JERZY STRASZKO, URSZULA GABRIEL, ANNA POŚWITAJŁO

Analysis of imission measurement results in Cracow region by instructural methods

Keywords: air monitoring, Cracow region, results analysis, instructural methods;

Processes and phenomena are usually described by structural methods. Models are built on the basis of theory. Experimental data in this case serve for verification of theory and determination of coefficients of the equations describing process or phenomenon. This method is time consuming and requires significant expenditures. Other, instuctural methods instead, consist in deduction on phenomenon directly on the basis of measurements results. There are known fundamentals of this depiction of the problem and implementations in form of computer software. These methods are classified as artificial computational intelligence. They are applied in various branches. In this work the methods are used for analysis of the results of air–monitoring stations measurements in Cracow region.

General part

In air quality assessment codified systems oblige. They use mathematical methods of modeling of atmospheric transport and methods of direct assessment based on the measurements results. In theory of contaminants propagation structural models of Gauss, Euler or Lagrange are used. They were described in the books [1–3], monographs [4–7], doctoral thesises [8–10] and numerous publications [11–16]. Applications of these models are also codified in the form of recommended computer software.

In case of analysis of imission measurements results other issues appear. They refer especially to determination of probabilistic characteristics of measured values. First works in this field were published by Larsen [17, 18]. He used the one–dimensional distributions and stated that the lognormal distribution was most often fulfilled by the experimental data. Many publications later on devoted attention to this issue.

The values recorded at monitoring stations are random variables. Measured in time they determine time series. In theories applied to this problem deterministic models, and more generally structural models, predominate. In case of complex processes or phenomena (air contamination propagation belongs to this type of processes) it is usually difficult to build sufficiently precise theory. It should be added that obtainment of data for calculations with use of structural models is expensive and sometimes also impossible.

This types of problems occur in many disciplines. For description of these phenomena instructural analysis systems, using directly the experimental data, have been elaborated [19, 20]. There exist good bases of these methods and their applications in the form of computer software. These methods are classified as artificial computing intelligence and they found many applications. They are still developed.

Conventional (traditional) statistical methods are insufficient nowadays. The aim of traditional data analysis is most often hypothesis verification or theory verification. However in practice it is in question to obtain as quickly as possible an easy hint necessary for making decisions. Methods of data mining provide this kind of information. New methods do not eliminate the old proven systems. They require however their more effective use.

Irrespectively of adopted method of analysis, adequate quality of data is the key factor influencing the results, and so the quality of the build model. Before the calculations the type of model, for instance deterministic equation or neural network has to be selected, and the results have to be prepared so that they could be useful i.e. preprocessing is performed. Then analysis of data correctness and unambiguousity is carried out. Gathered data cannot contain any failures. Occurrence of untypical observations needs to be investigated too. If gathered data are incomplete, the missing values can be filled up, for example by replacing them with average, median or removing the cases in which failure occurs.

Second stage consists in model building and its evaluation. Different models, for example different neural networks are tested. The best model is selected, usually basing on statistical evaluation. Applying the built models usually two kinds of tasks are solved: more direct one, namely prediction and more general consisting in knowledge discovering. Prediction refers to the situation when we

Prof. dr hab. inż. J. Straszko, dr inż. U. Gabriel, mgr inż. A. Poświtajło – Instytut Inżynierii Materiałowej, Politechnika Szczecińska

have complete data from the past, and we currently want to predict the values on the basis of incomplete data we have. In this way of predicting regression and classification are used. A specific case of regression is prediction of future values of time series. As mentioned before the values of variables measured at air monitoring stations determine the time series.

Knowledge discovery is the discerning description of data, indicating the structures, dependencies and information hidden in them. This was the main aim of this work. The results obtained by artificial computing intelligence can be useful as an essential complement of currently performed analyses during air quality assessment.

Experimental data

The calculations were based on the results of measurements performed according to the standing regulations in the period of September 2007 – April 2008 at the automatic air monitoring stations in Cracow region [21]. In figure 1 locations of all the stations, including the planned ones, are presented.

In figure 2 distributions of 1–hour concentrations of contaminants and distributions of meteorological parameters registered at Nowa Huta station are presented as example.

According to evaluations of Voivodeship Environmental Protection Inspectorate in Cracow, in year 2008 in Lesser Poland Voivodeship exceedances of permissible pollution target levels and permissible pollution levels extended by tolerance margin for the following substances occurred: particulate matter PM10 – average concentration per calendar year, particulate matter PM10 – 24–hour concentration, nitrogen dioxide – average concentration per calendar year and annual average concentration.

This work does not concern the assessment of air contamination according to the standing procedures. The aim was the analysis of the data by artificial computational intelligence methods indicating the structures, dependencies and regularities hidden in them. The first stage was determination of probabilistic characteristics of measured values by traditional statistical methods. Known one–dimensional distributions were tested [2]. It was demonstrated, that the lognormal distribution was most often fulfilled by the experimental data registered at monitoring stations.

Analysis of functional dependencies

Investigations of interdependencies between registered variables, i.e. knowledge discovering indicating the structures and regularities hidden in data, were called the functional analysis. To perform the analysis by artificial neural networks method the describing and described variables have to be defined. Time, wind speed, wind direction and temperature belong to describing variables. Contaminants concentrations should be classified in general to described variables.

While describing contaminants concentrations measured at monitoring stations usually we do not have important information such as number of emitters located in the examined area, their technical characteristics and the quantity of emission, which are the basis of structural models of atmospheric transport of contaminants.

During the construction of the way of analysis in this work the redundancy principle was used. It consists in that some of the measured values can complete the missing information. For example the temperature is connected with fuel combustion for heating purposes. Hence this could complete the results concerning the contaminants emission from these processes. It should be added that some contaminants react in air. Than their concentrations interdependency is natural.

Statistica Neural Networks computer software was used for calculations [23-26]. MLP (Multilayer Perceptron), RBF (Radial Basis Function) and GRNN (Generalised Regression Neural Networks) were tested. Good accuracy was usually obtained for GRNN networks. Therefore these networks were used the most often.

The way of calculations performing has been shown at the example of dependencies of 1-hour contaminants concentrations registered at Nowa Huta station. In figure 3 architecture of the built neural network is presented.

The first number stands for the number of describing variables, and the second one for the number of hidden neurons. Statistical evaluation of the network is given in table 1.

In figure 4 concentration values of NOx calculated by neural network and determined experimentally are compared.

The satisfactory accuracy was obtained. The built neural model could have been used in further calculations. The first stage was estimation of parametric sensitivity. The results are listed in table 2.

Table 1. Statistical evaluation of GRNN 9/2472 network

Parametr	Tr. NO_x	Ve. NO_x	Te. NO_x
Data Mean	86.29126	84.48058	84.17233
Data S.D.	101.0836	98.00799	96.04125
Error Mean	-0.6291	-3.01071	-2.15109
Error S.D.	21.01881	42.61291	41.78545
Abs E. Mean	13.08603	25.17274	24.43117
S.D. Ratio	0.207935	0.43479	0.435078
Correlation	0.979103	0.90169	0.90073

Table 2. Results of parametric sensitivity analysis for GRNN 9/2472 network

Parametr	SO_2	CO	PM10	WS	WD	Pressure	Temperature	Humidity	Time
Rank	9	1	2	7	3	8	5	6	4
Error	21.63131	72.14044	48.23425	32.29759	44.08896	28.29231	34.08931	32.69051	38.41671
Ratio	1.028888	3.431342	2.29425	1.536226	2.09708	1.345717	1.621449	1.554916	1.827281
Rank	**9**	**1**	**3**	**6**	**2**	**8**	**5**	**7**	**4**
Error	**42.887**	**73.81669**	**49.58733**	**47.06974**	**50.00605**	**43.88281**	**47.36082**	**46.0698**	**47.46838**
Ratio	**1.004334**	**1.72865**	**1.161243**	**1.102286**	**1.171049**	**1.027654**	**1.109102**	**1.078869**	**1.111621**

The rows marked in black refer to the training set, and the rows marked in red refer to the validation set. In both cases estimations are similar. This means that the calculations were performed properly.

Statistica Neural Network computer software enables visualization of obtained dependencies. In figure 5 the correlation of 1-hour concentration on NOx, CO and dependencies of NOx imission level on temperature and wind speed are shown.

Strong correlation between NOx and CO indicates, that most of these contaminants were emitted from the same sources. Temperature, in this case, does not fulfill physical parameter, but present influence of fuel combustion for heating purpose

High concentration of NOx occurred at low wind speed values. Low concentrations of nitrogen oxide in the area of the station was registered at wind speed values above 3 m/s.

The same calculations were performed for all contaminants registered at the automatic stations located in Cracow region. The obtained results were similar. Because of high influence of wind speed on aerosanitary conditions, this similarity was presented on example of results concerned dependencies of imission level on wind speed. In figure 6 results for Olkusz station are presented.

Similarly to Nowa Huta station low 1–hour concentration of NOx and CO was registered at wind speed values above 3 m/s.

In figure 7 results concerning 1–hour concentration of NOx and SO_2 registered in Trzebinia station are shown.

In this case low 1–hour concentrations of NOx and SO_2 were registered at wind speed values above 4 m/s.

In figure 8 results concerning 1–hour concentration of dust PM10 measured in Trzebinia and Kraków-Krowodrza station are presented.

In this case low 1–hour concentrations of PM10 occurred at wind speed values above 4 m/s. It is necessary to add, that wind speed at correct performed measurements, indicate ventilation of the region of station by air mass flow.

At the end, correlation between O_3 and SO_2 and NOx concentration are presented (Fig.9)

It is clearly visible that high O_3 concentration occurred at low SO_2 and NOx concentration. Moreover, O_3 is consumed during SO_2 and NOx oxidation in air.

Construction of imission fields

On the basis of measurement results the levels of imission in the region of the station are determined. Building of the station is expensive. For this reason the number of stations in the region is limited. To assess the concentration levels in the area located between the stations it is necessary to generate the imission fields on the basis of measured values. Such data are crucial for assessment of aerosanitary conditions in the region. The way of calculations is showed at the example of average annual concentration of SO_2. Computer software Statistica 6 was used. The data necessary for calculations are listed in table 3. In the first two rows coordinates describing the stations location are given. It should be added that the way of coordinates determination is generally arbitrary. Geographical coordinates or values read straightly from the map can be used. In table 3 such values are given. In the last column the average annual concentrations of SO_2 are placed. Automatic stations are indicated in bold.

The results of calculations are presented in figure 10.

Spatial graph (figure 10a) enables evaluation of matching degree of generated surface to experimental data. Some points deviate from the generated surface. This occurs in case of considerable differences of concentrations. Surface fitting to these points is obtained by performing the calculation in a few stages. In subsequent stage the values of concentrations calculated for the nearest neighboring area of the point are taken into consideration. In this work the results without correction are presented. The results shown in figure 10a were plotted on the map of region. Figure 10b shows projection of generated surface on the X–Y surface, figure 10c the result of overlapping the map of the region on this surface performed using Corel software, figure 10d – obtained map of imission field. The high concentration of annual SO_2 concentration occurred in the southern part of voivodenship (from Cracow to Zakopane) and in surroundings of Tarnow.

In the same way the imission fields for other contaminants were obtained. The results were showed in figure 11.

Emission fields of PM10, NO_2, NO, SO_2 are similar what indicates that primary amounts of contaminats were emitted from the same sources. The high concentrations occurred in the middle part of the region, so in a different way that for SO_2.

In the Cracow region the energy sources are traditional fuels. Industrial plants are equipped in treating installations. Monitoring network is well organized. For this reason the Cracow region can be treated as model region.

Conclusions

The results of analysis of contaminants concentration and meteorological parameters measurements performed in Cracow region in the period September 2007 – April 2008 are presented. In this region traditional fuels are used. Industrial plants are equipped in treating installations and monitoring network is well organized. For this reason the Cracow region is a good example for carrying out the analysis.
Interdependencies between variables registered at each station and in region were investigated. Artificial neural networks were applied. Sufficiently accurate models were obtained (correlation coefficient 0,90 and higher). Ranking of describing parameters for considered contaminants was determined and dependencies of contaminants concentrations on describing variables were defined. Occurrence of maximal imission at low temperature values, connected with fuel combustion processes for heating purposes, was stated. It was demonstrated that at low wind speed values high concentrations of all contaminants occurred. Low concentrations were registered at wind speed values higher than 3 m/s or 4 m/s. For the stations and for the region strong correlation between NOx, NO, PM10 and CO was stated. This means that primary amounts of the contaminants were emitted from the same sources (fuel combustion processes for heating purposes and motor transport). In case of SO_2 no clear interdependency on other contaminants concentrations was noted.

Presented way of analysis of measurement results at air monitoring stations enables obtaining of valuable information about aerosanitary conditions in industrial regions. This could be useful during air quality assessment in such regions.

LITERATURA:

[1] Juda J., Chróściel S.: Ochrona powietrza atmosferycznego, WNT, Warszawa 1974
[2] Zonnatti P., Air pollution modeling. Theories computational method and available software. Van Nostrad Reinhal, New York 1974
[3] Pasquill F., Smith F. B.: Atmospheric diffusion, Holsed Press, Ihon Wiley and Sons, New York 1986
[4] Bartnicki J.: Heavy Metal Eulerain Transport Model – HMET. Model Description and Results, DNMI Research Raport No 65, June , Oslo 1998
[5] Skrzypski J.: Analiza i modelowanie pól imisji powietrza w dużych miastach, Wyd. PAN Oddział w Łodzi, Łódź 2002
[6] Zwoździak J.: Metody prognozy i analizy stężeń zanieczyszczeń w powietrzu w regionie Czarnego Trójkąta., Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2002
[7] Kamiński W., Strumiłło P., Toczek E.: Zastosowanie systemów sztucznej inteligencji obliczeniowej w rozwiązywaniu wybranych problemów ochrony atmosfery, PAN w Łodzi, Łódź 2005
[8] Paulo L.A.: Analiza pól imisji zespołów źródeł punktowych, praca doktorska, Politechnika Szczecińska, Szczecin 1996
[9] Dziubakiewicz D.: Dynamika skażania powietrza atmosferycznego, praca doktorska, Politechnika Szczecińska, Szczecin 2004
[10] Bykowszczenko N.: Identyfikacja modeli skażania powietrza w regionach przemysłowych, praca doktorska, Politechnika Szczecińska, Szczecin 2007
[11] Straszko J., Paulo L.A., Witkowski A.: Models of distribution of pollutions in the air, *Env. Protec. Eng.*, 1-2, 71, 1996
[12] Pyta H., Zwoździak.: Hybrydowe modele rozkładu stężeń zanieczyszczeń powietrza, *Arch. Ochr. Środ.*, 23 (3/4), 59, 1997.
[13] Straszko J.: Stochastyczne modele rozprzestrzeniania zanieczyszczeń w powietrzu, *Inż. Chem. Proc.*, 21, 393, 2000
[14] Straszko J., Dziubakiewicz D.: Ocena skażania atmosfery w regionach na podstawie analizy zmienności pól imisji zanieczyszczeń, *Problemy Ekologii*, 1(7), 21, 2003
[15] Madany A., Bartochowska M.: Przegląd polskich modeli rozprzestrzeniania zanieczyszczeń. Prace naukowe Politechniki Warszawskiej, 19, 1995
[16] Łobacki L.: Wskazówki metodyczne dotyczące modelowania matematycznego w systemie zarządzania jakością powietrza, Instytut Systemów Inżynierii Środowiska Politechnika Warszawska., Warszawa 2003
[17] Larsen R.: Analizung Air Pollutant Concentration and Dosage Data, *Air Pollut. Control ASS*, 17, 85, 1967
[18] Larsen R.: New Mathematical Model of Air Pollutant Cocentration Axeraging Time and Frequecy, Air Pollut. Control ASS, 19, 24, 1969
[19] Gajda B.: Prognozowanie i symulacja a decyzje gospodarcze. Wydawnictwo CH Beck Warszawa, 2001
[20] Valnus A., Demski T.: Data mining – inteligencja biznesowa, *Magazyn Przemysłowy*, 3 , (56), 38, 2004
[21] Wojewódzki Inspektorat Ochrony Środowiska w Krakowie. *http/www.kraków.pios.gov.pl*
[22] Sobczak K.: Stochastyczne równania różniczkowe, WNT, Warszawa 1996
[23] *STATISTICA – System description*, Statsoft. Retrieved , November 1, 2007, from http/www.statsoft.pl
[24] *STATISTICA NEURAL NETWORKS™PL. Introduction to neural networks*, Statsoft. Retrieved, November 1, 2007, from *http/www.statsoft.pl*
[25] *STATISTICA NEURAL NETWORKS™PL. Problem guide*, StatSoft November 1, 2007, from *http/www.statsoft.pl*
[26] *STATISTICA NEURAL NETWORKS™PL. StatSoft*. Retrieved, November 1, 2007, from http/www.statsoft.pl

Table 3. List of data for calculations. Annual concentration of SO_2. Cracow region, September 2007 - April 2008.

Name of station	X	Y	SO_2
	cm	cm	µg/m^3
Krowodrza-Kraków, Szpital Jana Pawła II	**5.8**	**10.2**	**8.4**
Aleja Krasińskiego	**7**	**10**	**8.3**
Nowa Huta - Ul. Bulwarowa	**7.8**	**9.85**	**8.2**
Skawin, os. Ogrody	**5.25**	**8.8**	**11.4**
Tarnów Al.. Solidarności	**13.65**	**10**	**19.2**
Nowy Sącz , Pijarska	**11.85**	**5.2**	**13.4**
Zakopane Sienkiewicza	**6.6**	**1.15**	**10.7**
Olkusz Francesco Nullo	**3.6**	**12.3**	**3.3**
Trzebina	**3.6**	**10.5**	**9.6**
Szymbark	14.6	5.2	8.8
Nowy Sącz -Tarnowska	11.8	5.15	22
Bochnia -Kazimierza Wielkiego	9.65	9	19
Chrzanów -Grzybowskiego	2.5	10.65	31
Oświęcim -Więźniów Oświęcimia	1.2	9.3	64
Tarnów -Westerplatte	13.7	10.05	39
Brzesko -Głowackiego	11.15	9.2	33
Gorlice -Legionów	15.15	5.8	15
Nowy Targ -Szaflarska 5	6.9	3.35	16

Fig. 1. Locations of monitoring stations. Cracov region, year 2007-2008

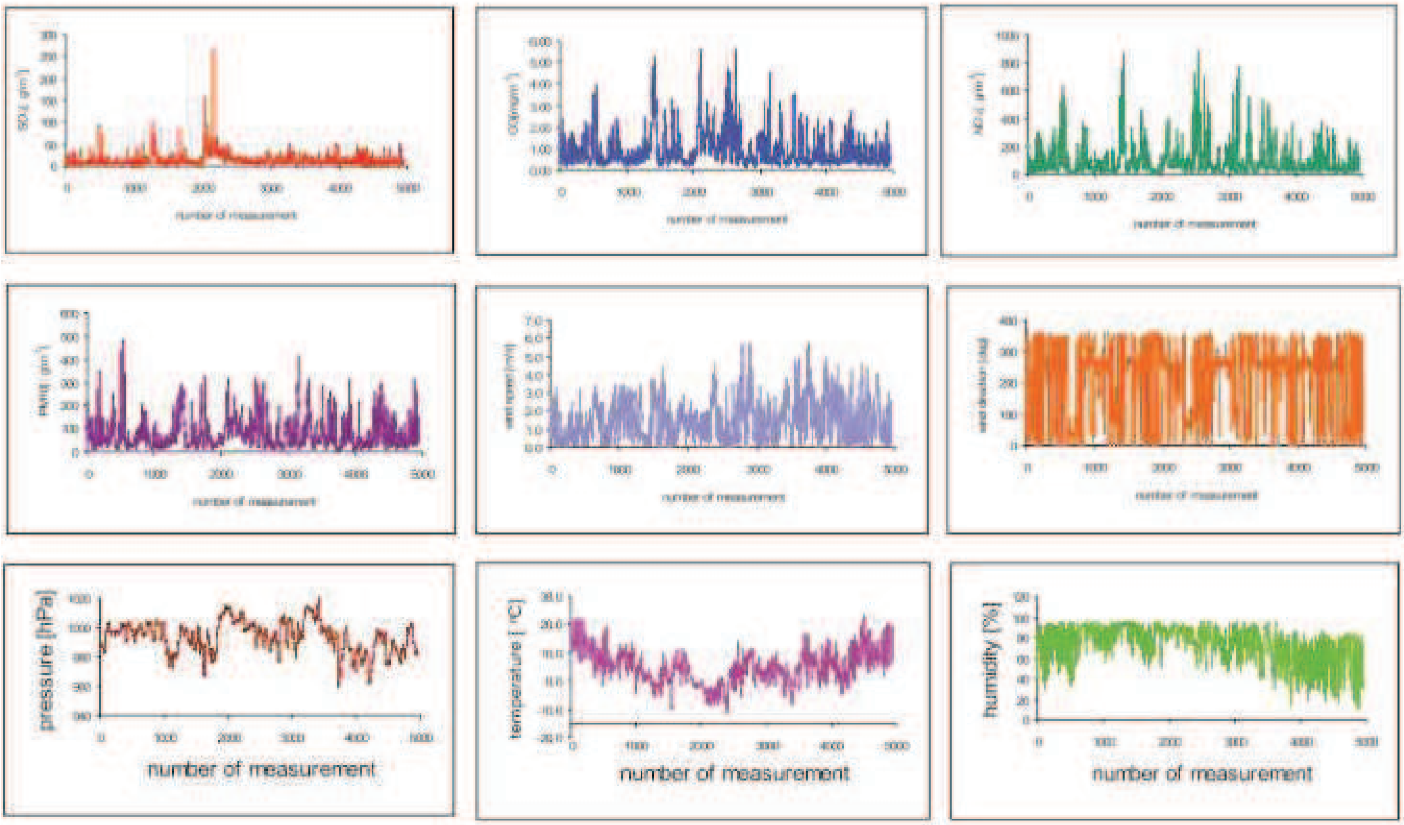

Fig. 2. Distributions of 1-hour concentrations od contaminants and meteorological parameters, Nowa Huta, September 2007 - April 2008

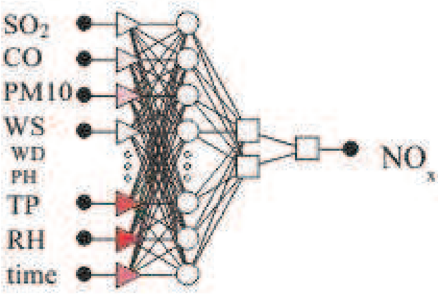

Fig. 3. Architecture of GRNN 9/2472 network

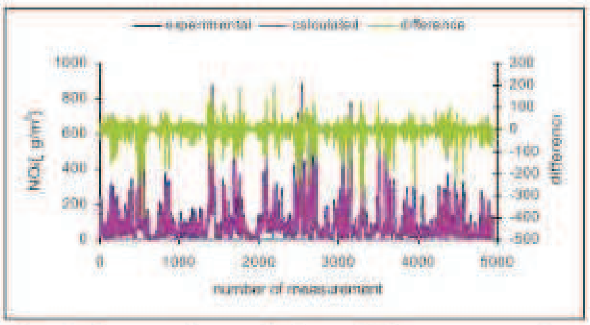

Fig. 4. Comparison of 1-hour NO_x concentrations calculated and determined experimentally, Nowa Huta station. September 2007 - April 2008

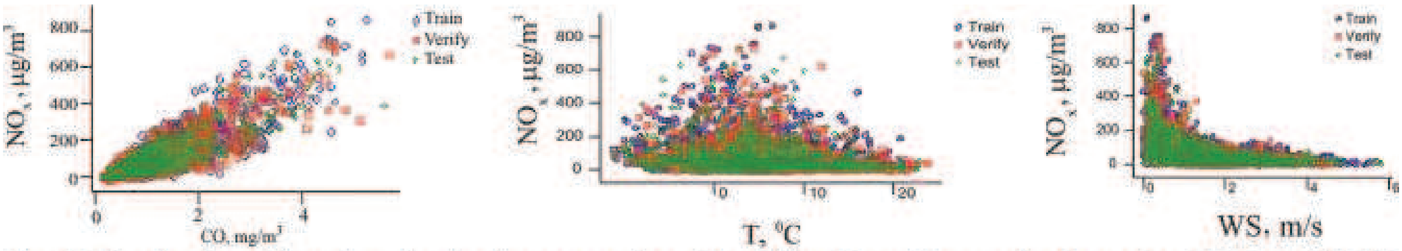

Fig. 5. Vizualization of interdependencies between registered variables. Nowa Huta station,September 2007 - April 2008

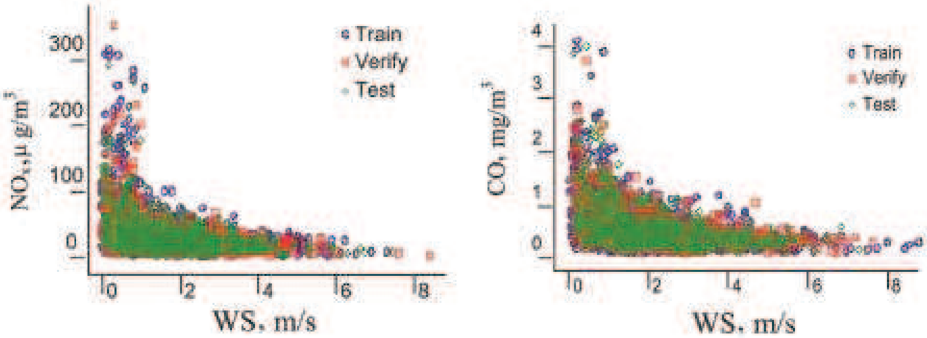

Fig. 6. Dependencies of 1-hour concentrations of NOx and CO on wind speed. Olkusz station, September 2007 - April 2008

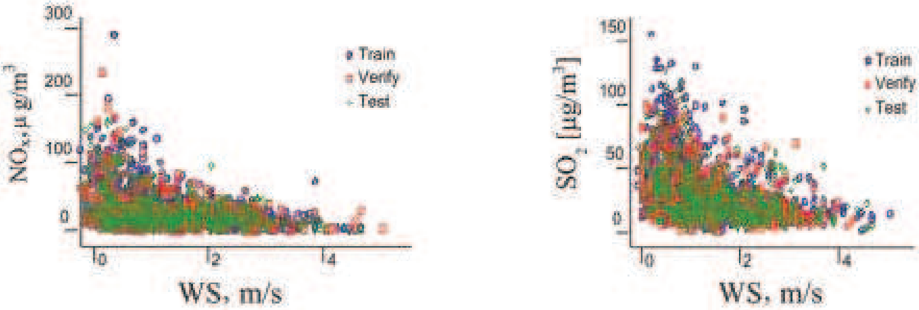

Fig. 7. Dependencies of 1-hour concentrations of NOx and SO_2 on wind speed. Trzebinia station, September 2007 - April 2008

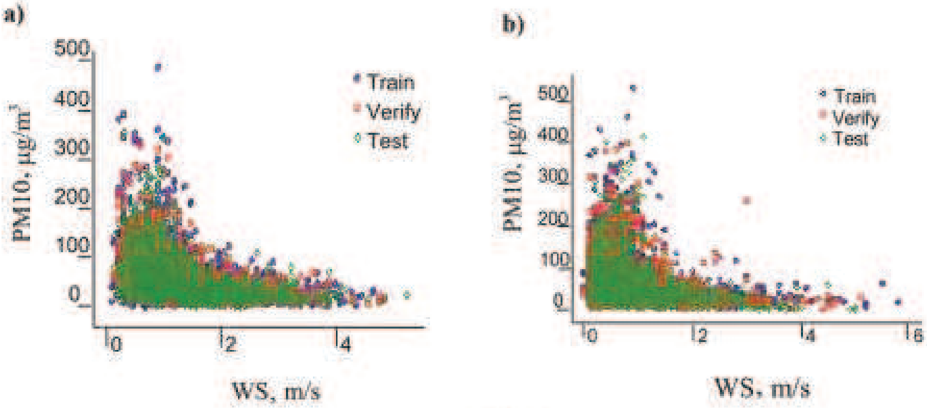

Fig. 8. Dependencies of 1-hour concentrations of PM10 on wind speed; a) Trzebinia station, b) Kraków-Krowodrza station, September 2007 - April 2008

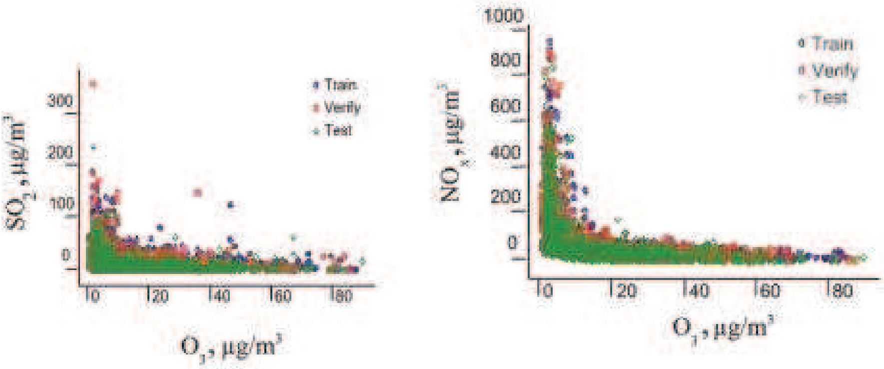

Fig. 9. Dependencies of O_3 on SO_2 and NOx concentration. Kraków-Krowodrza station, September 2007 - April 2008.

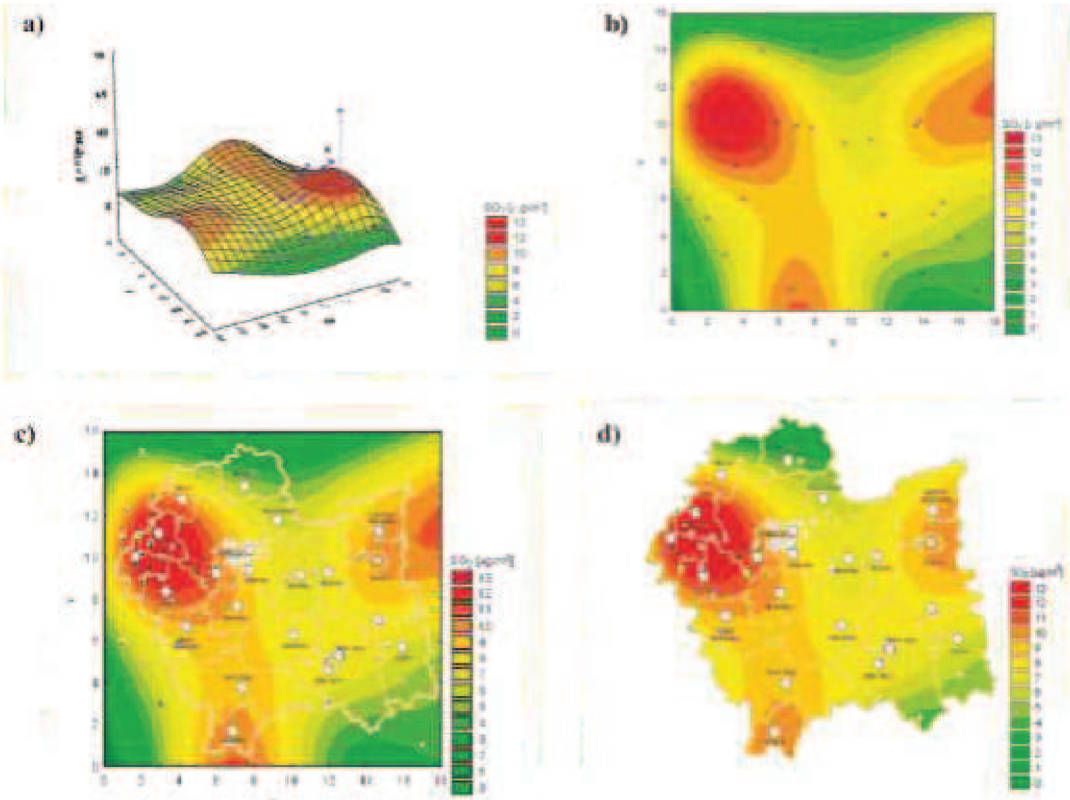

Fig. 10. Method of construction of concentration distributions. Cracow region. Annual concentration of SO_2.

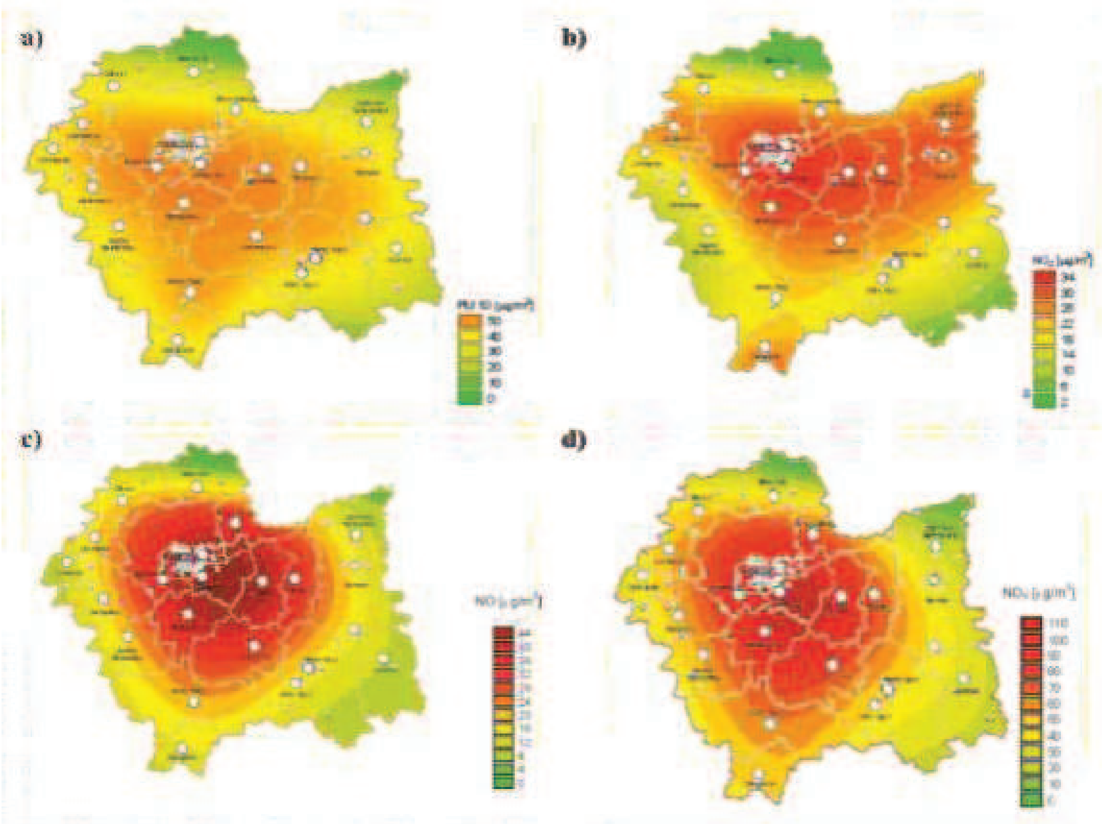

Fig. 11. Distributions of annual contaminants concentrations. Cracow region. Year 2008.
a) PM 10, b) NO_2, c) NO, d) NO_x