# Air Pollutants and Meteorological Parameters Influence on $PM_{2.5}$ Forecasting and Performance Assessment of the Developed Artificial Intelligence-Based Forecasting Model

**MARIAN POPESCU*, SANDA FLORENTINA MIHALACHE, MIHAELA OPREA**

Petroleum-Gas University of Ploiesti, Automatic Control, Computers and Electronics Department, 39 Bucuresti Blvd.,100680, Ploiesti, Romania

*Particulate matter with an aerodynamic diameter lower than 2.5 µm ($PM_{2.5}$) is one of the most important air pollutants. Current regulations impose measuring and limiting its concentrations. Thus, it is necessary to develop forecasting models programs that can inform the population about possible pollution episodes. This paper emphasizes the correlations between $PM_{2.5}$ and other pollutants, and meteorological parameters. From these, nitrogen dioxide and temperature showed have the best correlations with $PM_{2.5}$ and have been selected as inputs for the proposed forecasting model besides four $PM_{2.5}$ concentrations (the values from current hour to three hours ago), the output of the model being the prediction of the next hour $PM_{2.5}$ concentration. Two methods from artificial intelligence were used to build the forecasting model, namely adaptive neuro-fuzzy inference system (ANFIS) and artificial neural networks (ANN). The comparative study between these methods showed that the model which uses ANN have better results in terms of statistical indicators and computational effort.*

*Keywords: air pollution, particulate matter, forecasting model, artificial intelligence techniques*

Climate change is an important problem nowadays. A main component of this problem is air pollution with its multiple causes. National and international management air quality control programs and strategies were developed to control and reduce air pollution, which is a major concern for human health.

According to Environmental Protection Agency from United States, there are six common air pollutants: particulate matter, ground-level ozone, carbon monoxide, sulfur oxides, nitrogen oxides, and lead. These pollutants are selected according to the importance of their influence on human health and environment.

Particulate matter is a critical air pollutant, especially in South-East Asia [1], where pollution episodes occur with relatively high frequency. Particulate matter is formed by all solid and liquid particles suspended in air. These particles vary in size, composition, and origin and are usually denoted according to their diameter as coarse particulates - $PM_{10}$ (less than 10 µm), fine particulates - $PM_{2.5}$ (less than 2.5 µm) and ultrafine particulates - $PM_1$ (less than 1 µm). From these categories, $PM_{2.5}$ is inhalable and can directly penetrate human lungs. $PM_{2.5}$ air pollutant originates from two types of sources i.e., primary (direct air pollution sources, as e.g. industry, traffic, combustion etc.), and secondary (as an effect of some processes that take place in the atmosphere). A pollution episode can be characterized by poor visibility and air quality and also generates health problems. The $PM_{2.5}$ impact indicators (health effects and legal compliance) are described in [2-4], health effects - on short term as an increase of respiratory problems are presented in [5-7] and on long term as chronic diseases can be found in [8-10].

The negative effect of $PM_{2.5}$ pollution episodes on human health required the implementation of a strategy for air quality management. Therefore, there are developed standards in EU, China and US that limit the $PM_{2.5}$ year average. A current problem is that there are no sufficient measurements of $PM_{2.5}$. The lack of fine particulate matter data generated the need of correlation models of $PM_{2.5}$ with other measured pollutants ($NO_2$, CO and $SO_2$), visibility or meteorological parameters.

Recent researches emphasized the correlation between $PM_{2.5}$ and $NO_2$ using data from China [11], India [12] and 20 European areas [13]. In the latter study it is also presented the correlation of $PM_{2.5}$ with nitrogen oxides $NO_x$. The identification of air pollutants emission source locations and transportation trends are studied in [14] for a US region, where industrial and transportation activities are inducing increasing levels of $PM_{2.5}$. The air pollutants that were considered are $SO_2$ and $NO_x$. Another correlation is between $PM_{2.5}$ and $PM_{10}$ [15]. It is also emphasized the role of meteorological parameters. The main meteorological parameters that influence $PM_{2.5}$ are temperature, relative humidity and wind direction [16-21]. Aerosol optical depth is a measure of the sunlight blocking by the particles in the atmosphere and consequently correlated with $PM_{2.5}$ [22-24].

Most of the obtained correlation models use multiple regressions in order to compute $PM_{2.5}$ values. Artificial intelligence based methods can be used in order to model correlations or forecast the $PM_{2.5}$ concentration. A research work that applies an artificial intelligence based approach (ANFIS) to forecast $PM_{2.5}$ during haze episodes in a case study from Delhi India, and uses temperature and relative humidity is described in [25]. Three case studies from Romania that use ANFIS model to forecast the PM concentration are presented in [26]. A complex study of artificial intelligence methods used for $PM_{2.5}$ forecasting can be found in [27]. Relative humidity and temperature correlations and their effect on $PM_{2.5}$ forecasting are described in [28]. An inductive learning approach for PM forecasting is presented in [29].

The objective of this study is to find a meaningful correlation of $PM_{2.5}$ concentration with other air pollutants' concentrations and meteorological parameters for a Romanian city (Ploiesti) and to develop a forecasting model using artificial intelligence techniques such as ANFIS and

* email: mpopescu@upg-ploiesti.ro

ANN. The solution presented in this study is a part of the decision support system of the ROKIDAIR research project (http://www.rokidair.ro). Under the ROKIDAIR project it is developed an intelligent system [30] for particulate matter concentration monitoring, analysis and forecasting in two cities from Romania (Ploiesti and Târgoviste), which provides expert early warnings [31] to inform the public when pollution episodes occur, in order to protect children with health problems (e.g. respiratory problems).

*Proposed method*

In this study, data from a Ploiesti (Romania) air quality monitoring station (PH-2) were used. Approximately 2000 samples of pollutants concentrations and meteorological parameters were considered. The pollutants taken into consideration in the initial part of the study were carbon monoxide (CO), nitrogen monoxide (NO), nitrogen dioxide ($NO_2$), nitrogen oxides ($NO_x$), ozone ($O_3$), sulfur dioxide ($SO_2$), particulate matter 2.5 ($PM_{2.5}$) and the meteorological parameters were the temperature (T) and relative humidity (RH).

The first part of the study tries to determine which pollutant and meteorological parameter are more correlated with $PM_{2.5}$. The second part describes the model used to forecast $PM_{2.5}$ short-term concentration evolution. A selection of the $PM_{2.5}$ concentration evolution compared to different pollutants or meteorological parameters evolutions is presented in figures 1-7.

Besides the comparisons from figures 1-7, in order to determine the correlation between $PM_{2.5}$ and the above mentioned parameters, Spearman's correlation coefficient - ρ [32] was calculated, because of the nonlinear dependence of $PM_{2.5}$ to the other considered parameters, the results being presented in table 1. For this coefficient, values closer to 1 or -1 indicate a moderate to good correlation.

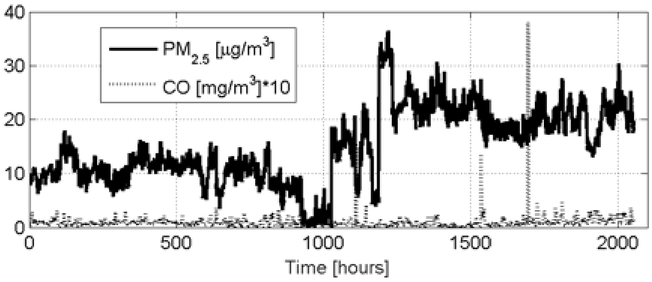

Fig. 1. Time series of $PM_{2.5}$ and CO concentrations

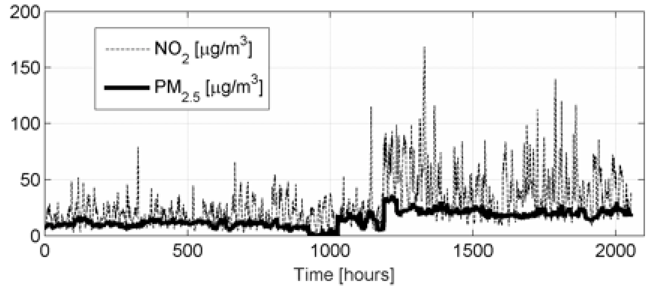

Fig. 2. Time series of $PM_{2.5}$ and $NO_2$ concentrations

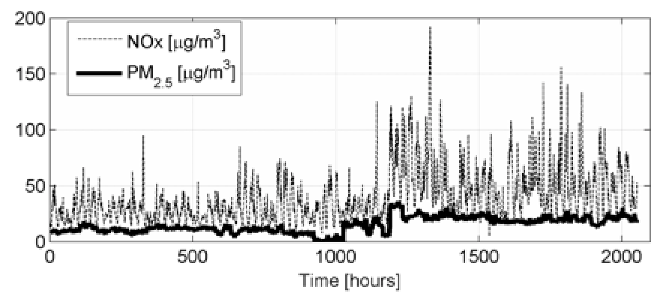

Fig. 3. Time series of $PM_{2.5}$ and $NO_x$ concentrations

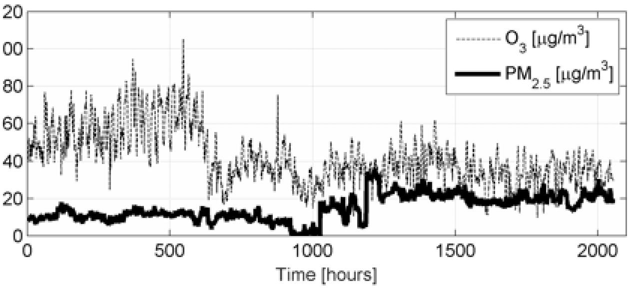

Fig. 4. Time series of $PM_{2.5}$ and $O_3$ concentrations

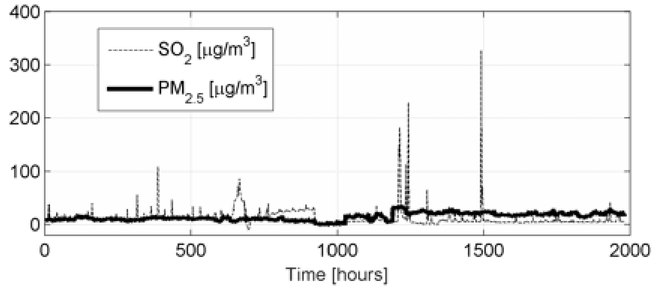

Fig. 5. Time series of $PM_{2.5}$ and $SO_2$ concentrations

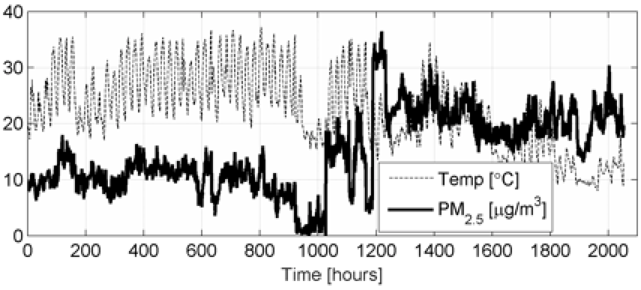

Fig. 6. Time series of $PM_{2.5}$ concentration and temperature

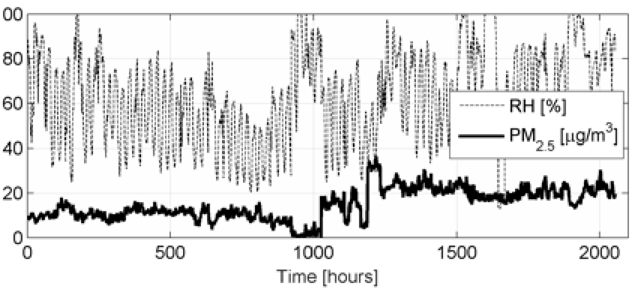

Fig. 7. Time series of $PM_{2.5}$ concentration and relative humidity

**Table 1**
SPEARMAN'S CORRELATION COEFFICIENT

| Parameter | ρ |
|---|---|
| CO | 0.073 |
| $NO_2$ | 0.573 |
| NO | -0.031 |
| $NO_x$ | 0.517 |
| $O_3$ | -0.232 |
| $SO_2$ | -0.365 |
| Temp | -0.379 |
| RH | 0.195 |

By analyzing the figures 1-5 and the values of the Spearman's correlation coefficient from table 1, it can be concluded that the most correlated pollutant with $PM_{2.5}$ is $NO_2$. Meanwhile, from the meteorological parameters, the temperature is more correlated with $PM_{2.5}$ than the relative humidity.

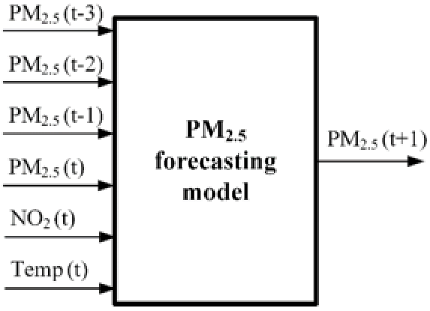


Fig. 8. Input-output representation of the $PM_{2.5}$ forecasting model

Taking into consideration these conclusions, the proposed $PM_{2.5}$ forecasting model in this study has the structure from figure 8.

The inputs of the forecasting model are: the $PM_{2.5}$ concentration measurements from current hour to three hours ago, current hour $NO_2$ concentration measurements, and current hour temperature. The output of the model is represented by the prediction of the next hour $PM_{2.5}$ concentration.

The proposed forecasting model is based on two artificial intelligence methods, namely: adaptive neuro-fuzzy inference system (ANFIS) and artificial neural networks (ANN). The results obtained with these methods are presented in the following section.

**Experimental part**

Both forecasting methods use normalized data for all inputs ($PM_{2.5}$ concentrations, $NO_2$ concentrations, Temperature). The data set was divided in training data (70%), validation data (15%) and testing data (15%).

In both cases, the next hour forecasted $PM_{2.5}$ concentration is compared with the measured value, the model accuracy being evaluated by calculating statistical indicators such as root mean square error (RMSE), index of agreement (IA) and coefficient of determination ($R^2$).

*ANFIS method*

The adaptive neuro-fuzzy inference system (ANFIS) is a general modelling technique, combining learning capabilities of ANNs and the linguistic approach of fuzzy logic theory [33]. Using the ANN technique to compute the parameters of the fuzzy inference system model, the FIS part is enhanced with the ability to learn from training data. The number of neurons in the hidden layer is fixed in the FIS structure of the ANFIS. This eliminates the experimental search of the best number of neurons in the hidden layer of ANN.

The ANFIS method from this study use triangular or Gaussian membership functions for inputs and constant type output. The fuzzy inference system structure generation use grid partition method, and the learning algorithm use back-propagation or hybrid optimization methods.

The shape of the inputs membership functions (Trimf or Gauss) and the type of the learning optimization methods (back-propagation or hybrid) are the parameters, which can be modified, thus resulting four combinations that will be tested through simulation. The results are presented in table 2.

**Table 2**

STATISTICAL INDICATORS FOR THE ANFIS MODEL

| ANFIS structure | RMSE [μg/m$^3$] | IA | $R^2$ |
|---|---|---|---|
| Gauss/Back-prop. | 2.7095 | 0.9555 | 0.8319 |
| Gauss/Hybrid | 3.5277 | 0.9327 | 0.7151 |
| Trimf/Back-prop. | 3.1143 | 0.9414 | 0.7780 |
| **Trimf/Hybrid** | **2.3892** | **0.9688** | **0.8693** |

From table 2, it can be seen that the best ANFIS structure use triangular membership functions for inputs and hybrid optimization method associated to the learning algorithm.

Figures 9 and 10 show the testing error for the best ANFIS structure and the comparison between the predicted and the actual values of the next hour $PM_{2.5}$ concentration.

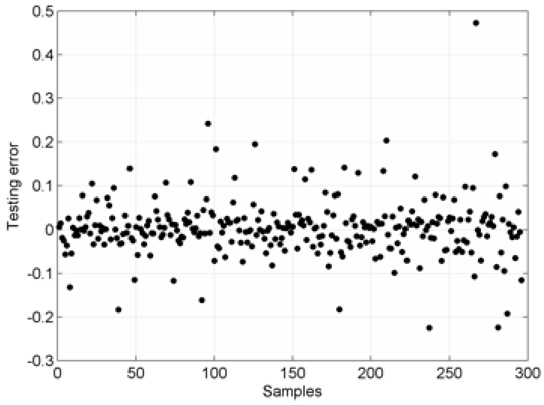


Fig. 9. Testing error for the best ANFIS structure

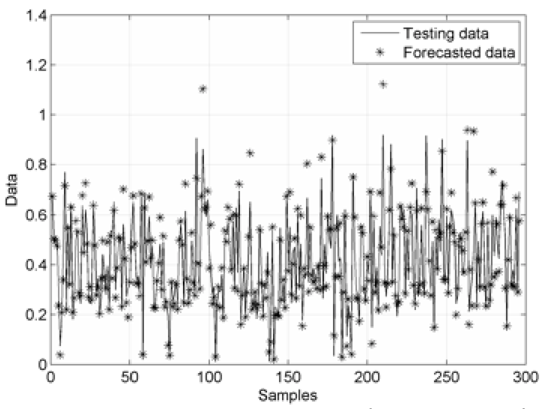


Fig. 10. Comparison between testing and forecasted data for the best ANFIS structure

*ANN method*

Artificial neural networks are models that estimate or approximate unknown functions with large number of inputs. Neural networks are usually organized in layers composed of interconnected processing elements, called neurons [34].

The ANN structure from this study contains an input layer with six neurons, a hidden layer (with different number of neurons) and an output layer with one neuron. The type of the neural network is feed-forward back-propagation with Levenberg-Marquardt training algorithm. The used adaptive learning functions were gradient descent weight and bias (Learngd) and gradient descent with momentum weight and bias (Learngdm).

The modifiable parameters in this case were the number of neurons in the hidden layer (from four to ten) and the adaptive learning functions. The results of the simulations, in terms of statistical indicators, are presented in table 3.

By analyzing the results from table 3, it can be observed that the best statistical indicators (smallest RMSE, biggest IA and $R^2$) are obtained for the ANN structure with 8 neurons in the hidden layer and gradient descent with momentum weight and bias adaptive learning function (Learngdm).

From figures 11 and 12, it can be observed that the testing error is much smaller and is distributed more compactly than the testing error from the ANFIS case, and the predicted values of the next hour $PM_{2.5}$ concentration are much closer to the actual values in the case of ANN compared to the ones from figure 10.

| ANN structure | RMSE [μg/m³] | IA | R2 |
|---|---|---|---|
| 6x4x1/Learngd | 1.2812 | 0.9906 | 0.9624 |
| 6x4x1/Learngdm | 1.2668 | 0.9908 | 0.9633 |
| 6x5x1/Learngd | 1.2600 | 0.9908 | 0.9637 |
| 6x5x1/Learngdm | 1.2628 | 0.9908 | 0.9635 |
| 6x6x1/Learngd | 1.2553 | 0.9909 | 0.9639 |
| 6x6x1/Learngdm | 1.2752 | 0.9906 | 0.9628 |
| 6x7x1/Learngd | 1.2550 | 0.9909 | 0.9639 |
| 6x7x1/Learngdm | 1.2645 | 0.9908 | 0.9634 |
| 6x8x1/Learngd | 1.2745 | 0.9907 | 0.9628 |
| **6x8x1/Learngdm** | **1.2546** | **0.9909** | **0.9640** |
| 6x9x1/Learngd | 1.2907 | 0.9904 | 0.9619 |
| 6x9x1/Learngdm | 1.2570 | 0.9909 | 0.9638 |
| 6x10x1/Learngd | 1.2695 | 0.9907 | 0.9631 |
| 6x10x1/Learngdm | 1.2760 | 0.9906 | 0.9627 |

**Table 3**
STATISTICAL INDICATORS FOR THE ANN MODEL

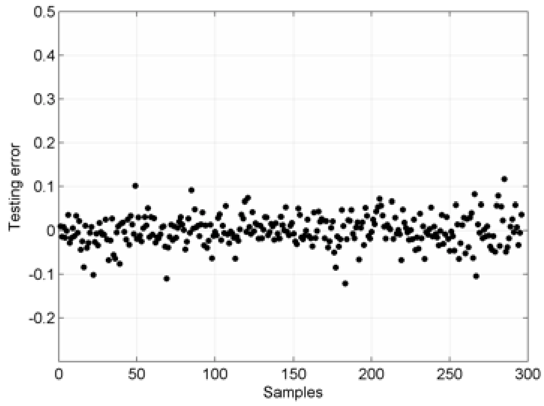

Fig. 11. Testing error for the best ANN structure

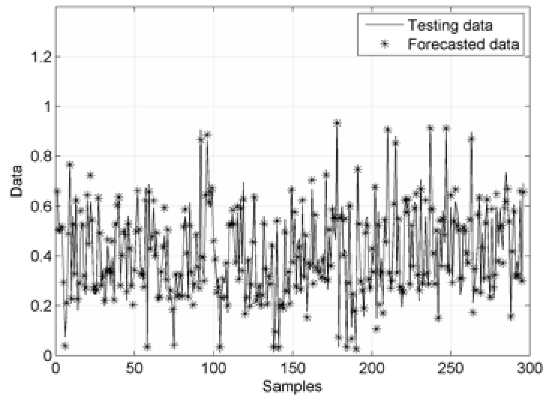

Fig. 12. Comparison between testing and forecasted data for the best ANN structure

## Conclusions

The paper proposes a forecasting model for $PM_{2.5}$ concentration. Among the most important factors that could influence $PM_{2.5}$ were considered the pollutants CO, NO, $NO_2$, $NO_x$, $O_3$, $SO_2$, and the meteorological parameters temperature and relative humidity. In order to determine the correlations between $PM_{2.5}$ and the above mentioned parameters, the time series of PM2.5 and these parameters were compared and Spearman's correlation coefficient was calculated. This analysis determines that $NO_2$ and temperature have the best correlations with $PM_{2.5}$. Consequently, these parameters are used as inputs of the proposed $PM_{2.5}$ forecasting model besides four values of $PM_{2.5}$ concentrations (from current hour to three hours ago), the output of the model being the prediction of the next hour $PM_{2.5}$ concentration. The forecasting model is tested using two artificial intelligence techniques, ANFIS and ANN. From the simulation results, it can be concluded that the artificial neural network method provided better results than ANFIS in terms of statistical indicators (RMSE, IA and $R^2$) calculated by comparing the predicted values of the next hour $PM_{2.5}$ concentration with its actual values. Also, the computational effort is much smaller (minutes) in the case of ANN compared to ANFIS effort (hours).

*Acknowledgement: The research leading to these results has received funding from EEA Financial Mechanism 2009-2014 under the project ROKIDAIR Towards a better protection of children against air pollution threats in the urban areas of Romania\ contract no. 20SEE/30.06.2014. Special thanks to Prof. 'tefania Iordache and Mr. Trond Bøhler for transforming a breakthrough idea into a research project.*

## References

1.SHEN, Z., CAO J., ZHANG, L., ZHANG, Q., HUANG, R., LIU, S., et al., Atmospheric Environment, **126**, 2016, p. 15.
2.KIESEWETTER, G., SCHOEPP, W., HEYES, C., AMANN, M., Environmental Modelling & Software, **74**, 2015, p. 201.
3.MIRCEA, I., FALUP, O., IVAN, R., SAMOILA, I., IONEL, I., REV. CHIM. (Bucharest), **66**, No. 2, 2015, p. 247.
4.CAPATINA, C., SIMONESCU, C.M., Rev. Chim. (Bucharest), **63**, No. 12, 2012, p. 1289.
5.KAUR, S., NIEUWENHUIJSEN, M., COLVILE, R., Atmospheric Environment, **39**, No. 20, 2005, p. 3629.
6.LUO, J.Y., CHEN, Z., HU, J.Q., Respirology, **20**, No. 12, 2015.
7.DUNEA, D., IORDACHE, S., LIU, H.-Y., BØHLER, T., POHOATA, A., RADULESCU, C., Environ. Sci. Pollut. Res., **23**, 2016, p. 15395.
8.MIMURA, T., ICHINOSE, T., YAMAGAMI, S., FUJISHIMA, H., KAMEI, Y., GOTO, M., et al., Science of the Total Environment, **487**, No. 1, 2014, p. 493.
9.LIN, C., LI, Y., LI, Z., LI, C., LAU, AKH, DENG, X., et al., Remote Sensing of Environment, **179**, 2016, p. 13.
10.HU, Z., International Journal of Health Geographics, **8**, No. 1, 2009, p. 1.
11.LIU, C., HENDERSON, B.H., WANG, D., YANG, X., PENG, Z., Science of the Total Environment, **565**, 2016 p. 607.
12.KUMAR, R., JOSEPH, A.E., Environmental Monitoring and Assessment, **119**, No. 1, 2006, p. 191.
13.PENNANEN, A., KATSOUYANNI, K., BELLANDER, T., LANKI, T., INEICHEN, A., GRIVAS, G., et al., Atmospheric Environment, **62**, 2012, p. 303.
14.YERRAMILLI, A., DODLA, V.B.R., CHALLA, V.S., MYLES, L.T., et al., Air Quality Atmospheric Health, **5**, 2012, p. 401.

15.ZHOU, X., CAO, Z., MA, Y., WANG, L., WU, R., WANG, W., Chemosphere, **144**, 2016, p. 518.
16.DEGOBBI, C., LOPES, F.D.T.Q.S., CARVALHO-OLIVEIRA, R., MUÑOZ, J.E., SALDIVA, PHN., Atmospheric Environment, **45**, No. 13, 2011, p. 2277.
17.TAI, A.P.K., MICKLEY, L.J., JACOB, D.J., Atmospheric Environment, **44**, Issue 32, 2010, p. 3976.
18.PATERAKI, S., ASIMAKOPOULOS, D.N., FLOCAS, H.A., MAGGOS, T., VASILAKOS, C., Science of the Total Environment, **419**, 2012, p. 124.
19.ZHANG, Z., ZHANG, X., GONG, D., QUAN, W., ZHAO, X., MA, Z., et al., Atmospheric Environment, **108**, 2015, p. 67.
20.WANG, J., WANG, Y., LIU, H., YANG, Y., ZHANG, X., LI, Y., et al. , Atmospheric Environment, **81**, 2013, p. 158.
21.ZHANG, C., NI, Z., NI, L., Physica A: Statistical Mechanics and its Applications, **438**, 2015, p. 114.
22.LI, H., FARUQUE, F., WILLIAMS, W., AL-HAMDAN, M., LUVALL, J., CROSSON, W., et al., Atmospheric Environment, **43**, No. 28, 2009, p. 4303.
23.KONG, L., XIN, J., ZHANG, W., WANG, Y., Environmental Pollution, **216**, 2016, p. 350.
24.XIN, J., ZHANG, Q., WANG, L., GONG, C., WANG, Y., LIU, Z., et al., Atmospheric Research, **138**, 2014, p. 179.
25.MISHRA, D., GOYAL, P., UPADHYAY, A., Atmospheric Environment, **102**, 2015, p. 239.
26.MIHALACHE, S.F., POPESCU, M., OPREA, M., 19th International Conference on System Theory, Control and Computing (ICSTCC), 14-16 oct. 2015, p. 895.
27.OPREA, M., MIHALACHE, S.F., POPESCU, M., Proceedings of the 6th International Conference on Computers Communications and Control (ICCCC), 10-14 May 2016, Oradea, Romania, p. 103.
28.OPREA, M., MIHALACHE, S.F., POPESCU, M., 12th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2016), 16-18 sept. 2016, Thessaloniki, Grecia, p. 204.
29.OPREA, M., DRAGOMIR, E.G., POPESCU, M., MIHALACHE, S.F., Rev. Chim. (Bucharest), **67**, no. 10, 2016, p. 2075.
30.OPREA, M., IANACHE, C., MIHALACHE, S., DRAGOMIR, E., DUNEA, D., IORDACHE, S., SAVU, T., Proceedings of the 19th International Conference on System Theory, Control and Computing (ICSTCC), Cheile Gradistei, Romania, 2015, p. 711.
31.IORDACHE, S., DUNEA, D., LUNGU, E., PREDESCU, L., DUMITRU, D., IANACHE, C., IANACHE, R., Proceedings of the 20th International Conference on Control Systems and Computer Science (CSCS20–2015), Bucharest, Romania, 27-29, 2015, p. 789.
32.MYERS, J.L., WELL, A., LORCH, R.F., Research Design and Statistical Analysis, Routledge, 2010.
33.JANG, R., IEEE Trans. on Systems, Man and Cybernetics, **23** (3), 1993, p. 665.
34.OPREA, M., Inteligenã artificiala- Elemente fundamentale, vol. I, Ed. Universal Cartfil, Ploiesti, 1998.

Manuscript received: 10.01.2017