimum error in testing. The

MAE, VE and RMSE were 20.81, 0.66 and 0.73, respectively. The RBF, with a hidden layer, contained a minimum error in training, validation and testing in comparison with the MLP neural network. The RMSE was 0.007. The coefficient of determination between the observed and predicted benzene data for both MLP and RBF neural networks was 0.868 and 0.9077, respectively. The results of the sensitivity analysis indicated that temperature, wind speed and humidity are the first, second and third factors affecting the prediction of benzene, respectively. From the comparison between the results of the MLP with the RBF neural networks in predicting benzene indicates that the forecasting of the RBF is closer. The need for air quality forecasts is clear and the main advantages to use this type of system are: that it is very easy to put in practice, it is not expensive and it uses certified data from the national air quality monitoring system. The main disadvantage of the proposed model is the accuracy in case of increased wind speed.

R E F E R E N C E S

[1] *Schiavon, M., Antonacci, G., Rada, E.C., Ragazzi, M., Zardi, D., 2014. Modelling human exposure to air pollutants in an urban area, Revista de Chimie, 65(1), pp. 61-64EEA* in the report "Air quality in Europe", internet site from EEA.

[2] *Gorai, A.K., Kanchan, Upadhyay, A.,Tuluri, F., Goyal, P., Tchounwou, P.B.*, 2015. An innovative approach for determination of air quality health index . Science of the Total Environment, 533, pp. 495-505.

[3] *Burney, J., Ramanathan, V.*, 2014. Recent climate and air pollution impacts on indian agriculture, Proceedings of the National Academy of Sciences of the United States of America, 111 (46), pp. 16319-16324.

[4] *Bridges, A., Felder, F.A., McKelvey, K., Niyogi, I.* , 2015. Uncertainty in energy planning: Estimating the health impacts of air pollution from fossil fuel electricity generation, Energy Research and Social Science, 6, pp. 74-77.

[5] *Stroe, C.-C., Panaitescu', V.N., Ragazzi, M., Rada, E.C., Ionescu, G.*, 2014. Some considerations on the environmental impact of highway traffic, Revista de Chimie, 65(2), pp. 152-155.

[6] *Hykyšová, S., Brejcha, J.,* 2009. Monitoring of PM10 air pollution in small settlements close to opencast mines in the North-Bohemian Brown Coal Basin , WIT Transactions on Ecology and the Environment ,123, pp. 387-398.

[7] *Ioana Benga*, 2012."Raport anual privind calitatea aerului 2012", www.anpm.ro/anpm_resources/migrated_content/files/APM%20BRASOV/2013/CallitateaAerului2012.pdf, pp.9-56, accessed in may 2013.

[8] *Rahman, N. H. A., Lee, M. H., & Latif, M. T.*, 2014. Artificial neural networks and fuzzy time series forecasting: an application to air quality. Quality & Quantity, pp. 1-15.

[9] *Ragosta, M., M. D'Emilio, and G. A. Giorgio*, 2015."Input strategy analysis for an air quality data modelling procedure at a local scale based on neural network." Environmental monitoring and assessment pp. 1-8

[10] *Pai, Tzu-Yi, Moo-Been Chang, and Shyh-Wei Chen*, 2015."Application of Computational Intelligence on Analysis of Air Quality Monitoring Big Data." Information Granularity, Big Data, and Computational Intelligence. Springer International Publishing. 427-441.

[11] *Kohonen, Teuvo* 1982 ."Self-Organized Formation of Topologically Correct Feature Maps".Biological Cybernetics 43(1): 59–69.