Multicomponent analysis on polluted waters by means of an electronic tongue

C. Di Natale [a,*], A. Macagnano [a], F. Davide [a], A. D'Amico [a], A. Legin [b], Y. Vlasov [b], A. Rudnitskaya [b], B. Selezenev [b]

[a] *Department of Electronic Engineering, University of Rome 'Tor Vergata', via della Ricerca Scientifica, 00133 Rome, Italy*
[b] *Department of Chemistry, University of St. Petersburg, St. Petersburg, Russia*

Abstract

In this paper the simultaneous measurements of the concentrations of a number of chemical species in solutions performed by a sensor array of ion sensitive electrodes are presented and discussed. By analogy with the well known electronic nose this sensor array operating in solutions, will be here called electronic tongue. In order to extract optimized information from the electronic tongue output data, many different techniques have been applied; they were based on chemometrics, non-linear least squares and neural networks. The best results have been achieved by the introduction of modular models which make use, at the same time, of both qualitative and quantitative information. © 1997 Elsevier Science S.A.

1. Introduction

Environmental monitoring requires on-site simultaneous measurements of a number of different chemical species. A classic approach to this problem is to consider a number of selective sensors one for each chemical species. Along with this in presence of complex solutions, sensors tend to lose their own specificity and their response is no more directly related to the concentration of the species for which they have been designed, and they become rather influenced by the presence of the other species. This process can be represented as a sort of convolution between the sensor selectivity (namely the sensitivity to the species present in the environment) and the chemical pattern occurring in the environment under measurement.

Multicomponent analysis is an analytical procedure allowing the extraction of qualitative and quantitative information from an array of non-selective sensors. It is based on the utilization of an array of sensors matched with a suitable data analysis procedure.

Multicomponent analysis provides a sensor array model from a calibration data-set, which should be large enough to cover the concentration range of each species, and to cope the non-linearity in the sensor responses.

Many data analysis techniques can be utilized to disentangle the information from sensor array outputs; they can be grouped in four classes: chemometrics, artificial neural networks, non-linear least squares and some other exotic methods (such as genetic algorithms or multi-dimensional splines).

In past years the possibility to perform measurements of concentrations of a number of chemical species in complex solution has been attempted [1–3]. By analogy with the natural olfaction, where sensor arrays operating in air which are labelled as electronic noses, sensor arrays operating in liquid media will be here called electronic tongues.

In this paper the measurement of quantities relevant for pollution monitoring of internal waters by means of an array of chemically sensitive electrodes will be presented and discussed. To this scope samples of waters from Neva river (a river flowing through the town St. Petersburg) has been artificially polluted with ionic metals in order to simulate generic industrial waste pollution.

---

* Corresponding author. Tel.: +39 6 7254408; fax: +39 6 2020519; e-mail: dinatale@eln.utovrm.it

Table 1
Ranges of the total concentrations of the eight chemical species in solution

Species	Concentration range ($-\log[C]$)
Cu	6–8
Cd	6–7
Zn	3–5
Cr	4–7
Fe	4–7
Cl	3–5.3
SO$_4$	3–3.9
H	2.6–8

Different data analysis methods have been utilized in order to select the most fitting one. Substantial improvement in accuracy has been obtained with modular models that make use, at the same time, of qualitative and quantitative information.

It is worth noting that regression methods generally do not make any use of qualitative information on the data. This means that an investigation of the data distribution may reveal, sometimes, the existence of different qualitative classes. In these cases it is straightforward to suppose that the development of a many regression model, one for each class, should perform better relative to a unique regression model holding over the whole concentration ranges. Data classification can be achieved in several ways, among the others in this paper principal component analysis and self organizing maps have been considered.

2. Experimental

To simulate real conditions samples of Neva river waters have been artificially polluted with ionic metals in order to reproduce generic industrial wastes. Neva river waters were taken at three different locations. In each sample the following elements were added in ionic form: Cu, Cd, Fe, Cr and Zn. The solutions were left at room temperature for one week to approach, as much as possible, an equilibrium state. The scope of the work was to measure the total concentrations of: Cu, Cd, Fe, Cr, Zn, Cl, SO$_4$ and H. It should be noted that no

addition of Cl, SO$_4$ and H was done and that therefore their concentrations actually occurred, in the Neva river, at the sampling locations. The range of concentrations, for each species, is shown in Table 1.

Samples have been measured by an electronic tongue formed by 22 electrodes mainly based on chalcogenide glasses variously doped and conventional electrodes.

The following chalcogenide glass systems have been used for the sensor preparation: AgI–Sb$_2$S$_3$, Ag$_2$S–As$_2$S$_3$, GeS–GeS$_2$–Ag$_2$S, Ge–Sb–Se–Ag with different components content. Also commercial copper-, lead- and cadmium-selective chalcogenide glass sensors were incorporated. Glasses were prepared in evacuated quartz ampoules at 1000 K from high purity components. 5–7 mm diameter and 3–5 mm thickness disks were cut and sealed into plastic tubes. Both liquid and solid-state inner contacts were used. The details of glass synthesis and sensor preparation are described in Ref. [4].

Commercial chalcogenide glass sensors were provided by Analytical Systems (St. Petersburg, Russia). Their compositions are summarized in Ref. [4]. Conventional crystalline sensor was used basically for chloride ion determination. Two sensors of every type have been used in the array.

In order to determine a robust regression model for the sensor array about 150 chemical solutions of different concentrations were prepared and measured.

All sensors, including those with a nominal ion selectivity, have shown a strong cross-selectivity that did not allow any direct measurement of concentrations.

3. Data analysis

In order to utilize the sensor array to retrieve an estimate of the concentrations of the relevant species in unknown samples a calibration of the sensor array is required. For this scope it is necessary to collect a calibration data-set of measurements performed at known environmental conditions and then carry out a data analysis procedure.

As mentioned in the previous paragraph about 150 calibration measurements were performed. The whole data-set has been split in two parts, one for calibration

Table 2
Mean relative absolute error (RAE) obtained analysing the data with various methods

	Cu (%)	Cd (%)	Zn (%)	Cr (%)	Fe (%)	Cl (%)	SO$_4$ (%)	H (%)
MLR	40	25	47	208	73	6	11	11
PLS	7	5	7	15	8	2	4	4
NLLS	9	4	7	13	6	3	3	6
BP–NN	7	5	7	15	7	1	1	4

It is worth to note that, apart MLR which is only a reference to evaluate the non-linearities, the other methods show about the same perfromances.
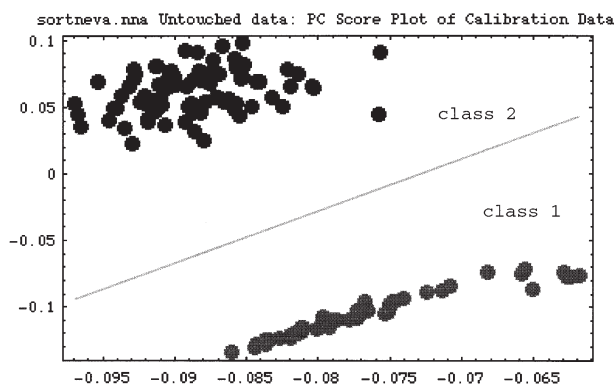
Fig. 1. PCA plot of concentration values in which the occurrence of two classes of data is observed.

(to determine the model) and one to test the capability of the model to predict correct concentrations from unknown samples.

Many different approaches to the data analysis have been utilized. An extensive comparison of the performances of many techniques have been accomplished in order to obtain the best possible estimation of the concentrations. The following methods were utilized: multiple linear regression (MLR), partial least squares (PLS), non-linear least squares (NLLS) and back-propagation–neural network (BP–NN).

MLR approach has been utilized to evaluate how far from linearity the sensor response was. Electrodes operating in single-component solutions behave according to the Nernst law, and cross-selectivities are taken into account for mixed solutions by the Nikolskij extension. In solutions characterized by the presence of many different compounds a deviation form the linearity is expected also if models for the sensor response are not currently available. So that the errors found by linear modelling give a figure for the entity of the deviation from linear behaviour.

PLS is a tool extensively utilized, for quantitative analysis, in chemometrics and in multisensors applications. Although it is based on a linear approach it achieves results which are substantially better than those obtained by MLR. Nevertheless the non-linearities involved can, sometimes, be so large that different approaches, non-linear in character, are required.

From an analytical point of view the non-linear approach is represented by the non-linear least squares. The sensor array is modelled as a system of non-linear equations where the number of equations is equal to the number of sensors and the variables are the concentrations.

Since no theoretical models are available each electrode has been fitted by a polynomial function of the third order.

$$f(c_1,...,c_8) = k_0 + \sum_{j=1}^{8} k_{1,j} c_i + k_{2,j} c_j^2 + k_{3,j} c_j^3 \qquad (1)$$

The calibration of the array consists then in the estimation of 25 parameters for each electrode for a total of 550 parameters. The whole array is then represented by a vectorial function describing the relation between each sensor output and the concentrations of the eight species. Since each electrode is represented by Eq. (1) the array function is written as:

$$\vec{F}(c_1,...,c_8) = k_0 + \begin{bmatrix} k_{1,1} & \cdots & k_1 \\ \cdots & & \cdots \\ k_{22,1} & \cdots & k_{22,25} \end{bmatrix} * \begin{bmatrix} 1 \\ c_1 \\ \cdots \\ c_8 \\ c_1^2 \\ \cdots \\ c_8^2 \\ c_1^3 \\ \cdots \\ c_8^3 \end{bmatrix}$$

$$(2)$$

The inversion of this equation allows to estimate, from the output of the sensors, the concentrations of interest.

These operations (parameters estimation and array function inversion) can be performed using the Levenberg–Marquardt algorithm which provides an iterative solution for the solution of redundant systems of non-linear equations. [5]

Non-linear least squares approach (NLLS) requires much effort in terms of calculus and furthermore the Levenberg–Marquardt algorithm can also have convergence problems which gives rise to a limitation of the accuracy of the estimations.

In recent years artificial neural networks have gained popularity to solve non-linear modelling problems. Their successes are mainly due to the fact that, from the user point of view, they can be utilized as a regression machine able to establish correlations between blocks of data.

Among the neural networks able to perform regression, back-propagation based networks are currently widely utilized. Basically these networks provide non-linear models whose parameters are optimized by an algorithm whose convergence is extremely favoured by the particular arrangement of the network in feed forward layers [6].

The network here utilized was a 22:15:8 feed forward network where the neuron transfer function was the hyperbolic tangent.

MLR, PLS and NLLS were implemented on Mathematica™ For the ANN the Professional II Software tool by NeuralWare™ was utilized. All software ran on Apple PowerMacintosh™ 7500. PLS was implemented following the kernel approach described in Ref [7]. The
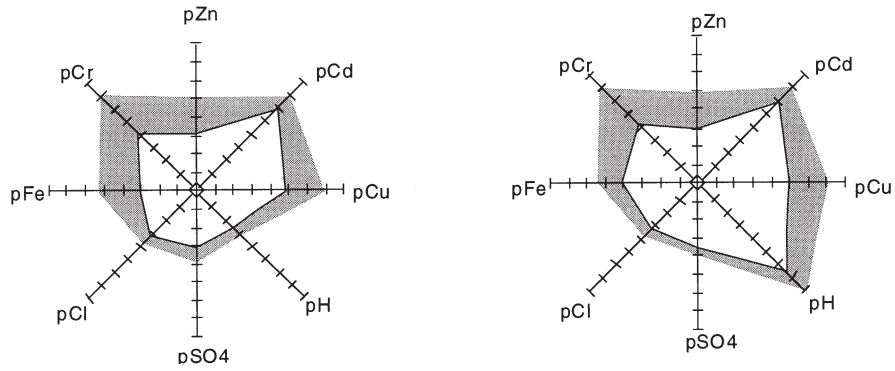
Fig. 2. Radar plot of the concentration ranges in the two classes. The main difference among the classes is represented by the value of pH.

performances of the three methods in predicting unknown concentrations have been evaluated through the percentage relative absolute error (RAE) in retrieving the concentration in the test data-set. Results are listed in Table 2.

Disregarding MLR which, as said before, gives only information on the entity of non-linearities involved in the array, the other three methods exhibit similar performances towards all the concentrations. It is remarkable the fact that although the results are comparable the effort, in terms of calculation and theoretical complications, is rather different. Indeed PLS can be considered as the less expensive technique, it can be implemented in any mathematical oriented programming language (such as Mathematica or Matlab) and the calculation time is rather fast also on medium sized personal computers. NLLS and BP–NN, on the other hand, require longer time of calculus and, due to their iterative nature, present convergence problems.

A substantial improvement of the data analysis performances can be obtained taking into consideration the qualitative character of the data. This can be easily performed by using PCA, a method widely utilized in chemometrics and then on electronic nose data analysis to display multidimensional data in a sub-space formed by the principal components, namely those directions along which the variance of the data is maximum.

Fig. 1 shows the PCA plot of the concentrations; the distribution of points reveals that the concentrations were not homogeneously distributed in the eight-fold space of concentrations but that they are rather clustered in two classes. The concentration range, inside each class, is shown as a radar plot in Fig. 2. The main difference among the two classes of concentrations lies in the pH value that, in one case, spans from 2.6 to 3.5 and, in the other class, from 6.5 to 8.

Fig. 3 shows the PCA plot of the sensors outputs. It is clear that, from a qualitative point of view the sensors are able to discriminate between the two classes. This circumstance suggests that it should be possible to obtain a better performing sensor array model including these qualitative information and taking into account one sub-model for each qualitative class.
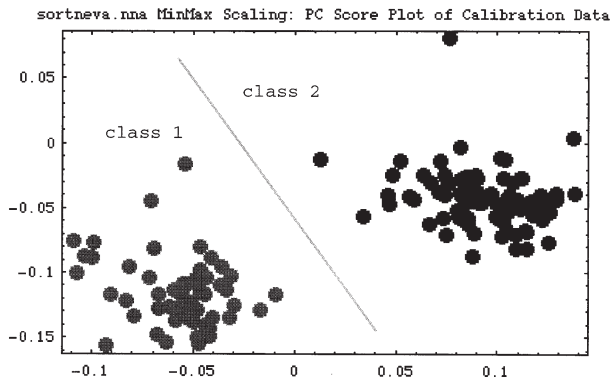

Fig. 3. PCA plot of the sensor outputs. The electronic tongue correctly classify the data in two classes.
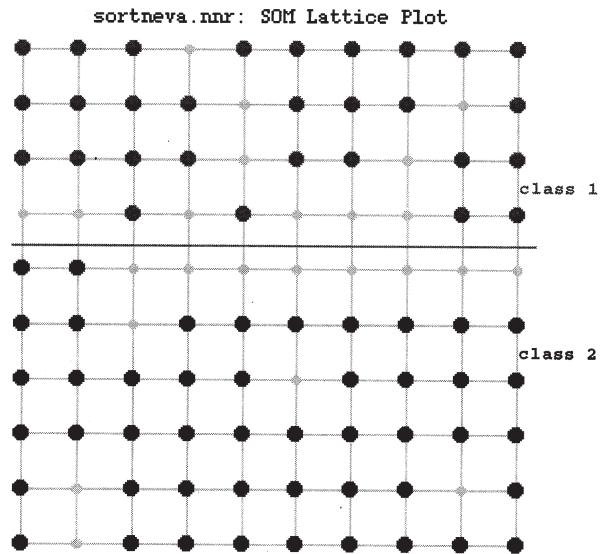

Fig. 4. SOM lattice plot of the sensor outputs. A net distinction of the two classes is observed.
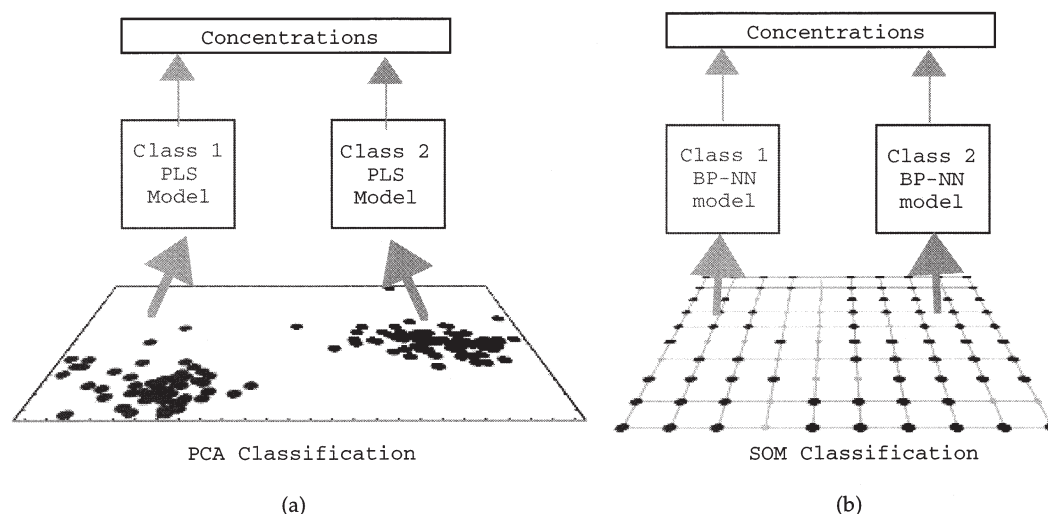
Fig. 5. Architecture of the two modular models. On the left side the chemometric based one is shown. It is formed by a PCA layer as classifier and on two PLS block one for each classes. Depending on the side of the PCA plot plane on which a data falls one of the two block is activated. On the right side the neural network based model is represented. In this case the classification is accomplished by a SOM while two BP–NN produce the estimation of the concentrations.

Two modular models have been considered and compared. The first was based on chemometrics blocks, composed by one PCA and two PLS blocks, and the second was based on neural networks, using a Self Organizing Map (SOM) and two BP–NN [8].

In the neural networks model a SOM is utilized as a classifier. Beside its many features [9,10] SOM can be considered as a data modelling tool which gives a bidimensional representation. It has been demonstrated that SOM is a neural implementation of a sort of principal curve analysis. Fig. 4 shows the SOM lattice plot of the sensor outputs; as in PCA, also in this case the two classes are clearly separated.

Fig. 5 shows the architecture of the two modular models, while their results, expressed by the mean RAE, are listed in Table 3.

Modular models strongly reduces the error in the concentration estimates in respect to the conforming models operating without the qualitative information. Furthermore the modular model formed by two kinds of neural networks behaves better than that based on chemometrics blocks.

4. Conclusions

The electronic tongue, developed in the St. Petersburg–Rome collaboration, has been shown to be an useful tool for simultaneous measurement of several species in environmental applications, such as the waters of a river. The accuracy of the measurement can be improved choosing the proper data analysis technique. A comparison among several methods, based on conceptually different approach, shows that the performances obtained by PLS, NLLS and BP–NN are basically similar.

A decisive improvement of the predictions is obtained enlarging the amount of considered information taking into account also the qualitative aspects of the data. To this regard two different modular models based on chemometrics (PCA + PLS) and neural networks (SOM + BP–NN) have been introduced. It has been proved that these models significantly improve the accuracy of the estimations in respect to models operating without qualitative information.

Table 3
Mean RAE obtained by modular models

	Cu (%)	Cd (%)	Zn (%)	Cr (%)	Fe (%)	Cl (%)	SO$_4$ (%)	H (%)
PCA + PLS	4	2	3	6	4	1	2	2
SOM + BP–NN	2	<1	1	3	1	<1	<1	1

A factor two of gain in respect to the value listed in table 2 is obtained for (PCA + PLS) model, while more than a factor four is reached by the (SOM + GP–NN) model.

References

[1] C. Di Natale, F. Davide, A. D'Amico, A. Legin, Y. Vlasov, Multicomponent analysis of heavy metal cations and inorganic anions in liquids by a non-selective chalcogenide glass sensor array, Sensors and Actuators B 34 (1996) 539–542.

[2] M. Otto, J.D.R. Thomas, Model studies on multiple channel analysis of free magnesium, calcium, sodium and potassium at physiological concentration levels with ion-selective electrode, Anal. Chem. 57 (1985) 2647.

[3] M. Bos, A. Bos, W.E. Van der Linden, Processing of signals from ion selective electrode array by a neural network, Anal. Chem. 233 (1990) 31.

[4] Yu.G. Vlasov, E.A. Bychkov, Chalcogenide glass ion-selective electrodes, Ion Selective Electrode Rev. 9 (1987) 5.

[5] C. Di Natale, F. Davide, in: A. D'Amico and G. Sberveglieri (Eds.), Proceedings of the 1st European School on Sensors, World Scientific, Singapore, 1995.

[6] D.E. Rumelhart, J.L. McClelland (Eds.), Parallel Distributed Processing: Explorations in the Microstructures of Cognition, vol. 1, Foundations, MIT Press, Cambridge MA, 1986.

[7] F. Lindgren, P. Geladi, S. Wold, A kernel algorithm for PLS, J. Chemometrics 7 (1993) 45.

[8] C. Di Natale, F. Davide, A. D'Amico, W. Göpel, U. Weimar, Sensor arrays calibration with enhanced neural networks, Sensors and Actuators B 19 (1994) 654–657.

[9] T. Kohonen, Self Organizing Map, Springer-Verlag, Berlin, 1995.

[10] F. Davide, C. Di Natale, A. D'Amico, Self organizing multisensor systems for odour classification: internal categorization, adaptation and drift rejection, Sensors and Actuators B 18 (1994) 244–258.