ANALYSIS OF IMISSION MEASUREMENT RESULTS IN CRACOW REGION BY INSTRUCTURAL METHODS

Jerzy Straszko [1)], Urszula Gabriel [2)], Anna Poświtajło [1)]

1) Institute of Chemistry and Environmental Protection, al. Piastów 42, 71-065 Szczecin, e-mail: Jerzy.Straszko@zut.edu.pl, tel. (+48) 91-449-4535

2) Institute of Materials Science and Engineering, al. Piastów 19, 70-310 Szczecin, e-mail: Urszula.Gabriel@zut.edu.pl, tel. (+48) 91-449-4054

SUMMARY

The results of imission analysis results recorded at monitoring stations in Cracow region in the period of September 2007 - April 2008 have been presented. Contamination levels of SO_2, NO_x, PM10, O_3 and meteorological parameters: wind direction, wind speed, temperature and atmospheric pressure were examined. The measured magnitudes determine time series with periodic oscillations and oscillations in shorter periods. Dynamics of imission and meteorological parameters changes in time were described and dependencies of contamination levels of pollutants on meteorological parameters were analyzed. Instructural methods have been applied.

1. INTRODUCTION

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, requires significant expenditures. Another, instuctural method consists 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 programs. 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.

2. 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 programs.

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 programs. 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 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.

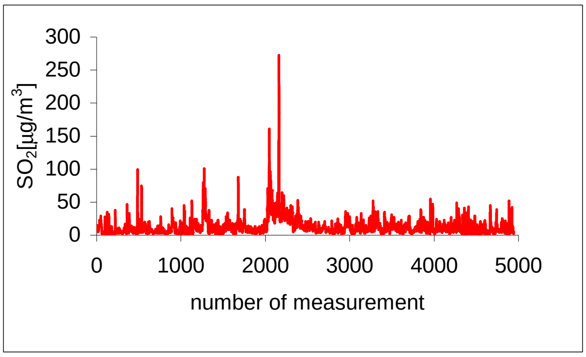
Knowledge discovery is the discerning description of data, indicating the structures, dependencies and information hidden in them. This was the main aim of this work. The results obtained by artificial computing intelligence can be useful as an essential complement of currently performed analyses during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

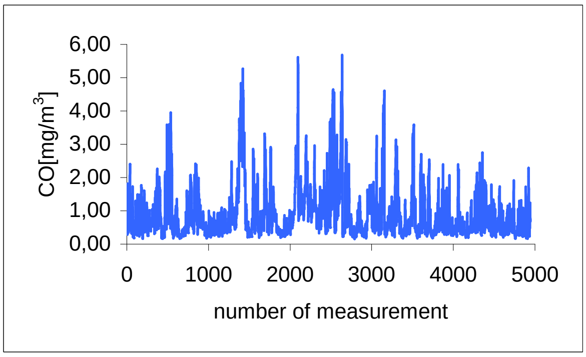
3. EXPERIMENTAL DATA

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.

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

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





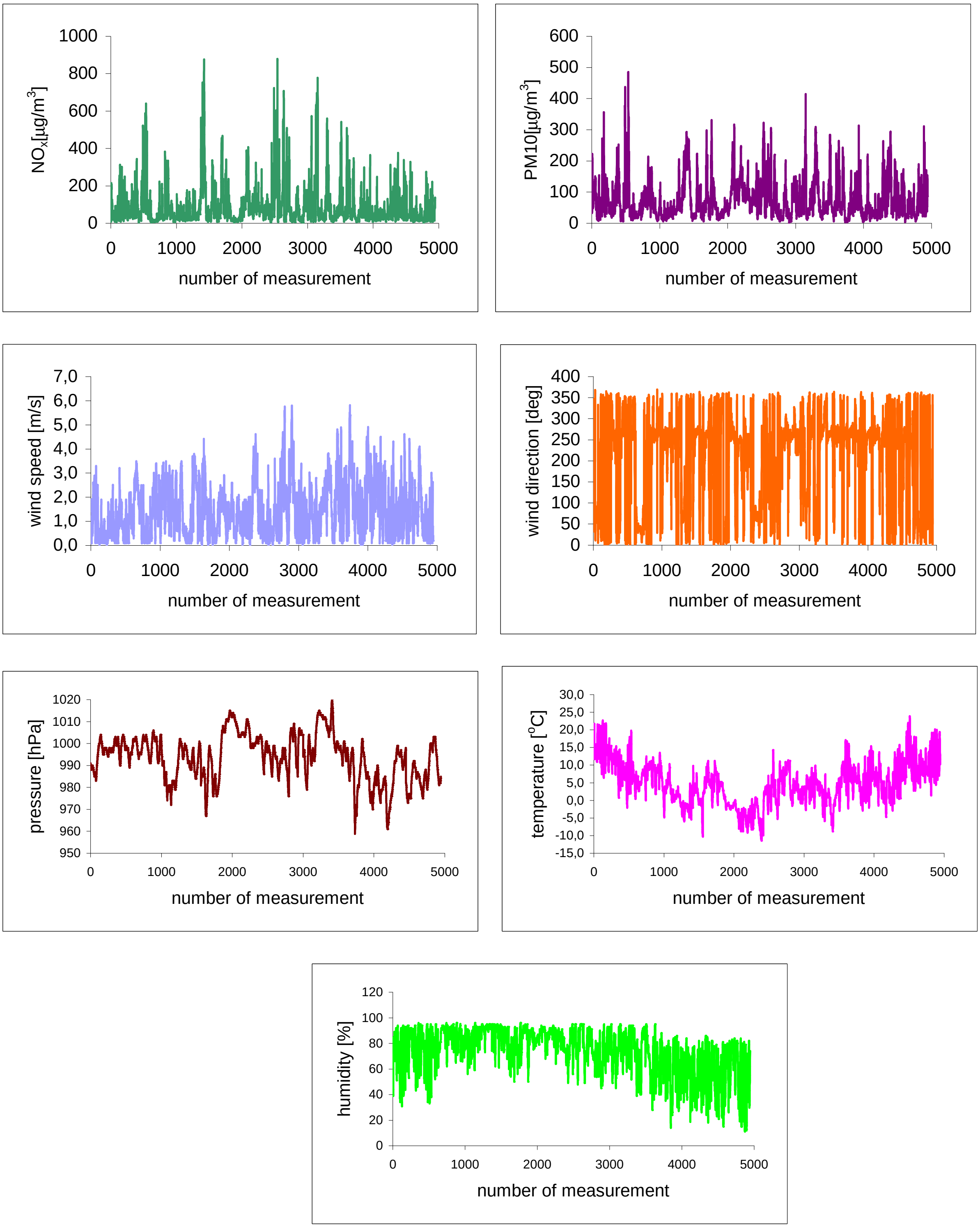


Fig. 2. Distributions of 1-hour concentrations of contaminants and meteorological parameters. Nowa Huta, September 2007 - April 2008.

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.

4. ANALYSIS OF FUNCTIONAL DEPENDENCIES

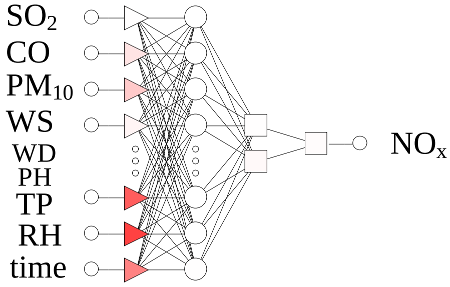
Investigations of interdependencies between registered variables, i.e. knowledge discovering indicating the structures and regularities hidden in data, were called the functional analysis. To perform the analysis by artificial neural networks method the describing and described variables have to be defined. 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.

Computer program Statistica Neural Networks 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.



Rys. 3 Architecture of GRNN 9/2472 network.

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.

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

In figure 4 concentration values of NO_x calculated by neural network and determined experimentaly are compared.

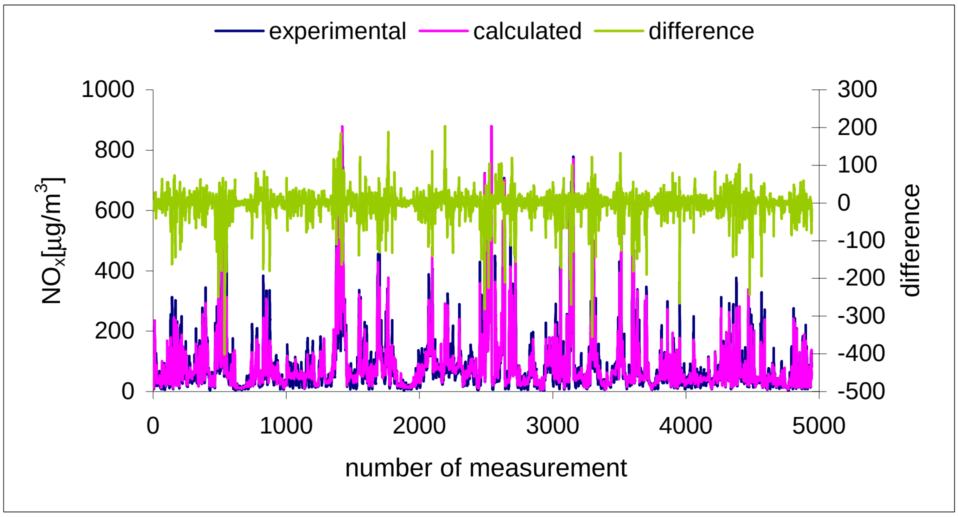


Fig. 4. Comparison of 1-hour NO_x concentrations calculated and determined experimentaly. Nowa Huta station. September 2007 – April 2008.

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 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 program enables visualization of obtained dependencies. In figure 5 such charts are presented.

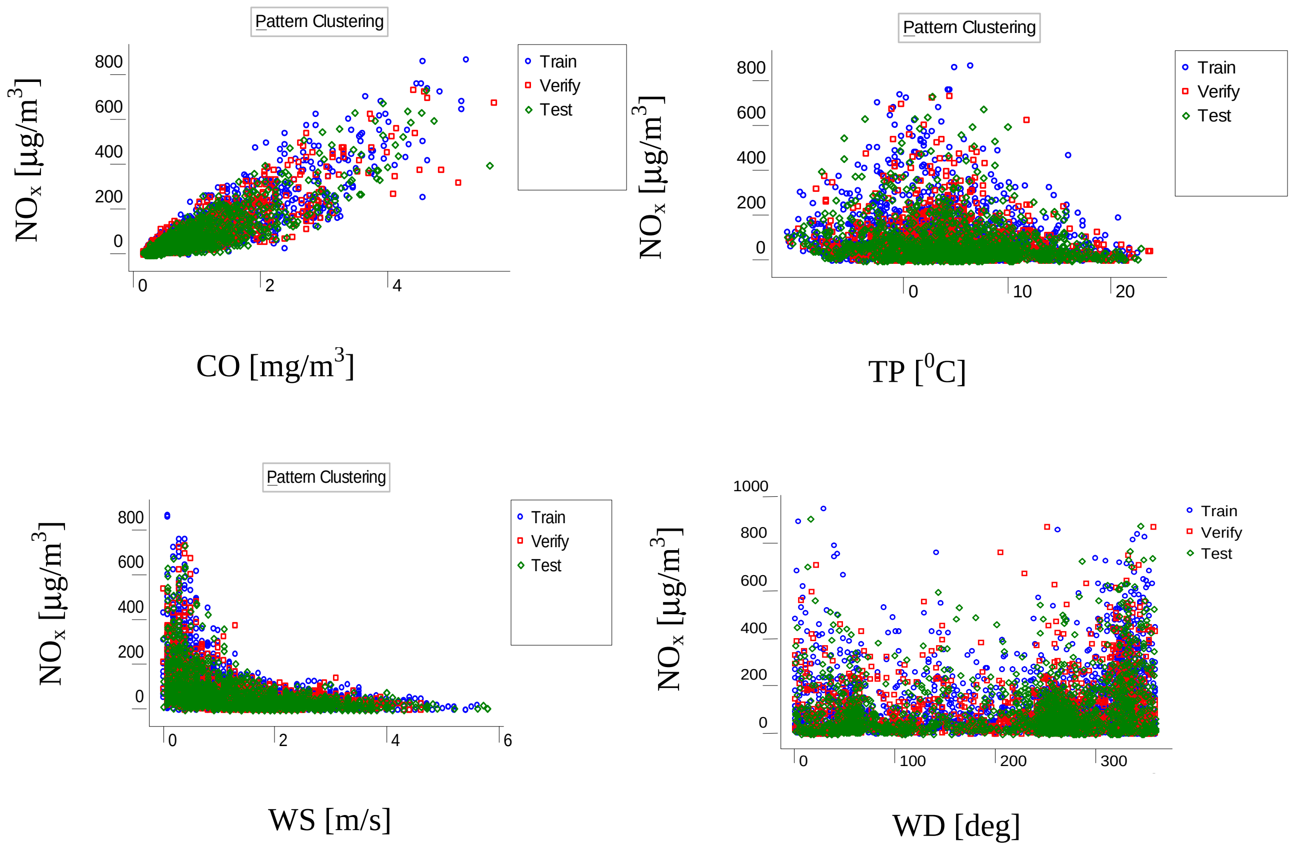
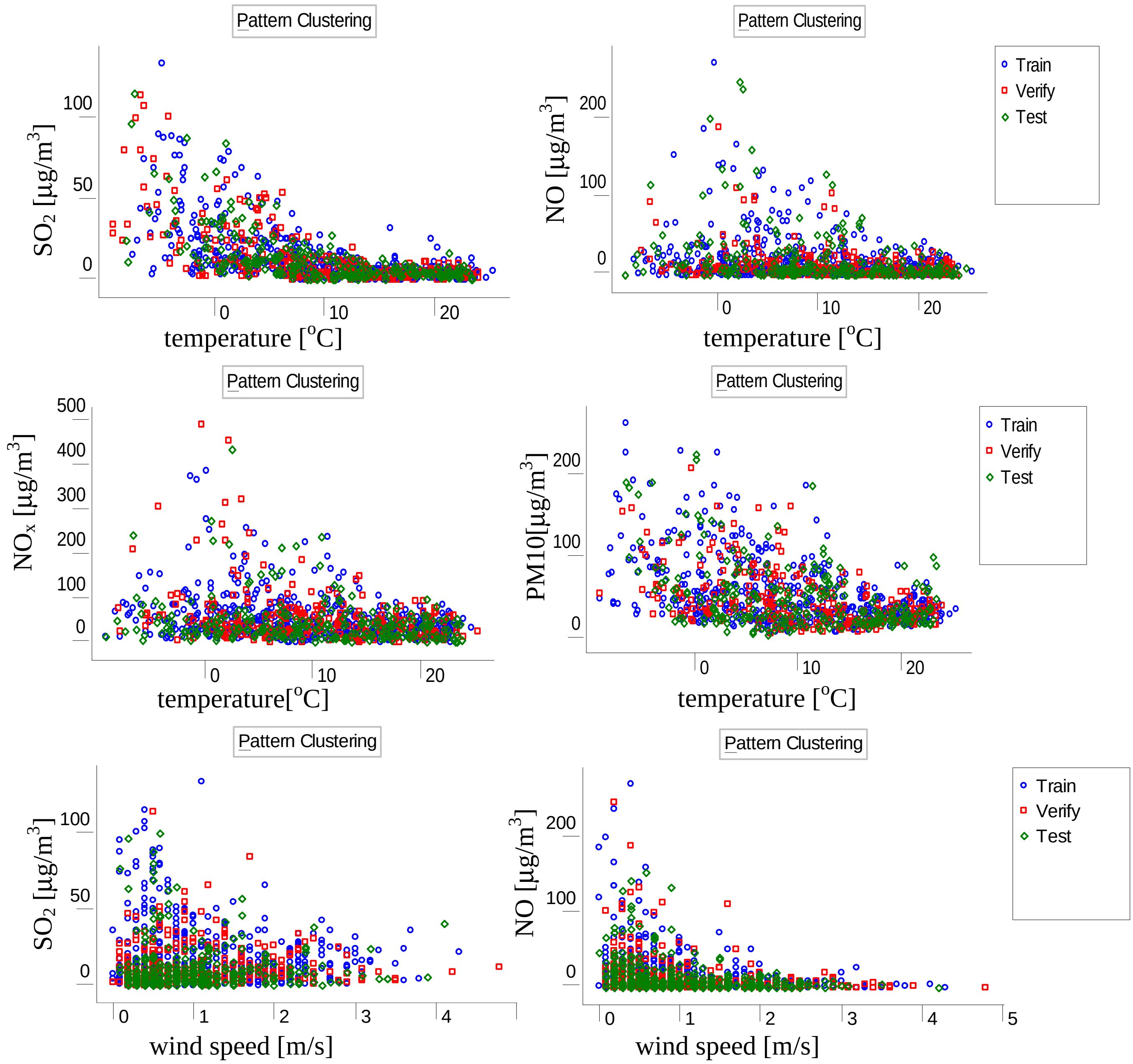


Fig. 5. Visualisation of interdependencies between registered variables. Nowa Huta station, September 2007 - April 2008.

Strong correlation between NO_x and CO indicates, that in the area of this station primary amounts of these contaminants were emitted from the same sources. High concentration of NO_x 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 last picture shows, that nitrogen oxides flowed into the area of the station from different directions.

The same calculations were performed for all contaminants registered at the automatic stations located in Cracow region. Correlations between daily concentrations of contaminants at stations took different forms. Different dependencies of imission levels on wind direction were also obtained. They were conditioned by location of imission sources against the stations. However for all the stations dependencies of contaminants concentrations on temperature and wind speed took same forms. The results were confirmed performing calculations jointly for the data registered at all the automatic stations. In figure 6 dependencies of daily concentrations of SO_2, NO, NO_x and PM10 on temperature and wind speed for the region are shown.



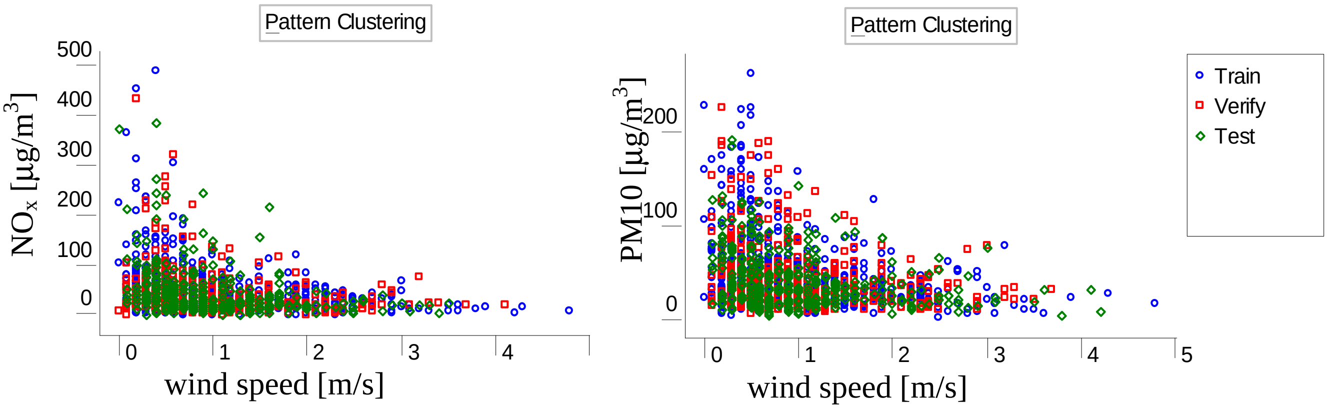


Fig. 6. Dependencies of daily concentrations of contaminants on temperature and wind speed. Cracow region, September 2007 - April 2008.

Considering the stations jointly similar charts were obtained. This means that determined dependencies for single stations refer also to the whole region.

CONCLUSIONS

1. The results of analysis of contaminants concentration and meteorological parameters measurements performed in Cracow region in the period September 2007 – April 2008 are presented.
2. It was demonstrated that the values measured at stations and in region are described with good accuracy by one-dimensional distributions, the most often by lognormal distribution.
3. 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).
4. Ranking of describing parameters for considered contaminants was determined and dependencies of contaminants concentrations on describing variables were defined.
5. Occurrence of maximal immision at low temperature values, connected with fuel combustion processes for heating purposes, was stated.
6. 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.
7. Contaminants dependencies on wind direction differed for the stations. They depended on locations of emission sources against the stations.
8. For the stations and for the region strong correlation between NO_x, 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.
9. 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.

LITERATURE

[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^{TM}PL. Introduction to neural networks,* Statsoft. Retrieved , November 1, 2007, *from http/www.statsoft.pl*
[25] *STATISTICA NEURAL NETWORKS^{TM}PL. Problem guide*, StatSoft November 1, 2007, *from http/www.statsoft.pl*
[26]] *STATISTICA NEURAL NETWORK^{TM},PL StatSoft.* Retrieved, November 1, 2007, *from http/www.statsoft.pl*