INFORMATION
SYSTEMS IN
MANAGEMENT

Information Systems in Management (2015) Vol. 4 (4) 251–263

THE USE OF SPATIAL DATA PROCESSING TOOLS FOR AIR QUALITY ASSESSMENTS - PRACTICAL EXAMPLES

DOMINIK KOBUS, JAKUB KOSTRZEWA

Air Quality Monitoring Department, Institute of Environmental Protection - National Research Institute (*IOS-PIB*)

In this article the potential applications of GIS systems in the management of air quality are discussed. In particular two specific issues are described: determining the spatial representativeness of air quality monitoring stations and areas of standards exceedances. The methods using spatial and emission data, as well as, the results of measurements, i.e. Land Use Regression method and so-called Beta parameter method are presented. The first one is based on the regression models in which mentioned information may be treated as independent data and, as a result, we obtain information on the levels of pollutant concentrations in point and space. The second method involves the parameterization of the spatial characteristics selected on the basis of the monitoring and emission data. These methods can find practical application in the field of air quality monitoring, assessment and the preparation of a public web presentation.

Keywords: air quality, environmental monitoring, geographic information systems, spatial data processing, decision support systems

1. Introduction

Air pollution is one of the main environmental problems occurring in Poland and in other countries. In Poland, this problem affects mainly large urban areas, but also concentrations of polluting substances are in excess of the established standards can affect smaller cities and rural areas. Industrial installations, road transport and municipal-household sector (small heating plants and individual heating systems) are the sources of air pollution. Particulate matter PM10 or PM2.5 and benzo(a)pyrene and, in smaller range, ozone or nitrogen dioxide, can be treated as problematic pollutants whose concentrations exceed the applicable standards [1]. An increased concentrations of these substances can cause adverse health effects, both short-term as a result of exposure during episodic occurrence of smog situation, and as a result of long-term exposure.

In 2013 exceedances of the limit value for PM10 (based on daily averages) occurred in 36 out of the monitored 46 zones in which an air quality assessment is performed [1]. It is associated mainly with the emission from the sources of communal-residential sector (household heating systems using solid fuels, often old and not efficient, especially active in winter period). An allowed number of exceedance of the level 50 $\mu g/m^3$ is 35, according to the Polish and European legislation. This standard is not achieved at many stations throughout the country, but in the south (region of Silesia and Lesser Poland) situation is the worst. Poland was reprimanded by the European Commission for its non-compliance and an infringement procedure was started at the European Court of Justice.

Various informatics systems are used for the purposes of air quality management at different levels European, national, regional and local. Assessment of air quality is an element of environmental management processes. Information systems are used, inter alia, for the collection, processing, visualization, transmission and reporting of various types of data and information, e.g. the results of measurements of pollutants concentrations or mathematical modeling. To the group of those systems we can include different types of systems and tools from the GIS family, from mobile applications or simple desktop applications to complex, multi-module systems for analysis and presentation, which use spatial databases, analyze information in real time and allow to work through an internet network (i.e. web-GIS).

2. Usage of spatial information processing in air quality management

2.1. Model of air quality management

The air quality assessment is one of the elements of environmental management. For the purposes of the analysis, including analysis the interplay between the environment and socio-economic activities, Driving forces - Pressures - State - Impact - Responses (DPSIR) framework is used. This approach can encourage and support decision-making processes, by pointing to clear steps in the causal chain of management. It has been adopted, e.g. by the European Environment Agency (EEA) and US Environmental Protection Agency (US EPA). It is extension of the Pressure - State - Response model developed by OECD and it has been applied to the organization of systems of indicators and statistics in relation to policy aims.

This model describes a dynamic situation, with attention for the various feedbacks in the system. By their nature, indicators take a snapshot picture of a constantly changing system, while the assessments that accompany the indicators can highlight the dynamic relations [2]. The existence of dynamic interrelations within a DPSIR framework makes it often a very complex web of many interacting factors. In many cases the change in the state of the environment or impacts has several causes, some of which may be immediate and of local origin, others may be exerting their influence on a continental or even global scale. Reductions in pressures often result from a mixture of policy responses and changes in various driving forces [3]. The framework is seen as giving a structure within which to present the indicators needed to enable feedback to policy makers on environmental quality and the resulting impact of the political choices made, or to be made in the future.

Information and communication technologies (ICT) may currently be used at each stage of the analysis and management of air quality. It includes, among others, the processing and analysis of the spatial information using GIS tools and systems. Fig. 1 shows the general scheme of the DPSIR model, while Table 1 gives brief definitions of individual elements of the model, along with examples of activities carried out with the use of GIS.

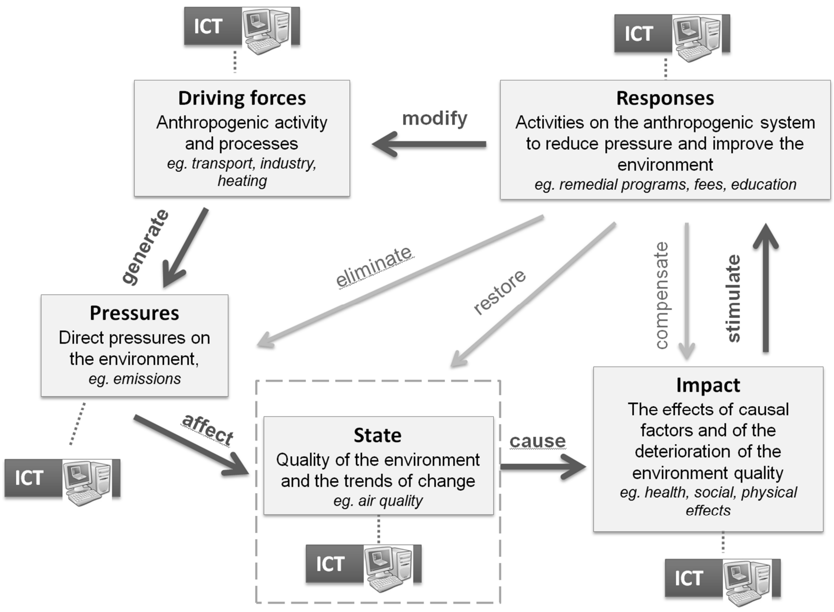

Figure 1. General schema of the DPSIR Framework
Source: own preparation on the basis of [2]

Table 1. An explanation of the levels of the DPSIR Framework with examples of application of GIS technologies

Level	Description	Examples
Driving forces	Human influences and activities which underpin environmental change (positively or negatively). The driving forces behind air quality change include increased burning of fossil fuels for transport, and industrial or domestic consumption of energy.	• Analysis of the spatial location of industrial plants • Analysis of the transport system and traffic distribution • Analysis of the structure and condition of the buildings (e.g. heat and energy demand)
Pressures	Direct or indirect pressures on the functionality and quality of the environmental system or resource, resulting from the driving forces, e.g. emissions of SO_2, NO_X, PM10 etc. constitute pressures on the natural atmospheric system.	• Analysis of the location of emission sources and spatial distributions of emissions • Emission inventory and modeling
State	Current status of the system or resources, in terms of quality of the environment and quantity or quality of resources, e.g. gaseous and particulate concentrations measured at particular stations - the state of air quality (national, regional, local, urban, etc.).	• Monitoring network management • Analysis of the results of measurements and mathematical modeling, including the use of geostatistical methods • Decision support in the air quality assessment
Impact	Environmental effects/responses to pressures on the current state, e.g. human health impacts (welfare of human beings , increased incidence of respiratory disease) and higher incidence of corrosion of infrastructure.	• Modelling and assessment of health risks
Responses	Responses to the pressures on the states and resultant impacts. Possible actions: • to mitigate, adapt to, or protect human induced negative impacts on the environment, • to halt or reverse environmental damage already inflicted, • to preserve and conserve natural resources. e.g. implementation of air quality standards, monitoring of air quality, installation of clean-air technologies, changing of heating systems, limitation of traffic, a policy to change mode of transportation, e.g. from private cars to public etc.	• Forecasting and visualization of the effects of corrective measures • Public information and education - geoportals

Element "State" in Fig. 1 has been marked because the process of air quality assessment concerns mainly the diagnosis and description of its state. In addition, of course, it includes indicate the reasons of the described state (Pressures), and also shows the possible consequences - "Impact".

Under the current rules, relevant services and institutions use various types of data and information from the following sources in the frame of air quality assessment: measurements of air pollution concentration (automatic or laboratory), mathematical modeling of the pollutant distribution and objective estimation methods.

2.2. The concept of GIS-based Decision Support System for air quality assessment

The GIS systems are used for the measurement networks management, for example to analyze and evaluate the location and characteristics of the stations. Their spatial representativeness is an important feature, which allows proper interpretation of measurement results. This is an attribute that is also subject to information reporting to the European Commission within the description of the measurement system functioning in a Member State. Methods for determining boundaries of representativeness area are the subject of various studies [4, 5], including ones presented in this article. An analysis of regional and local dispersion conditions of gas and dust substances is required to assess the representativeness and evaluation of potential public exposure to measured pollutant concentrations. Processing and visualization of spatial information can much help in such kind of tasks.

The processing and analysis of the measurements, modeling results, identification and evaluation of the situation of standards exceedances, including the identification of exceedances areas, constitutes another group of tasks for air quality assessment. The GIS tools are becoming more widely used also in this field. Measurement and modeling can be supported with objective estimation methods. So far, they have been based often on expert assessments e.g. with the supposed analogy between different areas or time periods. Increasingly, these methods use a computer support, primarily based on the GIS, e.g. the analysis of land use or correlation between spatial distribution and activity of emission sources and levels of measured concentrations. Various types of models, e.g. stochastic regression model or multi-source Gaussian dispersion model, as well as, the geostatistical interpolation methods are used to evaluate the exposure to pollution. An important element of air quality management is analysis of the trends of the past and forecasted changes. This is often done in conjunction with the assessment of changes in terms of spatial use and projection of possible implementation of corrective actions. This can be a support for planning, designing and evaluation of strategies and actions to control emissions and air quality management (investments or organizational activities - for example: reduction of car traffic in a given area, changes in traffic infrastructure or domestic heating systems in the analyzed area). Monitoring of the implementation and effectiveness of the applied measures is a very important issue in the frame of this management process [6, 7].

Implementation of the listed tasks can be performed using standard GIS applications and also through the use of own, dedicated tools and solutions. An example of the latter approach is a Decision Support System for air quality assessments, tentatively called AirQualGIS, designed and built in the Air Quality Monitoring Department in the Institute of Environmental Protection - National Research Institute. Its general scheme is presented in Fig. 2. It consists of a group of internal data processing modules, combined with the internal database of descriptive and spatial data, as well as, a dedicated module used for information presentation and user interaction, based on access via a web browser. The data obtained from external systems are processed by the internal interface, which implements also functions of export of data obtained in the system. Currently the system is used, inter alia, for the development and testing of methods, which are described later in this article.

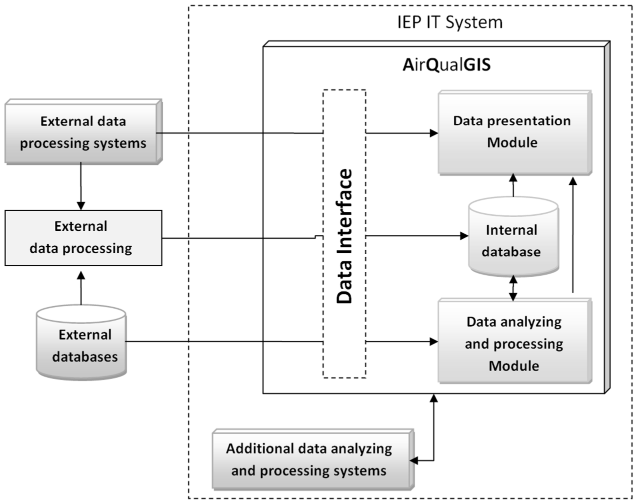

Figure 2. General schema of developed GIS decision support system for air quality assessment

The list of technologies and tools used for the construction of a prototype decision support system, along with the interrelationships diagram, is illustrated in Fig. 3.

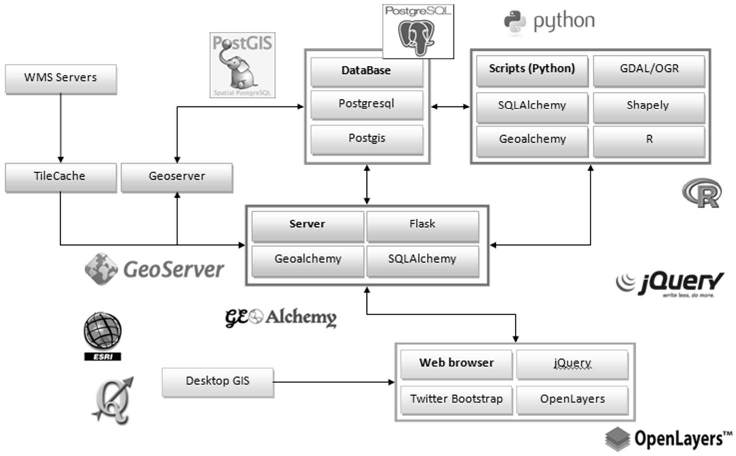

Figure 3. Technologies and tools used in the decision support system

3. Practical examples of the use of spatial information processing tools

Two analytical methods which employ the use of spatial data processing and which can be practically applied in the processes of air quality assessment are described below: Land Use Regression (LUR) method and the so-called Beta parameter method. The authors investigated the possibility of their adaptations and applications in Polish conditions, e.g. for determining the stations' spatial representativeness or ranges of exceedances of the limit levels. These methods were used within the development of previously described decision support system. As mentioned, representativeness of monitoring station is an important element of proper interpretation and analysis of results. The premise, which was adopted in the analysis and modification of the foregoing methods, it was easy availability of the required input data, so that it would be possible to apply them for different regions and periods. Another condition is the speed of data processing and average requirements of hardware resources for example for real-time analysis and visualization. Both these features distinguish these methods from the modeling of chemical pollutants transformation and transport, that may indeed produce more reliable results, but they are in great demand in relation to data and hardware performance, and require highly qualified personnel.

3.1. Land Use Regression method

The Land Use Regression method (LUR) has been more and more widely used in recent years and its popularization is results from the development of the GIS software, often equipped with tools for application of geographically weighted regression methods. This kind of methods can be used in e.g. real estate market analysis, the studies of the health risks and exposure and environmental analysis related to the soil, water or air pollution [8]. The term "regression" is most commonly used in relation to the method for the prediction of unknown values of one of the variables on the basis of the relationship between knowledge of other variables and their values [9]. A variable whose value is sought is called dependent variable. The relationship between the explanatory variable (or variables) and dependent variable is used to forecast the value of the latter. In case of more complex dependencies, the regression equation containing only one explanatory variable proves not sufficient. In that case, the multiple regression, allowing for the effect of a larger number of variables can be used. Multiple regression equation can be written as (1).

$$y_i = a_0 + a_1 x_{i1} + ... + a_k x_{ik} + \varepsilon_i \tag{1}$$

The parameters a_0, a_1,..., a_k are unknown and must be estimated on the basis of a random sample, using e.g. the least squares method. Parameter ε_i represents random components that play the role of random error. The use of multiple linear regression for analysis of air quality is justified by a significant number of explanatory variables that can be used in the description of the spatial variability of concentrations. These include spatial data describing the analyzed area. Potential variables are presented in Table 2.

Table 2. The potential explanatory variables, which can be used in LUR modeling

Source of the variables	Description	Spatial range
Land use	Sum of the areas of particular classes of land use in the vicinity of the analyzed points.	Circle with a given radius of the analyzed points.
Road network	The sum of the lengths of road types (divided by the volume of traffic or road class) in the neighborhood of the point, the distance from the road.	Circle with the radius of the analyzed points, the distance from the object, etc.
Traffic information	The number of vehicles per day on roads in the vicinity of the analyzed points.	Circle with a given radius around the analyzed points.
Population density	The average population density in the area surrounding the analyzed points.	Circle with a given radius around the analyzed points.
Number of house-holds	The density of households in the vicinity of the analyzed point.	Circle with a given radius around the analyzed points.
Emission Cadastre	Point, surface and linear emission. The sum of emissions in a given area or emission weighted with a distance of source from the point.	The sum of the emissions in a circle with a given radius or account of distance from the point.
Location	Geographical coordinates and altitude	-

The dependent variables in the construction of the model can be measured values on the basis of which a model is constructed. These data can be derived from fixed pollution concentrations measurement stations or be collected on passive measurement stations located at selected points within the measurement campaign aimed at developing LUR model. Depending on the scale of the analyzed area such data can be derived from the city, region, country etc. At a later stage of the analysis the dependent variables store the values calculated using the constructed model.

The Corine Land Cover data, road network with traffic information and concentrations of air pollutants included in the AirBase (air quality database managed by European Environment Agency) were used to testing of the method and its implementation in the mentioned earlier Decision Support System. The calculations were performed for computational grids designated within the Mazovian Voivodeship using pre-adopted three grids with different scales and ranges. Fig. 4 shows an example of the average annual concentration distribution of PM10 in the Warsaw and the surrounding area, obtained by modeling based on the LUR method.

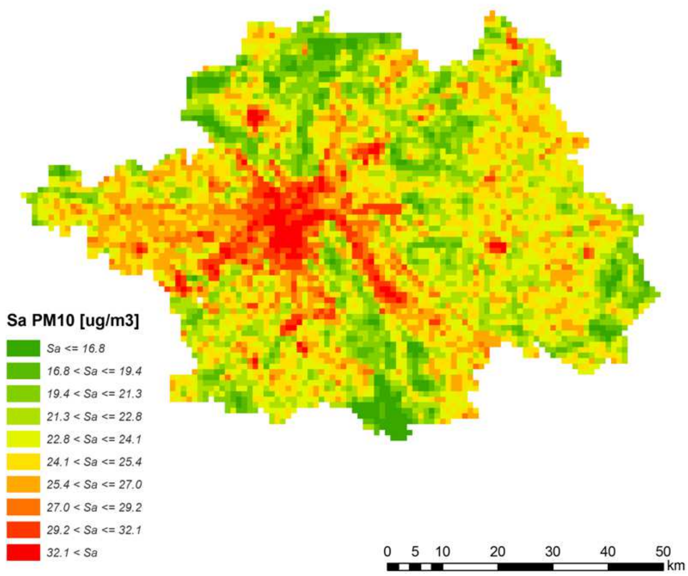

Figure 4. Annual average of PM10 in Warsaw and the surrounding area

3.2. Beta parameter method

So called Beta parameter method can be potentially used for the objective assessment and analysis of spatial representativeness of air quality measurement stations. Similarly, as in case of LUR, it is based on the use of publicly available data and information, such as land use map: Corine Land Cover and much more detailed Urban Atlas. It is based on the assumption that there is a relationship between the annual average values of pollution concentrations, measured at a given position, and calculated parameter characterizing the cover and land use in the neighborhood of the measurement point. Its foundations were developed as part of the work presented in [4] and [10]. This method involves determination of the parameter β, characterizing the impact of land use on concentration of a selected pollutant in analyzed area. This indicator is calculated using the formula (2).

$$\beta = \log\left[1 + \frac{\sum_i a_i \times n_{RCLi}}{\sum_i n_{RCLi}}\right] \quad (2)$$

where:
i – RCL class index (land use class), occurring in the area,
n_{RCLi} – number of pixels with the class in the area,
a_i – inpact factor of the considered area with specified RCL class for air pollution.

The β parameter is the logarithm of the sum of weighted and normalized distributions of RCL classes related to land use. The coefficients a_i are used to determine the weight of the impact of area with a particular RCL class on the pollutant concentration. Procedure for determining the coefficients for individual pollutants consist of two phases. It is assumed that the classes representing areas with a negligible impact on the emissions, i.e. semi-natural forests, green areas, water and wetlands have assigned values about zero. The classes associated with discontinuous urban fabric, which can associated with emissions coming from individual heating systems of buildings and traffic, have the coefficient with value one, for the execution of the subsequent normalization. In further calculations the emissions of air pollutants from various sectors of the economy and of human activity are taken into account, which allows the calculation of the relative emissions from different land classes, and the determination of the initial values of the a_i coefficients. The next step is to optimize the values of the coefficients, taking into account the parameters β, calculated for each measurement station using the long-term (e.g. five-year) average concentration values. The optimization process is based on perfect matching of the coefficients to the designated trend line parameter that minimizes the RMSE (root mean squared error), using algorithms implemented in the solver type tools. Then the value of β parameter can be calculated for all cells of a computational grid covering the analyzed area. Fig. 5 shows an example of the results of the parameter β initial calculation, made with a prototype of decision support system with respect to PM10 pollution. Results obtained with use this method allow, inter alia, on the analysis of the representativeness of measuring stations in air monitoring network. An example for NO2 measurement site in Piastów is presented in Fig. 6.

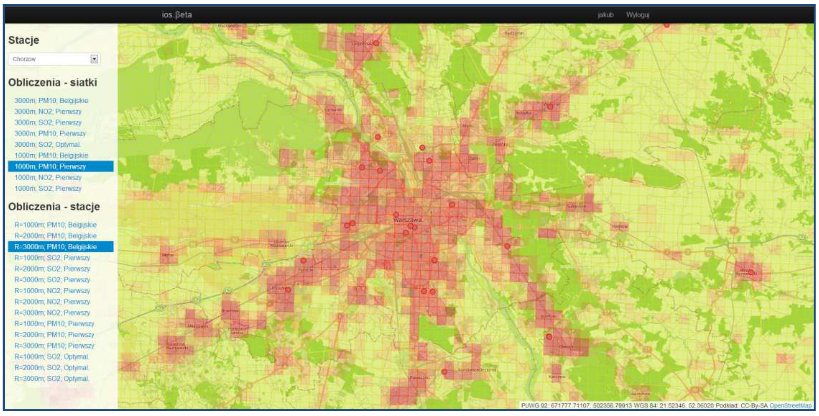

Figure 5. Sample screen of the prototype system - β parameter calculation results for Warsaw and the surrounding area

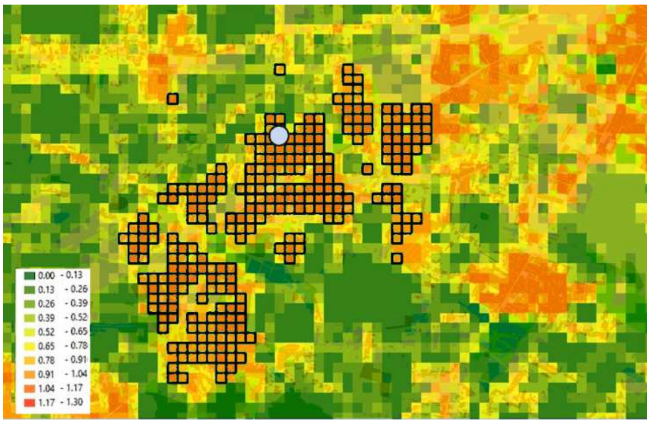

Figure 6. Estimation of spatial representativeness of NO2 measurement station in Piastów using calculation of β parameter based on Urban Atlas (grid 250 m)

This method, in addition to determining the differences in the impact of land use on air pollution, also allows to obtain the spatial distribution of substance's concentration in the air. For this purpose a geostatistical interpolation of measurements from the stations located in the study area are processed on the basis of the β parameter. Another way to use this method is to increase the resolution of the results obtained by means of mathematical dispersion models.

4. Conclusion

The article presents examples of the practical possibilities of using spatial data processing systems and tools for the management of the air quality monitoring network or analysis and interpretation of its results. The purpose of these methods is, inter alia, standardization and objectifying of determination of the representativeness of stations. The limitations that affect the calculation results must be taken into account, e.g. that meteorological conditions or the advection of pollutants and greater accumulation in certain parts of the studied areas are not included into the analysis. The quality and timeliness of input spatial information is very important as well as and the number of available measurement results. An important feature of these methods is the speed of obtaining the results of calculations and their ability to perform for the various areas of the country. These methods can also be used, for example, in automatically executed generation of pollution distribution maps, based on data from the fixed and mobile monitoring and spatial data - for the purpose of public presentation of current air quality information on geoportals or mobile devices.

REFERENCES

[1] Iwanek J., Kostrzewa J., Kobus D., Mitosek G., Parvi R., *Ocena jakości powietrza w strefach w Polsce za rok 2013. Zbiorczy raport krajowy z rocznej oceny jakości powietrza w strefach wykonywanej przez WIOŚ wg zasad określonych w art. 89 ustawy-Prawo ochrony środowiska*, Inspekcja Ochrony Środowiska, Warszawa 2014, http://www.powietrze.gios.gov.pl

[2] Bosch P., Gabrielson P., *Environmental indicators: Typology and use in reporting*, EEA, Copenhagen 2003

[3] Kristensen P., *The DPSIR Framework*, National Environmental Research Institute, Denmark Department of Policy Analysis, European Topic Centre on Water, European Environment Agency, 2004

[4] Janssen, S., Dumont, G., Fierens, F., Deutsch F., Maiheu B., Celis D., Trimpeneers E., Mensink, C.,. *Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation*, Atmospheric Environment 2012 no. 59, p. 492-500

[5] Spangl, W., Schneider, J., Moosmann, L., Nagl, C., *Final report, Representativeness and classification of air quality monitoring stations*, Umweltbundesamt GmbH (Ed.), Report REP-0121, Vienna, 2007

[6] Kobus D., *Systemy wspomagania decyzji w zarządzaniu środowiskiem i informacji publicznej oparte na monitoringu jakości powietrza*, Ochrona Środowiska i Zasobów Naturalnych, nr 34, Instytut Ochrony Środowiska-PIB, Warszawa 2008

[7] Kobus D., Skotak K., *The conception of decision support system for assessment and management of ambient air quality*, Information Systems in Management, 2012, Vol.1, nr 4, p. 305-317

[8] Mukerjee S., Smith L., Johnson M., Neas L., Stallings C., *Spatial analysis and land use regression of VOCs and NO2 from school-based urban air monitoring in Detroit/Dearborn, USA* , Science of the Total Environment 2009 nr 407, p. 4642–4651

[9] Górecki T., *Podstawy statystyki przykładami w R*, Wydawnictwo BTC, Legionowo 2011

[10] Janssen, S., Dumont, G., Fierens, F., Mensink, C., *Spatial interpolation of air pollution measurements using CORINE land cover data*, Atmospheric Environment 2008 no. 42, p. 4884-4903