dr inż. Monika Maciejewska
Wroclaw University of Technology
Faculty of Environmental Engineering

Wroclaw, 5th November 2012

Self-report

I graduated from Wroclaw University of Technology (WUT), Faculty of Environmental Engineering in 1996, with distinction. My MSc diploma work was submitted in Specialization: *Atmosphere Protection Systems*. One year earlier (1995), I received MSc degree in *Energy and Environmental Systems* from Glasgow Caledonian University. It was the final result of one year long MSc studies, which I completed in the framework of student exchange program TEMPUS. The diploma was validated in Poland.

In 2001 I was awarded the PhD degree in Technical Science. My PhD thesis *Neural predictive models and neural interpolation method for the imission concentrations* was defended with the distinction in the Institute of Environmental Protection Engineering, WUT. The supervisor of this work was prof. dr hab. inż. Jerzy Zwoździak. On receipt of PhD degree I was employed as an academic teacher at the Faculty of Environmental Engineering, WUT (2001). This position is held by me until now.

My appreciable research achievements came in the period after receiving the PhD degree. Actually, several works with my contribution were published earlier. However, I do not value them high regarding creativeness as well as the rank of the publishing houses. Therefore these publications were not included in the presentation of my scientific achievements which I submit for the evaluation now, on the occasion of applying for the habilitation.

The principal domain of my scientific interest is data analysis in sensor measurements. The research effort in this field resulted in my major scientific achievements. From the perspective of an environmental engineer there were good reasons for choosing sensor data analysis as the domain of research activity.

The information about the state of the environment is highly valued in the world today. Wide spectrum of information about the environmental pollution is needed continuously and in real time. Currently, sensor systems are the most promising measurement approach, which offers such functionalities. There are two main reasons for the competitive character of this solution. These are the progress in sensor technology and the development of data analysis methods. The importance of the second factor results from the necessity to provide methods of information retrieval from the complex data delivered by multisensor systems.

The choice of data analysis as the main field of my research activity was inspired and largely encouraged by dr hab. Andrzej Szczurek, prof. WUT, the renowned specialist in the methods and techniques of air pollutants analysis, with particular emphasis on the application of gas sensors in environmental analysis. The great part of my publications was written in cooperation with him.

Based on the extensive literature study, I realized that the conversion of sensor data into information requires a multistage process. It is realized by the set of mathematical methods and techniques. Together, they form the data analysis system. Asking a question about the principle of the functioning of such a system I came across the concept of *machine learning*. In this framework there are developed statistical systems, which learn how to transform the data on the objects into the information about the properties of these objects. After being taught with carefully selected exemplary data, the systems are able to process the unknown data and retrieve the information from it. Moreover, systems may be periodically retrained. Thinking about the potential users of sensor systems in terms of the environmental information they may wish to obtain I singled out the following categories: the defined qualitative information (what kind of pollution), the defined quantitative information (how much of the pollution), undefined information (something changed in the environment). This division matches quite well the distinct groups of data analysis methods, namely: classification, regression and data exploration methods. Based on my own experience with data analysis I became convinced about the principle role of a *pattern* in the information retrieval. The pattern is the way the particular information is represented in the data. I found the valuable suggestions on the methods of patterns construction in the works dedicated to feature selection and extraction.

In course of practicing the data analysis for the array composed of many partially selective gas sensors I noticed quite early, that the data source is equally important for the selection of data analysis strategy, as the kind of information and the required precision of its retrieval. Moreover, the invention while working on data analysis provides the knowledge which is highly appreciated by the developers of the sensor part of the measurement system.

Traditionally, sensor systems utilize sensor response in the steady state. Although in this case the measurement is relatively easy, the approach restricts the information content in the sensor data. From the data analysis point of view, using steady state responses of partially selective sensors does not leave much freedom in choosing the *pattern*. Therefore the major effort is allocated to finding methods of classification or regression, which best operate in the available feature space. I analyzed the data of this kind, in the context of applying gas sensor arrays in environmental measurements, but not exclusively. There resulted a number of contributions from these studies. The examples of the addressed topics were: the recognition of air pollutants, such as benzene, toluene and xylene [21]*), the recognition of volatile compounds which are emitted while applying solvents [8], the determination of organic solvents composition [6], the fraud detection due to mixing diesel fuel with the heating oil [12], the recognition of coatings on wooden materials [19], the odor intensity assessment [20], the production process monitoring [16]. While searching for qualitative information I frequently applied artificial neural networks (ANN). From amongst available architectures I chose the most appropriate ones, e.g. radial basis function ANN [21], probabilistic ANN [19] generalized regression ANN [20]. As a means of reference I applied linear discriminant nalysis (LDA) [12, 19, 21]. In simple cases this method was used exclusively [6, 8]. Quantitative problems I most willingly solved with multiple liner regression (MLR) [8, 16], and partial least squares regression (PLS) [6]. While studying the variability structure in the data I favored principal component analysis (PCA) [12, 16, 19, 21].

The exploitation of the sensor signal offers much wider possibilities, compared with the steady state sensor response. The particularly interesting sensor signals are obtained by modifying the parameters of sensor operation or their conditions of exposure to the gas under test. An example of the way to obtain the highly informative gas sensor signal is the *stop flow* mode of operation. It has been investigated in the Laboratory of Sensor Techniques and Indoor Air Quality, lead by dr hab. A. Szczurek, prof. WUT for several years. In one of the first contributions which addressed this theme [5], I analyzed the information content of the single gas sensor signal obtained in the *stop flow* mode, showing that it varied from the time point to the time point of exposure to the test gas. This finding inspired the concept of a feature, which is the sensor response at a single time point of the exposure. This kind of feature allows for constructing a large number of patterns. Therefore it allows for selecting the feature space, where the readout of the particular information is most effective. The data analysis presented in [7] revealed that there exist a number of patterns which allow for the errorless identification of a particular air pollutant based on a single time point of gas sensor array response. In [9] I showed that it is possible to find separate sets of patterns which are indicative for distinct volatile organic compounds (VOCs). These two observations contributed to forming the concept of data analysis for the gas sensor systems with broad applicability. I presented the principal elements of this idea in [2] (Attach. 3, Books...). The potential of data analysis in such system were discussed in [3] and [4] using an example of qualitative and quantitative assessment of numerous VOCs mixtures. In my work with multiple sensor signals I emphasized the search for best feature spaces. My favorite strategy was the complete wrapper-kind search with LDA serving as the assessment method

[3, 5, 7, 9]. The consequent use of this approach in several contributions of mine was intentional. I headed at the comprehensive reasoning about the possibility to recognize different VOCs and their mixtures based in the complex data provided by an array of semi-selective gas sensor array. In [5] I proposed to apply the best feature sets selection also in the context of the quantitative information. The robust multiple regression was chosen for the assessment.

A number of the above mentioned contributions were accomplished in the framework of the project *Detectors and sensors for measuring factors hazardous to environment - modeling and monitoring of threats*, POIG.01.03.01-02-002/08-00. In course of the entire duration of the project (2008–2012), I am involved in Task 3a.1: *Sensor system for measuring organic substances*.

In my development as a scientist, I owe much to the cooperation with Analytische Chemie - Elektroanalytik & Sensorik, Ruhr Univeristat Bochum, Germany, which is led by prof. dr. hab. Wolfgang Schuhmann. The cooperation was initiated during my postdoctoral fellowship in the NOVTECH Project (2004–5) and it continued in the following years (2005–8) during my 3–4 months long summer research visits in Bochum.

The starting point for the common research was the idea to adopt the sensor array concept to the set of elements of enzyme-polymer structure. The structures were built of materials investigated as the potential components of chemosensing layers in biosensors. Upon visualization with scanning electrochemical microscopy (SECM) there was revealed the spatially inhomogeneous activity of the structures in analyte solutions. Therefore it was justified to interpret the elements of the structures as the individual highly cross-selective sensors.

I analyzed the patterns of enzymatic-electrochemical activity of enzyme-polymer structures with an objective to attain the mathematical selectivity of glucose determination in the presence of the interfering compounds: ascorbic acid, maltose and 2-Deoxy-D-glucose in solutions. The scope of this research work was not directly related to *environmental engineering*. However, it defined a challenging context where I could practice data analysis in relation to the generic classes of problems which are also encountered e.g. in sensor measurement of air pollution. These were: i) retrieving information about many components of mixture simultaneously, ii) concentration assessment in the presence of interfering substances, iii) utilizing large set of sensors which have very similar chemosensing properties, iv) defining criteria of sensor selection for particular measurement tasks.

The cooperation with Analytische Chemie - Elektroanalytik & Sensorik resulted in six publications [10, 11, 13, 14, 17, 18], which are my most frequently cited scientific achievements. The postdoctoral fellowship in the NOVTECH project itself yielded two contributions [15, 16].

As the scientific achievement which is the basis for initializing the procedure of applying for habilitation under the Polish law I submit my monograph entitled *Data analysis in gas sensor measurements of air pollutants*, published by WUT Publishing House in 2012,.

From the perspective of my research activity in the domain of sensor data analysis, the monograph had two major objectives. These were: i) to systemize my concept of data analysis in air pollution measurements with partially selective gas sensors, ii) to demonstrate that the adequately profiled data analysis allows for retrieving wide spectrum of information about air pollution.

As the data source I chose the array composed of fifteen commercially available chemiresistors (TGS), which were operated in *stop flow* mode. The measurement data referred to volatile organic compounds. The choice of this group of air pollutants was intentional. It allowed for defining the pollution relevant information in a number of ways. I was interested in the traditional analytical information (identity and concentration of substances). However, I also considered other descriptors e.g. the category of pollutant, the concentration of organic carbon atoms. Such new ways of characterizing pollution are critical for the development of innovative sensor systems dedicated to environmental measurements. The methodological framework of the data analysis was defined by the *pattern* *recognition* approach. The literature on pattern recognition in gas sensor measurements and my own experience show that it is a very good strategy of data analysis in multisensor systems. In the monograph I analyzed the relationship between the composition of pattern recognition system and the efficiency of retrieving information about the pollution. Using my own concept of the feature, I studied various combinations of the methods to find best feature spaces together with classification or regression models. I tested a number of feature selection and extraction strategies. While searching for qualitative information I compared three different classifiers. Also the quantitative information was retrieved with several alternative regressions methods. My objective was to find a simple and effective solution for each kind of information. The results presented in the monograph demonstrate that the gas sensor array combined with an appropriate data analysis system, which utilizes pattern recognition approach, grants access to wide spectrum of information about air pollution. This coupling may become the best solution for the air pollution monitoring in the near future.

My favorite computational environment is Matalb. With this language I program the solutions of the majority of data analysis problems I face. Their predominantly complex character generates high memory and computational power requirements. Therefore I have an account in Wroclaw Center of Networking and Supercomputing (WCSS) and I use its resources on the regular basis.

Summarizing my scientific achievements after receiving the PhD degree; I am the co-author of 21 papers which were published in the international journals with *impact factor (IF)*. The total *IF* of these contributions is 57.924 (*IF* in the year of publication). I published one academic textbook in English and the earlier mentioned monograph. With my contribution there were written 3 book chapters, 4 publications in conference proceedings and 7 unpublished reports. The number of citations of these contributions is 72 (Web of Science). My Hirsch index is 6.

My scientific activity was appreciated by the University where I work. In 2008, 2009, 2010, 2011 and 2012 I was granted awards of Rector of Wroclaw University of Technology, for the distinctive contribution to the University achievements, with an emphasis on publications in renowned international scientific journals with high impact factor.

Monika Maciejewska

*) If not stated differently the literature reference sends to: Attachment 3 - **Papers in scientific journals with *impact factor*** *(after PhD)*