

A Relational Database Structure for Linking Air Pollution Levels with Children's Respiratory Illnesses

Daniel DUNEA¹, Stefania IORDACHE^{1*}, Mihaela OPREA², Tom SAVU³, Alin POHOATA¹, Emil LUNGU¹

¹Valahia University of Targoviste, Romania.

²Petroleum-Gas University Ploiesti, Romania.

³University Politehnica of Bucharest, Romania.

*Corresponding author: stefania.iordache@yahoo.com

Bulletin USAMV series Agriculture 71(2)/2014

Print ISSN 1843-5246; Electronic ISSN 1843-5386

DOI 10.15835/buasvmcn-agr: 10689

Abstract

Pollutants' data, meteorological data and medical data feeds the relational database with information from vulnerable urban areas (i.e., Targoviste and Ploiesti), which will help the running of algorithms between air quality, meteorology and health effects and later use of forecasted outputs to forecast health effects. Several computer-based tools were developed to facilitate the population of the database with specific data from various sources. One of these tools allows the automatic capturing of pollutants' concentrations from web-based official sources. The relational database structure integrates the fields for the required variables in the attributed data tables (PM_{2.5} and carried compounds/metals sub-database, meteorological sub-database and medical sub-database). The main criteria in selecting the respiratory illnesses that are linked to atmospheric pollution for children are the wheezing. The medical database contains as main fields: the number of wheezing episodes, number of asthma attacks (with hospitalization), the response to inhalation medication, medication controller, eosinophil count, serum level of E immunoglobins (IgE), and residential address and school/kindergarten address of the children. The presented database structure and adjacent tools are expected to improve the current monitoring methodology of air pollutants, mainly respirable dusts, and their content in various compounds in correlation with children's health.

Keywords: *air pollution, computer-based tools, population exposure, relational database*

INTRODUCTION

Environmental monitoring systems and early warning systems have been improving the quality of life and health in many places throughout the world. The spatial variation in the pollution concentration and the differences between the areas covered by the monitoring and the areas where the population is located create problems in using the air quality data generated by routine monitoring networks (Liu *et al.*, 2013). In addition, the method used to report air quality monitoring data restricts the availability of the collected information for assessing exposure to air pollution and, subsequently, its health effects (Henschel *et al.*, 2012).

In Romania, National Law no. 104/2011 regulates ambient air quality assessment. The air quality monitoring is performed continuously

using the automated stations of the National Network for Air Quality Monitoring (RNMCA) located in various areas of the country. At country level, the PM_{2.5} levels were monitored with discontinuities at 15 urban stations in 2012. PM_{2.5} monitoring is required to comply with the requirements of Directive 2008/50/EC on air quality and cleaner air for Europe. Measurements of concentrations are used to establish the average exposure indicator of the population, determined at country level, which implies three years of continuous monitoring. The annual limit for PM_{2.5} is 25 µg/m³, which is the value that must be reached on January 1, 2015. The limit value plus the tolerance margin was 27 µg/m³ for the year 2012 (RNMCA, 2014). The actual infrastructure of the Romanian RNMCA does not cover the entire area of pollutant receptors in the monitored urban

areas and provides data only for a limited portion of the residential areas.

Furthermore, some of the stations do not provide data in the periods of maintenance or equipment failures, and consequently a back-up monitoring solution would diminish this hiatus. The hardware configuration of many stations does not include equipment that is able to monitor pollutants, which are dangerous for children's health (e.g., PM_{2.5}, VOCs, aromatic compounds, etc.). They were deployed only in high-risk areas. However, the air quality monitoring should also provide reliable and complete data for the specific characterization of the impact on less polluted areas focusing on vulnerable receptors (e.g., schools, kindergartens, maternities, hospitals, residential areas, etc.). Furthermore, it is recommended to relate outdoor pollution with indoor levels (Rivas *et al.*, 2014) at these receptors as well.

The ROKIDAIR project aims to improve urban air quality monitoring and forecasting activities focusing on the critical areas' spatial delimitation based on the receptors' vulnerability and their detailed characterization in terms of PM_{2.5} effects on children's health in two towns of Romania, namely Targoviste and Ploiesti. These urban agglomerations serve as pilot areas in order to develop and deploy a monitoring network system that will provide synthesized information concerning PM_{2.5} levels obtained from simplified, but reliable monitoring micro-stations and artificial intelligence (AI) forecasting algorithms that will be developed in the course of the project. The two towns were selected because their residential areas are under the emissions' impact of industrial areas: metallurgical industry (Targoviste), petrochemical industry (Ploiesti), and heavy traffic. From the medical point of view, the ROKIDAIR project will focus on respirable dusts and their chemical content, estimating their aggravating or causative effect on respiratory diseases in children up to 10 years.

In this paper, we have presented the structure of the relational database for linking air pollution levels with children's respiratory illnesses integrating the required fields for variables in the attributed data tables (pollutant subdatabase, PM_{2.5} and carried compounds/metals subdatabase, meteorological subdatabase, medical subdatabase and receptor subdatabase). The GIS developed during the ROKIDAIR project will manage

past measurements and records, past cases of respiratory issues, current PM_{2.5} measurements, modelling results and will generate thematic maps of attributes extracted from the initial relational database, which is described in the results section.

MATERIALS AND METHODS

The research has been carried out to develop the prototype of the relational database structure based on past pollutants' data, meteorological data and medical data of the last five years, which were collected from the corresponding authorities of the envisaged cities (i.e., Targoviste and Ploiesti).

Because in these towns PM_{2.5} levels are not currently monitored (DB-1 station is not functioning and does not have a specific analyzer; PH-2 does have a PM_{2.5} analyzer but does not provide data), we have developed a monitoring plan using a top-down approach. Monitoring campaigns were performed using a portable monitoring system to determine PM_{2.5} levels in representative points. The sampled discs were analysed using laboratory methods to establish the heavy metal content and other compounds in the samples. Measurement sheets and field book with coordinates of the identified sampling points were inputted in the corresponding subdatabase. In Ploiesti, the monitoring campaigns were performed in 12 sampling points (Fig. 1), respectively 10 points in Targoviste (Fig. 2).

The experimental scheme allowed the monitoring of the PM_{2.5} levels using a calibrated micro-dust aerosol monitoring system (Fig. 3) insuring the objectives and the required parameters achievements. The portable monitoring system uses an infrared beam projected in a measuring chamber. The particles are captured on a filter being insured a constant airflow by the connected air sampler.

Detailed planning of the sampling dates, sampling interval, the sequence of measurements correlated with the availability of the location of past cases of children respiratory issues were also established. Figure 4 shows the time series of PM_{2.5} measurements recorded with 4 seconds sampling interval in one sampling point.

Based on the main database (i.e., pollutant emissions, number of children with respiratory illnesses, and scenarios modelling), critical areas will be identified. The prioritization of critical areas will be performed in order to facilitate



Fig. 1. PM 2.5 monitoring points in Ploiesti urban area (12 points)

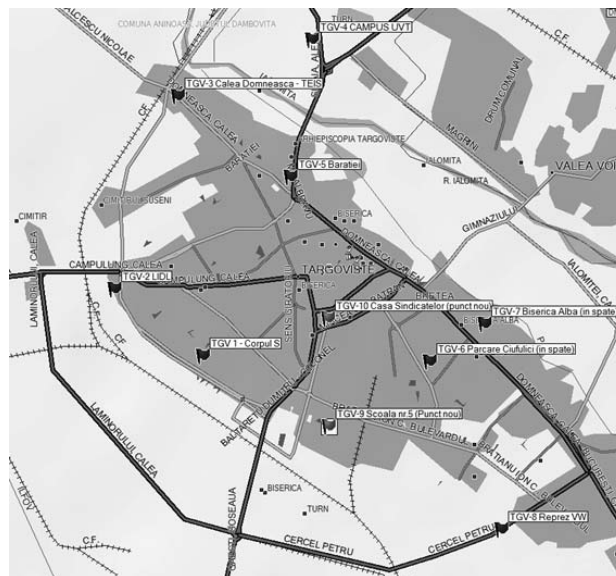


Fig. 2. PM 2.5 monitoring points in Targoviste urban area (10 points)



Fig. 3. Monitoring equipment of PM_{2.5} fraction deployed in Targoviste
 1 - data logger; 2 - aerosol measuring probe; 3 - adapter with filtering cassette of 37 mm; 4 - light shutter; 5 - filters for PM₁₀, Respirable (4µm) and PM_{2.5} fractions; 6 - filtering box of 25 mm; 7 - gravimetric adapter; 8 - APEX sampling pump; a - display of real-time particle concentrations on the data logger screen; b - side view of the probe with the filter box installed for continuous static monitoring of respirable dusts (Casella Microdust Pro).

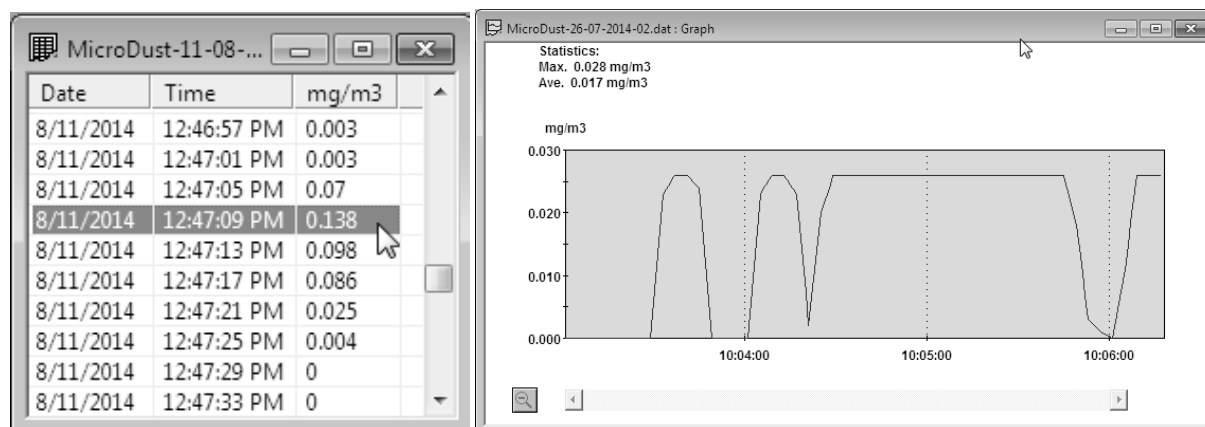


Fig. 4. Time series of PM 2.5 measurements recorded with 4 seconds sampling interval

the determination of the most relevant sites for automated PM_{2.5} microstations deployment. The microstations will be developed within the ROKIDAIR project.

RESULTS AND DISCUSSION

The relational database contains six subdatabases as follows:

1. Airbase EIONET pollutant database recorded by 8 stations located in the envisaged areas,
2. Current levels of air pollution database,
3. PM_{2.5} monitoring and PM_{2.5} content database,
4. Meteorological parameters database recorded at Targoviste (153750) and Ploiesti (153770) WMO stations,
5. Medical database, and
6. Receptor database (kindergartens, schools, and pediatric hospitals).

Figure 5 presents the structure of the relational database, which will form the core of the ROKIDAIR system. In the initial step, the first three subdatabases provide synthetic data filtered using various methods, i.e. statistical (Dunea et al., 2008), wavelet (Dunea et al., 2014), and/or artificial intelligence algorithms (Dunea et al., 2011, Oprea, 2008). Next step will integrate the meteorological influence based on correlation (Dunea, 2014) or other convenient methods (Iordache and Dunea, 2013).

All data are georeferenced in Romanian Stereographic 1970 and WGS 84 coordinate systems, which allows the spatio-temporal characterization of vulnerable areas in each town through geospatial analysis. Medical database and receptor database will supplement with specific infor-

mation the identification of the critical areas under PM_{2.5} impact. The subsequent ranking of the critical areas will allow the deployment of the microstations, whose measurements will be stored in a specific database. Furthermore, the deployment of microstations will be also considered for indoor measurements and indoor-outdoor comparisons.

In the following section, we have briefly described each of the component subdatabase.

The historical subdatabase is build based on data from Airbase EIONET (Airbase, 2014). These are official validated data reported by the national authorities. It is a MySQL database and the main tables are STATII (automated monitoring stations), CONFIG, VALORIORARE (hourly values), VALORIZILNICE (daily values). Eight stations were selected being located in the envisaged towns or at various distances from them (Fig. 6).

CONFIG data table contains descriptive information about the pollutants measured by the stations. The "station_european_code" field links this table with STATII data table which gives information about the location and type of the automated monitoring station.

Table VALORIORARE is created with the following command:

```
CREATE TABLE `valoriorare` (
  `ID` BIGINT NOT NULL AUTO_INCREMENT,
  `ID_Config` BIGINT NOT NULL,
  `DataOra` DATETIME NOT NULL,
  `Valoare` FLOAT NOT NULL,
  `Valid` TINYINT NOT NULL,
  PRIMARY KEY (`ID`)
)
```

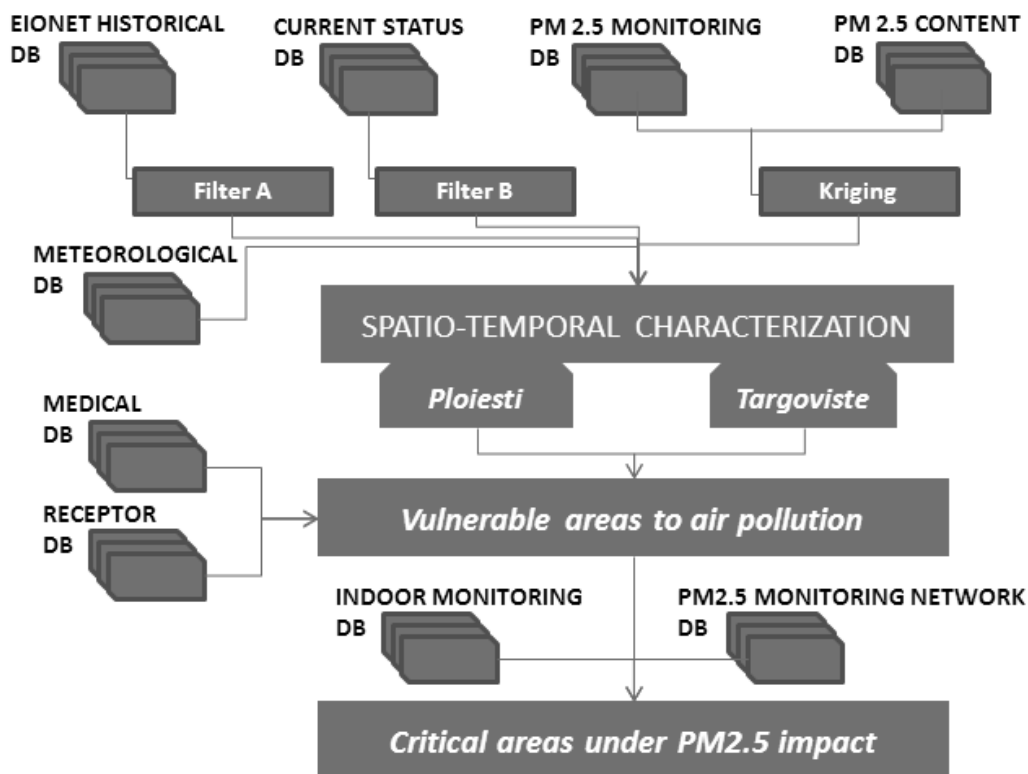


Fig. 5. The information flux within the ROKIDAIR relational database

Station_european_code	Station_local...	Cou...	Country_n...	Station...	Station_start_date	Type_of_station
RO0139A	RO-DB-1	RO	Romania	DB-1	2008-01-01	Industrial
RO0140A	RO-DB-2	RO	Romania	DB-2	2008-01-01	Industrial
RO0175A	RO-PH-1	RO	Romania	PH-1	2008-01-01	Traffic
RO0176A	RO-PH-2	RO	Romania	PH-2	2008-01-01	Background
RO0177A	RO-PH-3	RO	Romania	PH-3	2008-01-01	Background
RO0178A	RO-PH-4	RO	Romania	PH-4	2008-01-01	Industrial
RO0179A	RO-PH-5	RO	Romania	PH-5	2008-01-01	Traffic
RO0180A	RO-PH-6	RO	Romania	PH-6	2008-01-01	Industrial

Fig. 6. EIONET subdatabase with RNMCA stations table

ID	ID_Config	DataOra	Valoare	Valid
296	1	2009-01-13 07:00:00	13.23	1
297	1	2009-01-13 08:00:00	17.21	1
298	1	2009-01-13 09:00:00	17.14	1
299	1	2009-01-13 10:00:00	23.91	1
300	1	2009-01-13 11:00:00	51.82	1
301	1	2009-01-13 12:00:00	41.95	1
302	1	2009-01-13 13:00:00	30.49	1
303	1	2009-01-13 14:00:00	24.12	1
304	1	2009-01-13 15:00:00	25.31	1

Fig. 7. EIONET subdatabase with hourly values table recorded by RNMCA stations

The second field (ID_Config) links this table with CONFIG data table to identify the pollutant that the field Valoare refers to.

Similarly, VALORIZILNICE data table is created with:

```
CREATE TABLE `valorizilnice` (
  `ID` BIGINT NOT NULL AUTO_INCREMENT,
  `ID_Config` BIGINT NOT NULL,
  `DataCalendar` DATE NOT NULL,
  `Valoare` FLOAT NOT NULL,
  `Valid` TINYINT NOT NULL,
  PRIMARY KEY (`ID`)
)
```

and contains the field ID_Config that makes the connection between CONFIG and VALORIZILNICE data tables.

All data were imported using scripts in VBS to extract data from the text files provided in the archive downloaded from EIONET (Fig. 7). Stored data were recorded between 2008 and 2013.

The "Current levels of air pollution" subdatabase (BazaCalitateAer.s3db) is associated to the software application that downloads raw values of measured concentrations of pollutants and meteorological data from the dedicated website of RNMCA (www.calitateaer.ro). This Sqlite database allows an easy installation on any PC running Windows operating system without special requirements (dedicated database servers and corresponding connecting drivers). There are two tables defined in the database. One table (ParametruStatie) is used to store the information about the types of pollutants and meteorological data that are provided by each monitoring station that presents interest.

The table was created using the following SQL command:

```
CREATE TABLE ParametruStatie (
  Selectat NUMERIC,
  PS NUMERIC,
  Statie_ID TEXT,
  Statie_Nume TEXT,
  Parametru_ID TEXT,
  Parametru_Nume TEXT,
  Parametru_Formula TEXT,
  Parametru_UM TEXT
)
```

Although we can further normalize the table to separate information about stations and

pollutants we preferred to keep this form because the amount of data is relatively small and the table is populated with data before starting the service that will automatically download data from internet.

The installation of our application creates a service that needs to be started after this table is populated with data. The service queries this table and selects the lines for which the value of the first field (Selectat) is not equal to 0. For each value of the second field, the service will call the .exe file (GetTable.exe) that will collect the hourly data of the corresponding parameter and store them into the table Valori (values).

The executable file *GetTable.exe* works in two modes. If it starts with zero parameters then a window is showed that allows the user to add data in table ParametruStatie (parameter and station). If the .exe file is running with a numeric parameter, no window will be showed. The application will extract the values of Statie_ID and Parametru_ID cooresponding to the parameter and based on these values will build the necessary url to download data.

The second table of the current air pollution database is created with:

```
CREATE TABLE Valori (Valoare TEXT, DataOra TEXT, PS TEXT)
```

This table will store the hourly data for pollutants identified by the third field.

The data can be extracted from database for processing purposes using for example *SELECT * FROM Valori WHERE PS IN (*
SELECT PS FROM ParametruStatie WHERE
Statie_Nume='PH2' AND
Parametru_Formula='O3'

Figure 8 shows the window of the application for current measurements extraction from the RNMCA website. We have defined 55 combinations of Statie_ID and Parametru_ID in order to characterize the status of air pollution and some potential correlations of recorded pollutants concentrations with meteorological variables. Unfortunately, data were not available for several combinations due to analyzers malfunctions.

Figure 9 presents the air pollution status subdatabase, which was populated starting with July, 2014.

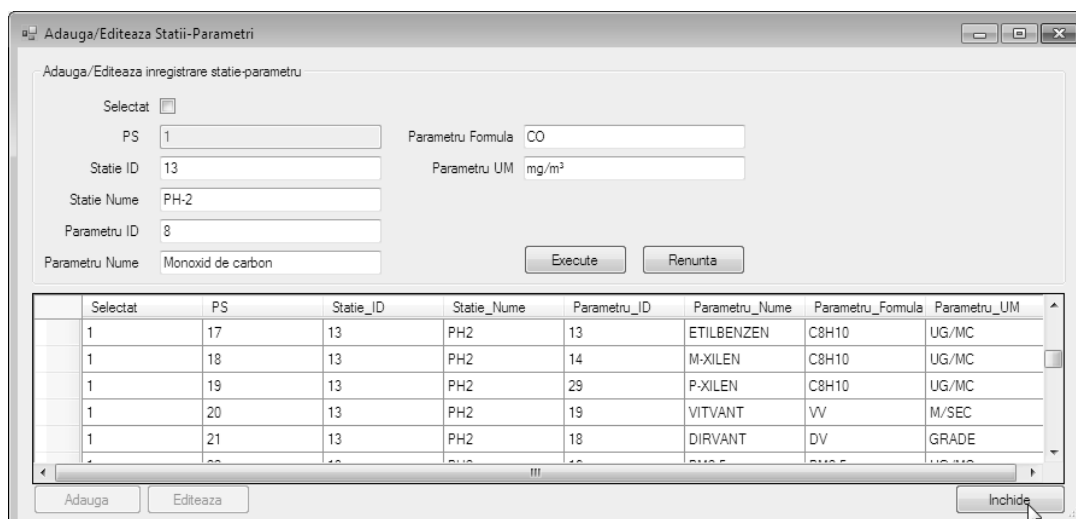


Fig. 8. Application for current measurements extraction from the RNMCA website

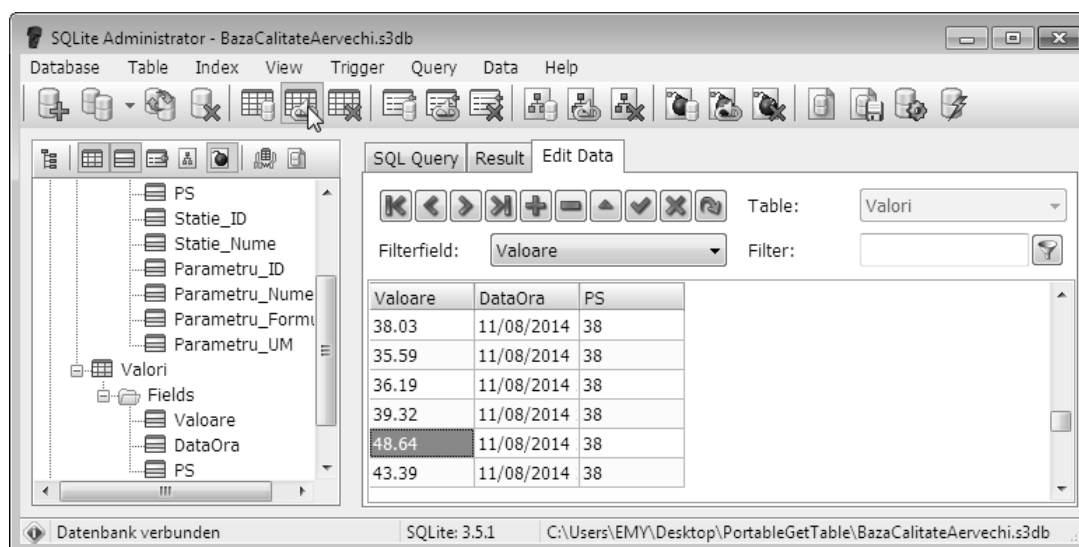


Fig. 9. Air pollution status subdatabase of ROKIDAIR DBMS

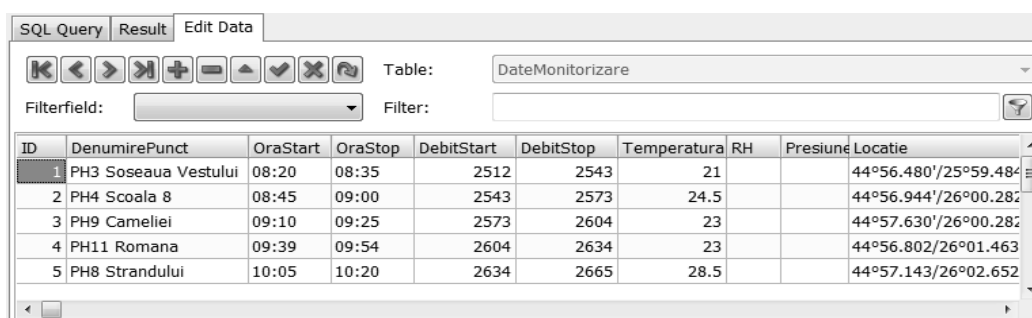


Fig. 10. PM_{2.5} monitoring subdatabase of ROKIDAIR DBMS

PM_{2.5} monitoring and PM_{2.5} content database (Fig. 10 and Fig. 11), contains information acquired from discontinuous field measurements and laboratory determinations providing PM_{2.5} concentrations, local meteorological conditions during measurement, sampling pump flows, spatial coordinates and particles content in heavy metals and other compounds. The population of the database improve the knowledge of heavy metals and VOCs carried by the fine particulate matters in both urban areas, which will bring new findings related to the health risk evaluation.

Meteorological parameters database contains data recorded at Targoviste (153750) and Ploiesti (153770) WMO stations (e.g. air temperature, wind speed and direction, relative humidity and precipitations – daily average values).

Receptor database contains information regarding the address of kindergartens, schools, and pediatric hospitals, number of children, age groups, number of sensitive children, and indoor PM_{2.5} levels.

The medical database contains as main fields the number of wheezing episodes, number of asthma attacks (with hospitalization), the response

to inhalation medication, medication controller (yes/no), eosinophil count (normal value = 1-3%), serum level of E immunoglobins (IgE) (normal value <60 units/ml), and residential address and school/kindergarten address of the children (Fig. 12). The number of episodes is counted starting from the child's first consultation. The main criteria in selecting the respiratory illnesses that are linked to atmospheric pollution for children are the wheezing. Children's respiratory illnesses will cover bronchiolitis, interstitial acute pneumonia with wheezing, recurrent wheezing, and asthma depending on the child age. Based upon cluster analysis, for each pollutant, in a grid-based GIS map, the areas of Targoviste and Ploiesti will be classified into different categories of pollution level (e.g., 1- *least pollution*, 4 - *highest pollution*), and coupled with a children's respiratory illness register of the participating hospitals.

In Austria, repeated examinations of children aimed to supplement the time series approach for detection of temporal relationships between changes in air pollution and in health status of risk groups, including signs and symptoms not leading to hospitalization and subclinical functional chan-

ID	PunctMas	Zona	Pb	Cd	Cr	Ni	Cu	Mn	Al	Zn
1	P1	Zona Lidl PM2.5	2.109	0.252	0.0601	0.872	0.871±0.1	0.792±0.1	3.134±1.0	0.932±0.1
2	P1	Zona Lidl TSP	2.672	0.281	0.0632	0.914	0.902±0.2	0.811±0.2	3.193±0.7	0.989±0.3
3	P2	Halta Teis PM2.5	2.269	0.244	0.0436	0.82	0.915±0.5	0.834±0.1	2.982±0.5	1.120±0.4
4	P2	Halta Teis TSP	2.311	0.264	0.0482	0.917	1.109±0.3	0.862±0.2	3.003±0.9	1.145±0.5
5	P3	Campus UVT PM2.5	2.572	0.242	0.0662	1.162	1.127±0.4	0.915±0.4	3.010±0.7	1.189±0.6

Fig. 11. PM_{2.5} content subdatabase of ROKIDAIR DBMS

ID	InitialeSiAnulN	Adresa	NrEpisoade	NrInternari	ValLgePerMI	ValEozinofile	MedicatieRe	MedicatieCon
1	A.D.2007	Str. Smaranda Ghegiu Nr. 2	20	4	465	2.3 +	+	
2	C.B.2007	Str. Tudor Vladimirescu 3A	12	3	241	1.4 +	+	
3	D.M.2007	Str. Ing. Gib Constantin	18	3	373	2.8 +	+	
4	F.A.2007	Str. Unirii	40	4	237	9 +	+	
5	L.M.2009	Str. Radu Popescu	25	3	251	1.8 +	+	

Fig. 12. Medical subdatabase of ROKIDAIR DBMS

ges. The main outcome variables for the analyses were as follows: asthma score, asthma attacks, and the overall assessment of respiratory health of the child during the last four weeks on a five-point scale by the parent (Neuberger *et al.*, 2004).

CONCLUSION

The developed relational database facilitates the running of statistics/AI models between air quality, meteorology and health effects, and later use of forecasted outputs to forecast health effects.

Acknowledgements. The research leading to these results has received funding from EEA Financial Mechanism 2009 - 2014 under the project "*Towards a better protection of children against air pollution threats in the urban areas of Romania*" contract no. 20SEE/30.06.2014.

REFERENCES

1. Dunea D (2014). An exploratory analysis of PM₁₀ particulate matter relationships with weather data and spatial variation, 14th GeoConference on Energy and Clean Technologies, Conference Proceedings, Nuclear technologies, Recycling, Air pollution and climate change, Vol. II: 273-280.
2. Dunea D, Iordache S, Pohoata AA (2014). Multiple characterizations of urban air pollution time series using a wavelet feedforward neural network integrative approach, Proceedings ITISE 2014, International work-conference on Time Series, Granada, June, 25-27 2014, Vol.2:804-815.
3. Dunea D, Pohoata AA, Lungu E (2011). Fuzzy inference systems for estimation of air quality index, ROMAI Journal, Romanian Society of Applied and Industrial Mathematics 7(2):63-70.
4. Dunea D, Oprea M, Lungu E (2008). Comparing statistical and neural network approaches for urban air pollution time series analysis, Proceedings of the 27th IASTED International conference on "Modelling, Identification and Control" (ed. L. Bruzzone), Acta Press, 93-98.
5. Henschel S, Atkinson R, Zeka A, Le Tertre A, Analitis A, Katsouyanni K, Chanel O, Pascal M, Forsberg B, Medina S., Goodman PG (2012). Air pollution interventions and their impact on public health, International Journal of Public Health, 57(5):757-768.
6. Iordache S, Dunea D (2013). Cross-spectrum analysis applied to air pollution time series from several urban areas of Romania, Environmental Engineering and Management Journal, April 2013, 12(4):677-684.
7. Liu H-Y, Skjetne E, Kobernus M (2013). Mobile phone tracking: in support of modelling traffic-related air pollution contribution to individual exposure and its implications for public health impact assessment, Environmental Health 2013, 12:93.
8. Neuberger M, Schimek MG, Horak Jr. F, Moshhammer H, Kundi M, Frischer T, Gomiscek B, Puxbaum H, Hauck H, AUPHEP-Team (2004). Acute effects of particulate matter on respiratory diseases, symptoms and functions: epidemiological results of the Austrian Project on Health Effects of Particulate Matter (AUPHEP), Atmospheric Environment 38:3971-3981.
9. Oprea M (2012). INTELEnvQ-Air: An Intelligent System for Air Quality Analysis in Urban Regions, International Journal of Artificial Intelligence, 9(A12):106-122.
10. Rivas I, Viana M, Moreno T, Pandolfi M, Amato F, Reche C, Bouso L, Alvarez-Pedrerol M, Alastuey A, Sunyer J, Querol X (2014). Child exposure to indoor and outdoor air pollutants in schools in Barcelona, Spain, Environment International, 69:200-212.
11. Romanian National Network for Air Quality Monitoring (RNMCA) – <http://calitateaer.ro/>
12. AirBase: public air quality database [online database]. Copenhagen, European Environment Agency, <http://www.eea.europa.eu/themes/air/airbase>