2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, Cheile Gradistei, Romania

On the Development of an Intelligent System for Particulate Matter Air Pollution Monitoring, Analysis and Forecasting in Urban Regions

Mihaela Oprea[1*], Cornel Ianache[1], Sanda Florentina Mihalache[1], Elia Georgiana Dragomir[2]
[1]Automatic Control, Computers and Electronics Department
[2]ITIMF Department
Petroleum-Gas University of Ploiesti
Ploiesti, Romania
mihaela@upg-ploiesti.ro

Daniel Dunea, Stefania Iordache
Faculty of Environmental Engineering and Food Science
VALAHIA University of Targoviste
Targoviste, Romania

Tom Savu
Centre for Advanced Technologies - CTANM
POLITEHNICA University of Bucharest
Bucharest, Romania

***Abstract*—The paper presents details on the development of an intelligent system for particulate matter (PM) air pollution monitoring, analysis and forecasting in two pilot cities, Ploiesti and Targoviste, in selected areas near schools, kindergartens and pediatric hospitals. The main purpose of the system is to provide expert early warnings to protect children with health problems. An in-situ PM_{10}, $PM_{2.5}$ monitoring network was designed and an online $PM_{2.5}$ monitoring network is under development for each city. Two case studies of PM air pollution analysis and forecast for the Ploiesti city are described. The PM air pollution analysis, forecasting, and children health impact analysis were made by using techniques from statistics and artificial intelligence.**

Keywords—intelligent system; PM air pollution monitoring and analysis; air pollution forecasting; artificial neural networks;

I. Introduction

Air pollution is a major problem in urban regions due to its potential important effects on population health, especially, during severe air pollution episodes that need real time expert monitoring, analysis, forecasting and alerting systems [1]. Among the main air pollutants, particulate matter (PM) has a greater impact on children health, especially the respirable PM fractions: PM_{10} and $PM_{2.5}$, i.e. PMs with a diameter smaller than 10 micrometers and 2.5 micrometers, respectively. Most of the countries have created regulations for the ambient concentration of PMs. For example, in European Union the emission standards include limits for PM_{10} (50 $\mu g/m^3$, daily average) and $PM_{2.5}$ (25 $\mu g/m^3$, yearly average). Several countries are monitoring PM_{10} and $PM_{2.5}$ under their national air quality monitoring network and provide real time forecasts of the air quality index (AQI) for their major cities (e.g. AirNow in USA – http://airnow.gov/; Air quality now in Europe – http://www.airqualitynow.eu/; Beijing Air Pollution – http://aqicn.org/city/beijing/). Current research trends are focused on the development of more accurate real time AQI forecasts and alerting software tools. In Romania, few stations from the National Air Quality Monitoring Network (RNMCA – http://www.calitateaer.ro/) are monitoring $PM_{2.5}$, and less of them are located in areas where children are living and studying. Moreover, there is no real time AQI forecasts and no online early warning/alert system that inform parents, teachers, schools and kindergartens managers, and medical personnel when PM air pollution episodes will occur, with a potential important impact on children health, especially, on sensitive children (e.g. with respiratory problems). The ROKIDAIR project (http://www.rokidair.ro/) is currently developing an intelligent system for $PM_{2.5}$ air pollution monitoring, analysis, forecasting and alerting in two pilot cities, Ploiesti and Targoviste, its main purpose being the protection of children health when severe $PM_{2.5}$ air pollution episodes occur in urban areas with schools, kindergartens and children hospitals. The cyberinfrastructure of the system was introduced in [2].

II. The ROKIDAIR Intelligent System Description

A. The ROKIDAIR System Architecture

Fig. 1 presents the block diagram of the ROKIDAIR intelligent system. The main components of the system are: a local monitoring and analysis system for each pilot city, Ploiesti and Targoviste, the web-based GIS module, the ROKIDAIR Decision Support System (DSS), the ROKIDAIR databases, the Web Service with a data server and a real time monitoring and service of the ROKIDAIR $PM_{2.5}$ micro-stations.

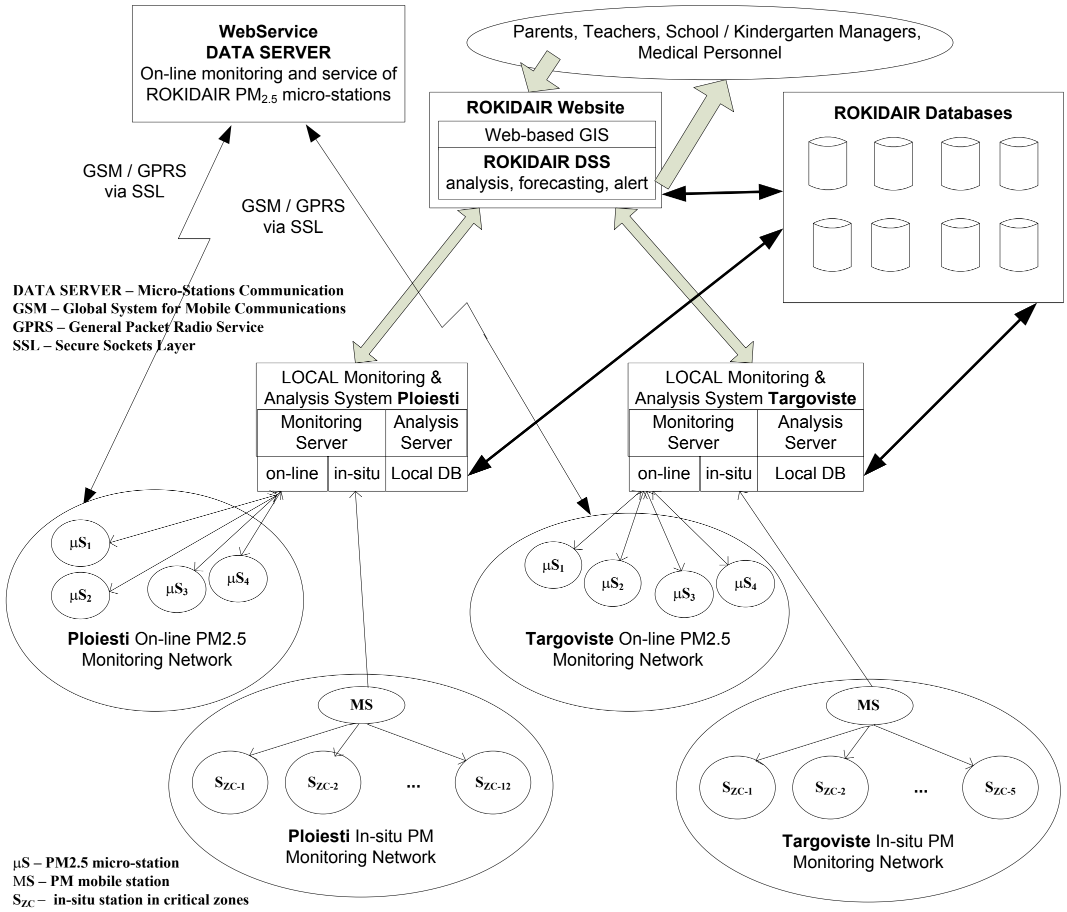

Fig. 1. The block diagram of the ROKIDAIR system

Eight databases will be included in the system structure: two databases with on-line (continuous) $PM_{2.5}$ monitoring data and two databases with in-situ PM monitoring data for both pilot cities, the meteorological database, a medical database [3], the AirBASE historical database and the ROKIDAIR potential users database.

B. The PM Monitoring Networks

The ROKIDAIR system will include four PM monitoring networks: two on-line $PM_{2.5}$ monitoring networks and two in-situ PM monitoring networks, for both pilot cities. The on-line (continuous) $PM_{2.5}$ monitoring network of each pilot city is under development and will contain four $PM_{2.5}$ micro-stations ($\mu S_{1,2,3,4}$) that will be located inside or nearby schools, and kindergartens. The general structure of a $PM_{2.5}$ micro-station is presented in Fig. 2. The modules of a micro-station are: the measurement module containing sensors for atmospheric pressure, indoor and outdoor temperature, relative humidity, $PM_{2.5}$ concentration level, and a data acquisition unit (DAU), the control module (a programmable automation controller – PAC), the communication module (GSM/GPRS) and the micro-station climate control module. The WebService of the ROKIDAIR system will perform a real time monitoring and service of all $PM_{2.5}$ micro-stations. In cases of some micro-stations malfunctions or faults, the WebService will detect them and will change remotely the micro-stations settings (e.g. it will do the internal clock setting, will perform corrections of some micro-station parameters, as well as auto-calibration) by using some expert rules. At present, a first prototype of a ROKIDAIR micro-station was developed. The in-situ PM monitoring network was designed so that it includes a set of measurement points selected from critical zones of each pilot city, situated near schools, kindergartens, children hospitals, residential areas and traffic or industrial air polluted zones.

The in-situ PM monitoring network for the Ploiesti city contains 12 measurement points, PH1÷PH12 (shown in Fig. 3), while for the Targoviste city it contains 5 measurement points. The $PM_{2.5}$ concentration level is measured with a mobile monitoring station (MS). During the monitoring campaigns performed in 2014 a Casella Microdust Pro mobile station was used, while currently it is used a DustTrak mobile station.

At each measurement point, the mobile station records the $PM_{2.5}$ concentration level (maximum and mean values for 15 minutes monitoring time) and some meteorological parameters (wind speed and temperature). Thematic maps are built for each city under the web-based GIS module of the ROKIDAIR system. The in-situ measurements are used to build isolines of $PM_{2.5}$ peak concentrations.

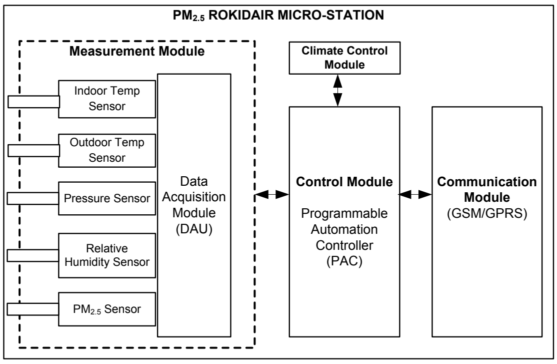

Fig. 2. The general structure of a ROKIDAIR $PM_{2.5}$ micro-station

Fig. 3. The 12 measurements points (in critical zones) from the in-situ PM monitoring network of the Ploiesti city (PH1÷PH12)

C. *The ROKIDAIR Decision Support System*

The current version of the ROKIDAIR DSS is composed by the following modules: the air pollution analysis module, a knowledge-based system (including a knowledge base with expert knowledge used for air quality analysis, forecasting and alerting modules decision support), a forecasting module and an alert / early warning module, as shown in Fig. 4.

The air pollution analysis module performs various analyzes: the air pollution standards checking (e.g. the law 104/15.06.2011), an elemental analysis (heavy metals), time series analysis (e.g. spectrum analysis). The knowledge based system was designed to provide knowledge under the form of facts (derived by a reasoning mechanism) for the forecasting module and for the alerting module. Also, it will perform an expert analysis of the impact of PM air pollution on children health, supervised by medical experts (pediatrist physicians). The forecasting module was designed to perform a prediction of the $PM_{2.5}$ concentration level for a short term (one, two hours or 24-hours in advance). Some artificial intelligence based techniques are currently implemented to be used: an artificial neural network (ANN), an adaptive neuro-fuzzy system (ANFIS), and a knowledge based system (KBS).

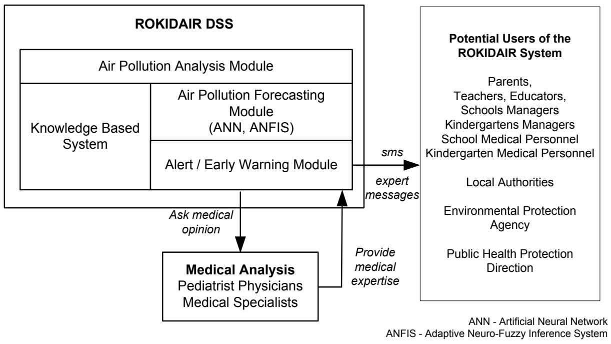

Fig. 4. The block diagram of the ROKIDAIR Decision Support System

The alert module will provide expert messages for sensitive children, with health problems (e.g. respiratory problems), from the areas with $PM_{2.5}$ air pollution episodes. The early warning messages are sent to parents, teachers, educators, school and kindergarten managers and medical personnel, only after a confirmation of air pollution analysis results and of the derived alert messages by medical specialists (i.e. pediatrist physicians).

III. CASE STUDIES OF PM AIR POLLUTION ANALYSIS AND FORECASTING

We have performed a statistical analysis of $PM_{2.5}$ air pollution (a spectrum analysis) and a PM_{10} short-term air pollution forecasting (a feed forward ANN - FFANN) and children health impact analysis (KBS). The two case studies used PM measurements datasets from the Ploiesti city.

A. *Airborne Fine Particulate Matter Spectrum Analysis*

The exploratory techniques of air quality data seek to analyze recorded time series by highlighting extremes, frequency distribution curves, and probability of peaks occurrence. The scalar accompanying the air pollution phenomenon is considered a probabilistic variable (with random evolution), which follows the law of large numbers and probability theory, which requires a large number of measured values in a time interval. Some of the techniques are factor analysis and regression to determine the influence of meteorological factors/other pollutants on observed daily $PM_{2.5}$ variability in various regions ([4], [5], [6]). Spectral analysis enables a robust estimate of the effect of climate change on $PM_{2.5}$ air quality ([7], [8]).

To establish potential $PM_{2.5}$ trends, it is important to study its multiannual time series. The current analysis considered daily time series recorded in Ploiesti city by a gravimetric instrument of the PH-2 automated urban station (N 44°56′21″, E 26°01′33″, altitude 150 m).

Table I presents the results of the statistical analysis using $PM_{2.5}$ time series recorded by PH-2 station in 2009-2012 period. $PM_{2.5}$ multiannual average at PH-2 station was 18.65 $\mu g/m^3$, and the annual average ranged between 16.9 and 20.7

μg/m^3. The resulted multiannual average was very close to the value determined for the *indicator of population exposure* at national level for the reference year 2010 i.e. 18.42 μg/m^3. This indicator was established by integrating the measurements of 15 stations, recorded in three consecutive years. The maximum amounts ranged between 62.1 and 88.9 μg/m^3, and the mode (most common value that occurs in the series) was between 8.9 and 16.4 μg/m^3. Time series of $PM_{2.5}$ daily averages showed coefficients of variation (CV%) between 57.8 and 66.7%. The distribution was of positive skewed type for all series. Kurtosis provided information regarding the flattening of distribution curves as follows: almost mesokurtic (2010), mesokurtic (2012) and leptokurtic (2009 and 2011).

In time series analysis, singular spectrum analysis (SSA) is a non-parametric spectral estimation method. It combines classic elements of time series analysis with multivariate statistical elements, dynamical systems, and signal processing.

TABLE I. STATISTICAL INDICATORS OF THE $PM_{2.5}$ TIME SERIES RECORDED IN PLOIESTI BETWEEN 2009 AND 2012 AT PH-2 AUTOMATED STATION

PH-2 urban station (years)	2009	2010	2011	2012
N (days)	*365*	*365*	*365*	*366*
Invalidated data (no. values)	154	10	5	2
Data capture (%)	57.8	97.3	98.6	99.5
Average (μg/m^3)	16.9	17.8	20.7	19.2
Median (μg/m^3)	14.5	15.1	17.1	16.4
Standard deviation	10.4	10.8	13.8	11.1
Coeff. of Var. (%)	61.7	60.8	66.7	57.8
Mode (μg/m^3)	8.9	14.0	11.8	16.4
Minimum (μg/m^3)	3.6	3.5	3.5	1.8
Maximum (μg/m^3)	72.7	62.1	88.9	77.0
Skewness	2.2	1.5	2.2	1.5
Kurtosis	6.5	2.6	6.2	3.0

SSA is used in the decomposition of a time series into a sum of components, each having a particular interpretation. Mathematically, SSA is a non-parametric spectral method that diagonalizes the delays' covariance matrix of variable $X(t)$ to obtain spectral information on the analyzed time series that are supposed to be stationary. Equations for obtaining the periodogram and spectral densities are given in [9].

The input in SSA was the daily multiannual time series obtained by computing pairwise the average of each day recorded in each of the 4 years (2009-2012). The number of valid observations was odd (365) and it should be even for Fourier analysis. The last valid observation was excluded from the analysis resulting 364 values after padding. The sine and cosine functions were orthogonal. The periodogram was obtained by summing the squared coefficients for each frequency (Fig. 5). The resulted values were interpreted in terms of variance of the data at the respective frequency or period (Figs. 6 and 7).

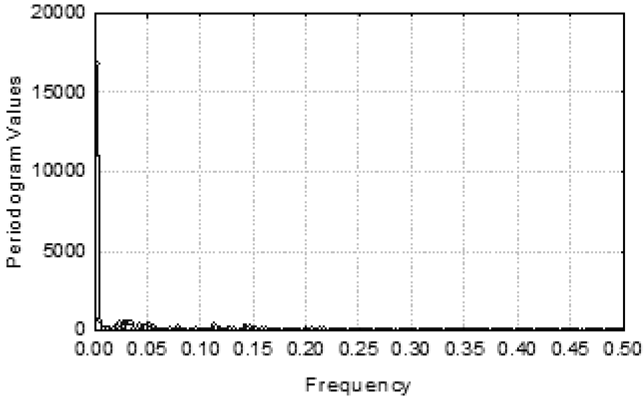

Fig. 5. Periodogram of $PM_{2.5}$ multiannual time series in Ploiesti city between 2009 and 2012

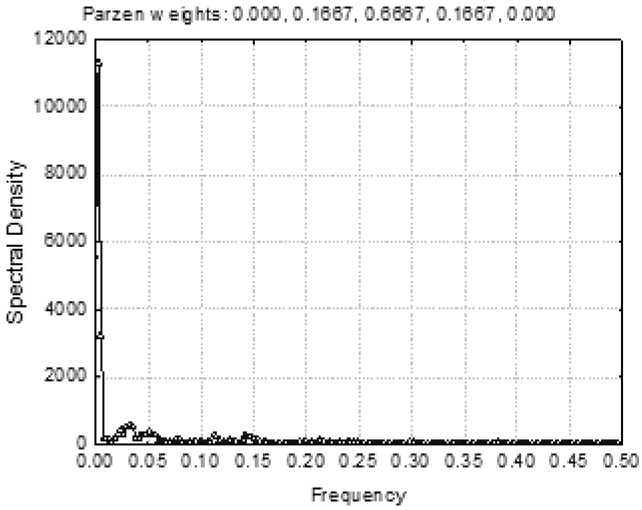

Fig. 6. Spectral density of $PM_{2.5}$ aggregated multiannual time series recorded in Ploiesti city at PH-2 urban station

Table II summarizes the results of SSA applied to the aggregated time series of $PM_{2.5}$ pointing out the potential periodicity of particulate matter data. Five largest periodogram peaks were extracted based on Value/Frequency ratio as follows: (1) 16,739.9 : 0.0027 at 364 period; (2) 631.4 : 0.0055 at 182; (3) 585.5 : 0.0357 at 28 (4) 575.7: 0.0302 at 33; and (5) 563.7 : 0.0330 at 30.

These results suggest some patterns of periodicities in $PM_{2.5}$ evolution at Ploiesti, respectively annual, half-year and monthly periodicities. Weekly periodicity (7) appeared in the tenth position. Fig. 8 shows an extract of the first 150 frequencies and the corresponding densities resulted from Fourier (Spectral) Analysis. Of particular interest was the median frequency (76.9×10^{-3}) that corresponded to a period of 13.

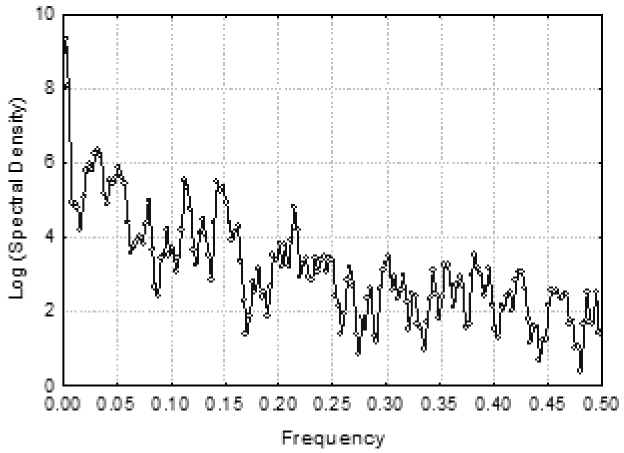

Fig. 7. Log of spectral densities at corresponding frequencies resulted from spectral analysis of $PM_{2.5}$ time series

This suggests that besides long-time intervals that influence the $PM_{2.5}$ periodicities, weekly and decadal periodicities might be present in particulate matter evolution at local scale in Ploiesti city.

TABLE II. SPECTRAL ANALYSIS RESULTS FOR THE MULTIANNUAL TIME SERIES OF $PM_{2.5}$ CONCENTRATIONS RECORDED AT PH-2 URBAN STATION IN PLOIESTI CITY (NO. OF CASES: 364)

Position	Frequency (cycles)*	Period (days)	Cosine	Sine	Periodogram	Density
1	0.00274	**364**	9.582	-0.382	**16739.99**	11265.73
2	0.00549	**182**	1.297	1.336	**631.42**	3212.08
13	0.03571	**28**	1.374	-1.152	**585.51**	499.22
11	0.03022	**33**	1.284	-1.230	**575.70**	514.57
12	0.03296	**30**	1.1641	-1.319	**563.68**	569.32
9	0.02472	40	0.898	-1.318	463.32	401.24
19	0.05219	19	-1.52	-0.173	425.97	364.46
8	0.02197	45.5	-0.371	-1.301	333.25	324.18
41	0.11263	8.8	1.314	0.0151	314.58	245.15
52	0.14285	7	0.34	-1.26	310.10	238.19

* *number of times a particular value for a data item has been observed to occur.*

B. *PM_{10} Short-Term Forecasting and Children Health Impact Analysis*

The PM_{10} short-term forecasting module uses hourly data recorded by the RNMCA network (http://www.calitateaer.ro) at the PH-2 station from Ploiesti, in the period January-December 2009. The forecasting model was developed by using a feed forward ANN that forecasts the next hour PM_{10} concentration based on the historical data from the past 8-hours (window time) of the following air pollutants concentrations: PM_{10}, SO_2, NO_x, and two meteorological parameters: relative humidity and temperature.

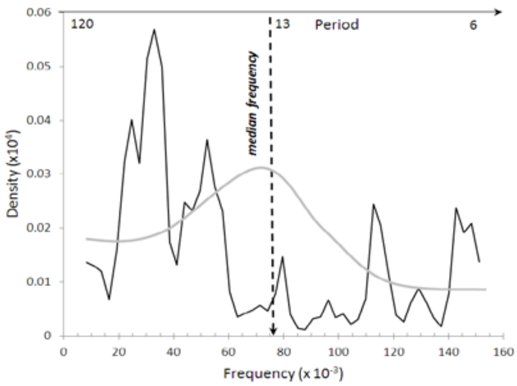

Fig. 8. Frequency spectrum of the $PM_{2.5}$ aggregated daily time series - extract of the first 150 frequencies and the corresponding densities resulted from Fourier (Spectral) Analysis; the grey line shows the smoothed spectrum from an autoregressive model applied to selected data

The inputs of the ANN are the most relevant parameters for the PM_{10} forecasting, selected with a principal components analysis (PCA), which derived eight principal components [10]. For the FFANN training step the back propagation algorithm with momentum was selected, being a versatile and robust technique able to solve prediction problems [11]. The data set was split into the training set (68%), the validation set (30%) and the testing set (2%). In order to avoid FFANN over fitting, experiments were done with a 10-fold cross-validation. Different configurations of the FFANN architecture were tested and the best one was chosen (8 nodes in the input layer, 4 nodes in the hidden layer and one node in the output layer). Fig. 9 shows the FFANN architecture (8×4×1) used in the Matlab implementation. Table III presents a selection of the experimental results obtained for some FFANN configurations (different no of hidden nodes, momentum and learning rate). The statistical parameters that were computed and analyzed are: the correlation coefficient (R^2), the epoch error (ERR), the mean absolute error (MAE) and the root mean squared error (RMSE). Their equations are given in [10]. The best values are bold highlighted.

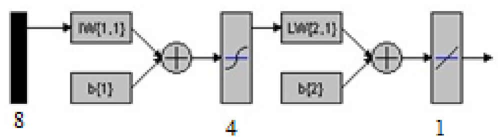

Fig. 9. The FFANN architecture (8×4×1) – Matlab implementation

An example of PM_{10} forecasting performed with FFANN:

Input PCA parameters:Time series (past 8 hours) with measurements for SO_2, NO_2, Temperature, Humidity, PM_{10};

ANN Input data: { PC1=227,79; PC2=-119,36; PC3=-2,96; PC4=-28,32; PC5=-2,24; PC6=-4,78; PC7=-1,51; PC8=-16,01} – the eight principal components provided by PCA

ANN Output (PM_{10} forecasted value) = 57.02 [$\mu g/m^3$];

PM_{10} measured value = 62.79 [$\mu g/m^3$].

A similar FFANN (named FFANN-24hF) was designed to forecasts the next day PM_{10} concentration level by using mean 24-hours values of the same parameters that were used by the previous forecasting model. Based on the output of this forecasting model, a PM_{10} AQI and children health impact analysis with corresponding alerting/early warning messages were performed by a prototyped knowledge based system, (KBS named PM10_AQI.kbs) that was implemented in VP-Expert, an expert system generator which provides easily and fast prototype versions of knowledge based systems.

TABLE III. EXPERIMENTAL RESULTS FOR DIFFERENT FFANN CONFIGURATIONS - NO OF EPOCHS = 5000

FFANN	Momentum	Learning rate	R^2	ERR	MAE	RMSE
8×2×1	0.2	0.3	0.7240	0.2093	5.9241	19.6859
8×3×1	0.1	0.1	0.7458	0.1499	10.9085	18.8446
8×4×1	0.1	0.1	0.7453	**0.1487**	**10.9033**	18.8892
8×5×1	0.3	0.2	0.6583	0.2005	6.0978	19.3452

The knowledge base contains rules for deriving the PM_{10} AQI status according to the ranges of PM_{10} forecasted concentration level, that are posted on the RNMCA monitoring network web site http://www.calitateaer.ro, and rules for possible children health effects for each PM_{10} AQI status. The following AQI status linguistic values were considered: excellent, very good, good, moderate, poor, very poor. Examples of rules used by the PM10_AQI KBS system:

```
RULE AQI_PM_4
IF F_PM > 29.9 AND F_PM <= 49.9 THEN   PM_AQI = 4;
RULE AQI_PM_S4
IF PM_AQI = 4 THEN   PM_AQI_S = moderate;
RULE HHE_PM_2
IF PM_AQI_S = very_good THEN   PM_HHE = none;
RULE HHE_PM_4
IF PM_AQI_S = moderate THEN
     PM_HHE = on_sensitive_population;
```

Several simulated PM_{10} air pollution episodes were run in order to test the KBS. Fig. 10 shows a screenshot of the KBS run for a simulated moderate PM_{10} air pollution episode in Ploiesti.

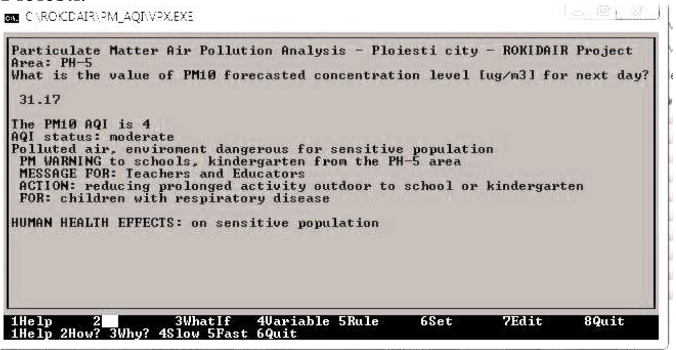

Fig. 10. PM10_AQI.kbs run screenshot with a simulated moderate PM_{10} air pollution episode in an area from the in-situ ROKIDAIR monitoring network of the Ploiesti city

IV. CONCLUSION AND FUTURE WORK

The paper presented the ROKIDAIR intelligent system for PM air pollution monitoring, analysis and forecasting, and two case studies of PM analysis: $PM_{2.5}$ air pollution spectrum analysis, and PM_{10} short-term forecasting with a children health impact analysis performed on Ploiesti datasets by the ROKIDAIR DSS. In situ and online PM monitoring networks are included in the system. Artificial neural networks and knowledge based systems are used for PM forecasting and health impact analysis. As a near future work we shall extend the knowledge base of the ROKIDAIR DSS with new heuristic rules derived by data mining techniques. Also, the online (continuous) $PM_{2.5}$ monitoring network will be developed.

ACKNOWLEDGMENT

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.

REFERENCES

[1] Y. Zhang, M. Bocquet, V. Maller, C. Seigneur, A. Baklanov, "Real-Time air quality forecasting, part I: History, techniques, and current status", Atmospheric Environment, vol. 60, pp. 632-655, 2012.

[2] St. Iordache, D. Dunea, E. Lungu, L. Predescu, D. Dumitru, C. Ianache, R. Ianache, "A Cyberinfrastructure for Air Quality Monitoring and Early Warnings to predict Children with Respiratory Disorders", Proceedings of the Int. Conf. CSCS20, May 2015, in press.

[3] D. Dunea, St. Iordache, M. Oprea, T. Savu, A. Pohoata, E. Lungu, "A relational database structure for linking air pollution levels with children's respiratory ilnesses", Bulletin USAMV series Agriculture, vol. 71, no.2, pp. 205-213, 2014.

[4] J. Hooyberghs, C. Mensink, G. Dumont, F. Fierens, and O. Brasseur, "A neural network forecast for daily average PM10 concentrations in Belgium", Atmos. Environ. 39: 3279–3289, 2005.

[5] St. Iordache and D. Dunea, "Exploring air quality data from an automated monitoring system in an urban-industrial area", Ann. Food Sci. Tech. 12 (1): 84-91, 2011.

[6] X.H. Liu, K.M. Wai, Y. Wang, J. Zhou, P.H. Li, J. Guo, P.J. Xu and W.X. Wang, "Evaluation of trace elements contamination in cloud/fog water at an elevated mountain site in Northern China", Chemosphere 88(5): 531-541, 2012.

[7] A.P.K. Tai, L.J. Mickley, D.J. Jacob, E.M. Leibensperger, L. Zhang, J.A. Fisher and H.O.T. Pye, "Meteorological modes of variability for fine particulate matter (PM2.5) air quality in the United States: implications for PM2.5 sensitivity to climate change", Atmos. Chem. Phys., 12, 3131–3145, 2012.

[8] St. Iordache and D. Dunea, "Cross-spectrum analysis applied to air pollution time series from several urban areas of Romania", Environ. Eng. Manage. J. 12: 677-684, 2013.

[9] D. Meko, "Spectral Analysis-Smoothed Periodogram Method", Notes_6, GEOS 585A, 2015. http://www.ltrr.arizona.edu/~dmeko/notes_6.pdf

[10] M. Oprea, E.G. Dragomir, S.F. Mihalache, M. Popescu, "Prediction methods and techniques for PM2.5 concentration in urban environment", In Methods for the assessment of air pollution with particulate matter to children's health, S. Iordache and D. Dunea (Editors), Matrix Rom, Bucharest, 2014 (in Romanian).

[11] H.J.S. Fernando, M.C. Mammarella, G. Grandoni, P. Fedele, R. Di Marco, R. Dimitrova, P. Hyde, "Forecasting P_{M10} in metropolitan areas: Efficacy of neural networks", Environmental Pollution, vol. 163, pp. 62-67, 2012.