

DATA ANALYSIS FOR CHEMICAL SENSOR ARRAYS

Corrado Di Natale, Eugenio Martinelli, Giorgio Pennazza, Andrea Orsini,
Marco Santonico

University of Roma "Tor Vergata"

Dept. of Electronic Engineer, Via del Politecnico, 1 00133 Roma

dinatale@uniroma2.it, damico@eln.uniroma2.it, andrea.orsini@psm.rm.cnr.it

Abstract

Arrays were introduced in the mid-eighties as a method to counteract the cross-selectivity of gas sensors. Their use has since become a common practice in sensor applications. [1]. The great advantage of this technique is that once arrays are matched with proper multivariate data analysis, the use of non-selective sensors for practical applications becomes possible. Again in the eighties, Persaud and Dodds argued that such arrays has a very close connection with mammalian olfaction systems. This conjecture opened the way to the advent of electronic noses [2], a popular name for chemical sensor arrays used for qualitative analysis of complex samples.

It is worth remarking that a gas sensor array is a mere mathematical construction where the sensor outputs are arranged as components of a vector. Arrays can also be utilized to investigate the properties of chemical sensors, or even better, the peculiar behaviour of a sensor as a component of an array. In this chapter, the more common sensor array methodologies are critically reviewed, including the most general steps of a multivariate data analysis. The application of such methods to the study of sensor properties is also illustrated through a practical example.

Keywords:

electronic nose; principal component analysis; pattern recognition; chemical sensors; sensor arrays; olfaction system; multivariate data analysis.

1. Feature extraction

In pattern recognition, a "feature" is any direct or derived measurement of the entities to be classified that helps differentiate between

classes. In chemical sensor arrays individual measurements are the entities assigned to classes while a measurement is a time sequence of sensor signals collected during the exposure of the sensor to the sample. Consequently, feature extraction for chemical sensors is the evaluation, from a sensor signal stream, of a number of parameters that can, as much as possible, represent the sensor experience containing that information related to the classification objective.

Feature extraction is of fundamental importance because sensor features are utilized in any successive elaboration to produce the output of the sensor system in terms of estimation of the measured quantities.

To define a feature extraction procedure it is necessary to consider that the output signal of a chemical sensor follows the variation of the concentration of gases at which it is exposed with a certain dynamics. The nontrivial handling of gas samples complicates the investigation of the dynamics of the sensor response. Generally, sensor response models based on the assumption of a very rapid concentration transition from two steady states results in exponential behaviour.

A straightforward solution of the feature extraction problem disregards the dynamic transitions and considers only the signal shift between two stationary states before the application of gas stimuli and during the exposure to gas after the transitory phase. This quantity has a straightforward meaning related to the equilibrium conditions established between the molecules in gas phase and those interacting with the sensor. Although the direct chemical and physical meaning of the steady state signal shifts, it is worth investigating if the dynamic properties can provide features with an extended information content.

Simple evidence about the information content of the dynamic properties can be obtained by considering the Langmuir model of adsorption of molecules from gas phase to a limited number of interacting sites on a sensor surface [3].

According to this model, the total rate of adsorbed molecules (dn/dt) is given by the algebraic sum of the two processes occurring at the interaction sites: adsorption and desorption. The equilibrium condition is reached once one equalizes the other.

In the case of N_s adsorbing sites, if P is the gas partial pressure held constant by some gas reservoir, and k_a and k_d are adsorption and desorption constants respectively, the following expression for adsorption and desorption rates holds:

$$\frac{dn}{dt} = \left(\frac{dn}{dt}\right)_{ads} - \left(\frac{dn}{dt}\right)_{des} = P \cdot k_a \cdot (N_s - n) - k_d \cdot n. \quad (1)$$

Solution of this equation gives rise to the time behaviour of the number of adsorbed molecules. For most transducers, this quantity is then linearly converted into a sensor signal.

$$n(t) = \frac{k_a NP}{k_d + k_a P} \left(1 - e^{-(k_d + k_a P)t} \right). \quad (2)$$

The number of adsorbed molecules evolves exponentially towards a steady value. In figure 1 the qualitative behaviour of the steady state amount of adsorbed molecules and the time constant are plotted versus the gas concentration. It is evident that the feature representing the steady state signal reaches a saturation value as the number of adsorbed molecules equals the number of adsorbing sites while the dynamics represented by the time constant or, even better, its inverse, is linear with the gas concentration. In practice, when the concentration rapidly changes from zero to a value exceeding the saturation limit, the dynamics of the sensor continues to provide a linear response. From the point of view of sensitivity, namely the derivative of the feature versus the concentration, the steady-state feature results in a progressively decreasing sensitivity, while, in the limits of the model, the dynamic property provides a constant sensitivity over a theoretically unlimited range.

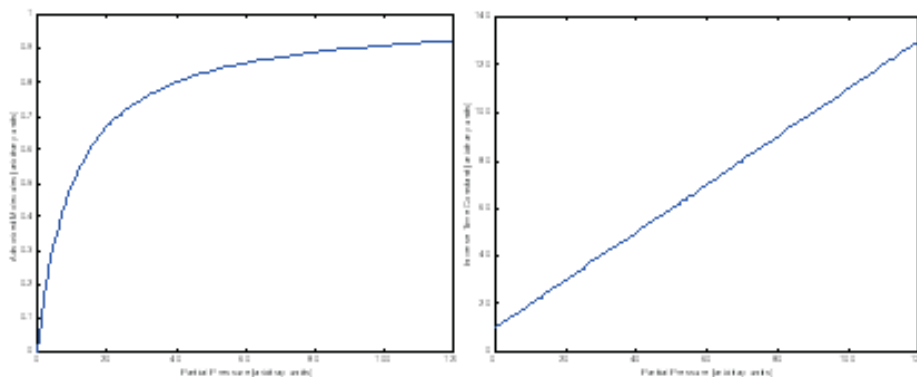


Figure 1. Qualitative behaviour with the gas concentration of the steady-state signal shift (a) and time constant inverse (b) for a pure Langmuir adsorption on a limited amount of equi-energetic adsorption sites over a sensor surface.

Some attempts to exploit sensor dynamics for concentration prediction were carried out in the past. Davide et al. approached the problem using dynamic system theory, applying non-linear Volterra series to the modelling of Thickness Shear Mode Resonator (TSMR) sensors [4]. This approach gave rise to non-linear models where the difficulty to discrimi-

nate the intrinsic sensor properties from those of the gas delivery systems limited the efficiency of the approach.

Other authors concentrated their attention on the evaluation of the time constants. It is well known that multiexponential fit is an ill-posed problem, and its solution, obtained with typical non-linear optimization methods (e.g. the Levenberg-Marquardt algorithm), may be affected by large errors. Di Natale et al [5] applied an alternative method based on multiexponential spectroscopy, originally developed by Samitier et. al. [6], to the determination of the time constants in a tin-oxide sensor array. An opposite approach to feature extraction considered completely empirical methods. Eklöv et al [7] investigated a number of parameters taking into account several descriptors of the behaviour of CHEMFET sensor signals. Once a feature is chosen, the signals may be transformed into data, which can be analyzed to find the correlation between sensor response and relevant sample characteristics.

2. Data Pre-processing: Scaling

In data analysis, data are seldom used without some preprocessing. Such preprocessing is typically concerned with the scale of data. In this regard two main scaling procedures are widely used: zero-centered and autoscaling.

Zero-centered data means that each sensor is shifted across the zero value, so that the mean of the responses is zero. Zero-centered scaling may be important when the assumption of a known statistical distribution of the data is used. For instance, in case of a normal distribution, zero-centered data are completely described only by the covariance matrix.

Autoscaling means to scale each sensor to zero-mean and unitary-variance. This operation equalizes the dynamics of all sensor responses, preventing a sensor having a larger response range from hiding the contribution of other dynamically limited sensors. Further, autoscaling makes the sensor responses dimension-less. This feature becomes necessary when sensors whose signals are expressed in different units are merged in one array. This is the case for hybrid arrays (different sensor technologies in the same array) and when chemical sensors are fused with other instruments, e.g. the fusion of electronic noses and electronic tongues [11].

3. Normalization

Horner and Hierold [12] showed that the application of a simple normalization of sensor data can greatly help in preventing quantitative information from masking qualitative aspects of the data.

Figures 2 and 3 show the situation in the case of an array of TSMR sensors exposed to different compounds at various concentration levels [13]. The cross-selectivity of the sensors makes their individual responses ambiguous. Namely different samples, due to a combination of qualitative and quantitative aspects, may give rise to similar sensor responses. In figure 2 the confusion among the data shows the ambiguity of the sensor response. Figure 3 shows a Principal Component Analysis (PCA) plot. PCA will be thoroughly discussed in a later section. It is a typical method to represent multivariate data in a bidimensional plot.

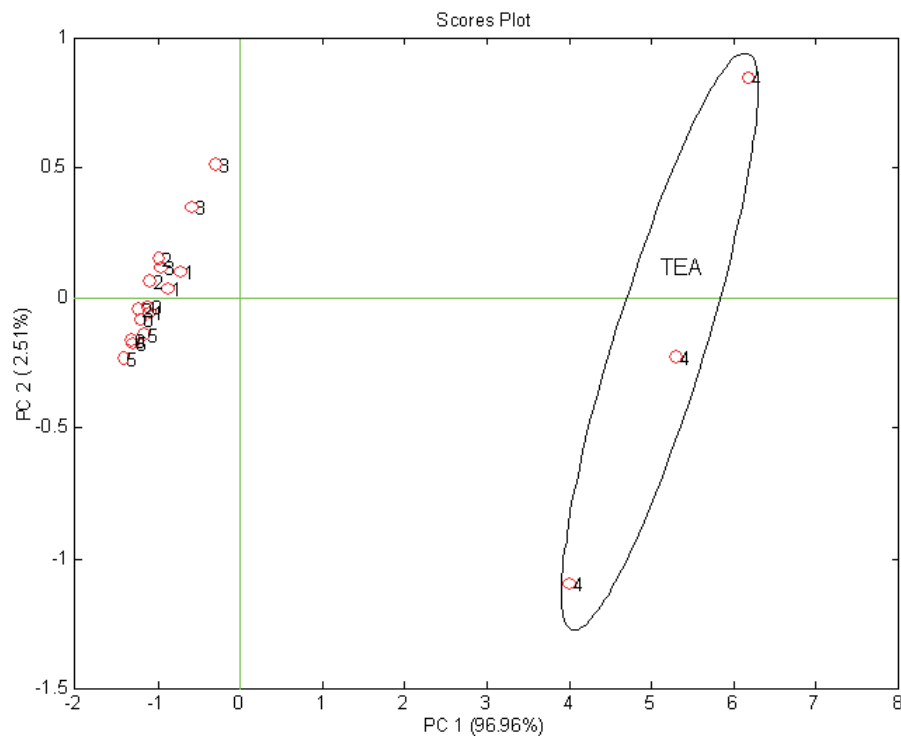


Figure 2. Examples of data characterized by strong concentration effects. Data are related to a quartz microbalance array exposed to six volatile compounds each measured three times at different concentrations. Only triethylamine (TEA in the plot) data emerge (experimental details in ref. [10]).

A simple way to disentangle the information is obtained when the relationship between sensor responses and concentrations of analytes is linear, such as found in polymer coated TSMR:

$$\Delta f_i = K_{ij} \cdot c_j \quad (3)$$

where Δf_i is the response of the i -th TSMR, K_{ij} is the sensitivity of the i -th sensor towards the j -th compound, and c_j is the concentration of the j -th compound.

The normalization consists of dividing each sensor response by the sum of all the sensor responses to the same sample, so that the concentration information may disappear.

$$\Delta f_i \Rightarrow \frac{K_{ij} \cdot c_j}{\sum_m K_{mj} \cdot c_j} = \frac{K_{ij}}{\sum_m K_{mj}} \quad (4)$$

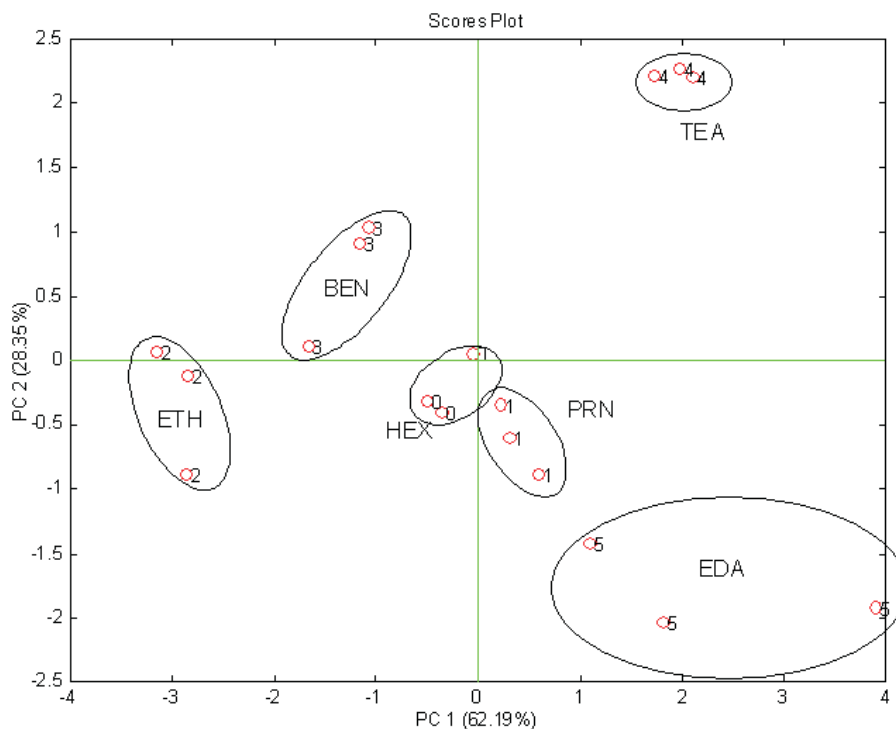


Figure 3. Data of figure 2 after the application of linear normalization of equation 4. Classes are now clearly separated. It is worth noting the condensation of TEA data (class 4) in a very restricted region. Meaning of the classes: HEX: hexane; PRN: propanal; ETH: ethanol; BEN: benzene; TEA: triethylamine; EDA: ethylenediamine.

Figure 3 shows the PCA score plot of the same data of figure 2 after the application of equation 4. The application of linear normalization to an array of linear sensors should produce, on the PCA score plot, one point for each compound, independent of its concentration, and achieve the highest possible recognition. Deviations from ideal behaviour, as shown in figure 3, are due to the presence of measurement errors, and to the non-linear relationship between sensor response and concentration.

In the previously quoted paper, Horner and Hierold also treated the case of sensors whose response is ruled by a power law relationship between sensor response and analyte concentration. This is the case for metal-oxide semiconductor gas sensors. Eq.4 can be extended to sensors described by a power-law ($z = c^\alpha$) simply linearizing, through the logarithm, the sensor response.

Normalization is, in practice, also useful to counteract any possible fluctuations in the sample concentration. These fluctuations are, in practice, mostly due to sample temperature fluctuations, and to instabilities of the sampling system and they may lead to variations of the dilution factor of the sample with the carrier gas. Of course, normalization is of limited efficiency because the mentioned assumptions strictly hold for simple gases and they fail when mixtures of compounds are measured. Furthermore, it has to be considered that in complex mixtures, temperature fluctuations do not result in a general concentration shift, but since individual compounds have different boiling temperatures, each component of a mixture changes differently so that both quantitative (concentration shift) and qualitative (pattern distortion) variations take place.

4. Multivariate data exploration

Given a set of data related to a number of measurements, after the application of proper feature extraction, pre-processing and normalization, exploratory techniques aim at studying the intrinsic characteristics of the data in order to discover eventual internal properties.

Exploratory data analysis shows the aptitude of an ensemble of chemical sensors to be utilized for a given application, leaving to the supervised classification the task of building a model to be used to predict the class membership of unknown samples.

Two main groups of exploratory analysis may be identified: representation techniques and clustering techniques.

Representation techniques are a group of algorithms aimed at providing a representation of the data in a space of dimension lower than that of the original sensor space. The most popular of these methods

are Principal Component Analysis, Self Organizing Map, and Sammon's mapping. Each of these techniques is based on specific hypotheses about the nature of the data and the sensor space. Each of them tends to preserve some particular characteristic of the data. It is worth remarking that the simplest, in terms of calculus and interpretation of results, is that based on the strongest assumption about the statistical distribution of the data. On the other hand, when assumptions about data distribution are removed a neural network is necessary for data representation.

Clustering techniques are mostly based on the concept of similarity expressed through the definition of a metric (distances calculus rule) in the sensor space. The most trivial and common choice is to express the similarity as a Euclidean distance. Other definitions, such as the Mahalanobis distance, are also used [14].

5. Principal Component Analysis

The scope of Principal Component Analysis (PCA) is a *consistent* portrayal of a data set in a representation space. Mathematically, PCA is a linear transformation that may be described as $S=WX$. Here X is the original data set, W is the transformation matrix, and S are the data in the representation space. PCA is the simplest and most widely used method of multivariate analysis. Nonetheless, most users are seldom aware of its assumptions and sometimes results are badly interpreted.

The peculiarity of PCA is in a representation of the data set onto a subspace of reduced dimensionality where the statistical properties of the original data set are preserved.

Although the PCA concept is used in many disciplines it was strongly developed in chemometrics, where it was introduced at the beginning to analyze spectroscopic and chromatographic data, which are characterized by a higher correlation among the spectra channels [8].

The possibility of a reliable representation of a chemical sensor array data set in subspaces of smaller dimension lies in the fact that the individual sensors always exhibit a high correlation among themselves. PCA consists of finding an orthogonal basis where the correlation among sensors disappears.

As a consequence, in a sensor space of dimension N the effective dimension of the sub-space occupied by the data is less than N . This dimension can be precisely evaluated using algorithms developed to describe dynamic systems. An example is the correlation distance that allows evaluating the fractional dimensionality of a data-set [16]. Correlation distance provides an independent way to evaluate the expected reduction of dimension.

Statistical properties of a data set can be preserved only if the statistical distribution of the data is assumed. PCA assumes the multivariate data are described by a Gaussian distribution, and then PCA is calculated considering only the second moment of the probability distribution of the data (covariance matrix). Indeed, for normally distributed data the covariance matrix ($X^T X$) completely describes the data, once they are zero-centered. From a geometric point of view, any covariance matrix, since it is a symmetric matrix, is associated with a hyper-ellipsoid in N dimensional space. PCA corresponds to a coordinate rotation from the natural sensor space axis to a novel axis basis formed by the principal direction of the hyper-ellipse associated with the covariance matrix. The reduction of an ellipsoid to its canonical form is a typical linear algebra operation performed by calculating the eigenvectors of the associated matrix.

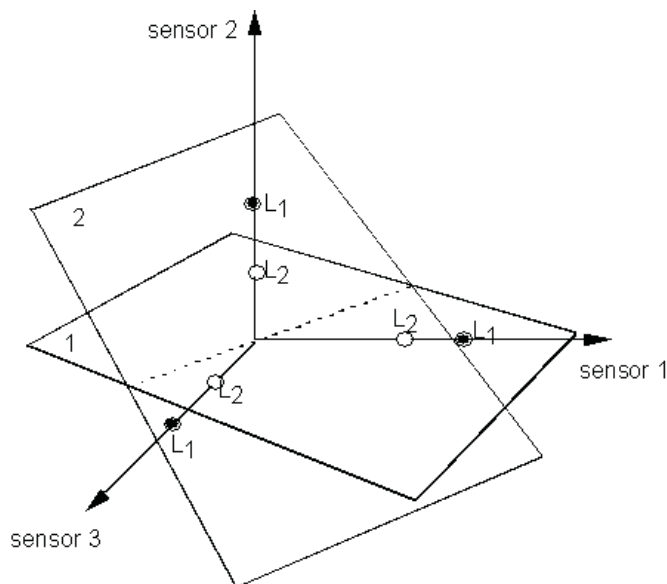


Figure 4. Sketch of the geometrical meaning of loading based feature extraction. Each measurement is represented by the loadings of a PCA decomposition limited to the most meaningful components. In the picture the case of three sensors is displayed limiting the feature to the first two principal components. Each measurement determines a plane. In terms of pattern recognition, the study of similarity and difference among measurements is translated into differently oriented planes.

In practice, PCA can be calculated with the following rule. Let us consider a matrix X of data, let $C=X^T X$ be the covariance matrix of X . The i -th principal component of X is $X^T \lambda(i)$, where $\lambda(i)$ is the i -th

normalized eigenvector of C corresponding to the i -th largest eigenvalue. A sketch of the geometric meaning is shown in figure 4.

Since the eigenvalues of the matrix associated with the quadratic form describing a hyper-ellipse are proportional to the extension of the solid in the direction of the corresponding eigenvector, the eigenvalues of the covariance matrix are directly proportional to the variance along the corresponding eigenvector. Therefore considering the relative values of the eigenvalues $\lambda(i)$ it is possible to reduce the representation to only those components carrying most of the information.

Given a matrix of data PCA results in two quantities usually called scores and loadings. Scores are related to the measurements, and they are defined as the coordinates of each vector measurement (a row of matrix X) in the principal components basis.

The loadings describe the contribution of each sensor to the principal components basis. A large loading, for a sensor, means that the principal component is mostly aligned along the sensor direction.

It is important to note that the highest eigenvalues correspond to components defining the directions of highest correlation among the sensors, while the components characterized by smaller eigenvalues are related to uncorrelated directions. Since sensor noises are uncorrelated the representation of the data using only the most meaningful components removes the noise of the sensors. In this way PCA is used to remove noise from spectral data [15].

When applied to electronic nose data the presence of various sources of correlated disturbances has to be considered. As an example, sample temperature fluctuations induce correlated disturbances, which may be described by principal components of highest order. When these disturbances are important the first principal component has to be eliminated in order to emphasize the relevant data properties. A set of algorithms called Minor Component Analysis (MCA) was introduced to take into account these phenomena mainly in image analysis [17].

The hypothesis of a normal distribution is a strong limitation that should be always kept in mind when PCA is used. In electronic nose experiments, samples are usually extracted from more than one class, and it is not always that the totality of measurements results in a normally distributed data set. Nonetheless, PCA is frequently used to analyze electronic nose data. Due to the high correlation normally shown by electronic nose sensors, PCA allows a visual display of electronic nose data in either 2D or 3D plots. Higher order methods were proposed and studied to solve pattern recognition problems in other application fields. It is worth mentioning here the Independent Component Analysis (ICA) that has been applied successfully in image and sound analysis

problems [18]. Recently ICA was also applied to process electronic nose data results as a powerful pre-processor of data [19].

Nonetheless, a sub-set belonging to one class may very likely be normally distributed. In this case a PCA calculated on one class cannot work in describing data belonging to another class. In this way, the membership of data to each class can be evaluated. This aspect is used by a classification method called SIMCA (Soft Independent Modelling of Class Analogy). It is a clever exploitation of the limitations of PCA to build a classification methodology [20].

Non linear PCA algorithms have also been developed to provide a representation along principle curves rather than principal directions [21]. Also neural networks were proposed to solve the problem of faithful representation of multidimensional data in representation spaces of lower dimensions [17].

Another limitation to the use of PCA comes from the fact that being a linear projection it may introduce mistakes. Indeed, in the projection, data separated in the original space may result in a similar score, a phenomenon like that producing a constellation in the starred sky.

A final consideration about PCA is concerned with its use as a preprocessor of non-linear methods such as neural networks [22]. The assumption of a normal distribution of the data requires all following analysis steps to adhere to this hypothesis. If positive results are sometimes achieved they have to be considered as serendipitous events.

6. Supervised Classification

Exploration analysis is not adequate when the task of the analysis is clearly defined. An example is the attribution of each measurement to a pre-defined set of classes. In these cases it is necessary to find a sort of regression able to assign each measurement to a class according to some pre-defined criteria of class membership selection. This kind of analysis is called supervised classification. The information about which classes are present have to be acquired from other considerations about the application under study. Once classes are defined, supervised classification may be described as the search for a model of the following kind:

$$\vec{c} = f(\vec{s}) \quad (5)$$

where \vec{c} is a vector describing the class assignment, \vec{s} is the vector of features of the sensors in the array and f is a generic function. This kind of problem is generally called pattern recognition. To solve pattern recognition problems using conventional algebra, class memberships must be encoded in a numerical form that allows treating the problem by numer-

ical methods. The most common way to represent class memberships is the so-called “one-of-many” code. In this codification, the dimension of \vec{c} is equal to the number of classes. The component corresponding to the class to which a sample belongs is assigned to 1, leaving to 0 the others. In chemical sensor array applications various sources of measurement errors may occur. As a consequence, equation 5 is written in a more realistic form as:

$$\vec{c} = f(\vec{s}) + \vec{e} \quad (6)$$

where the vector \vec{e} contains anything not related to the classification scheme expressed by the vector \vec{c} . Equation 6 is formally similar to the general problem of regression where the scope is the determination of the function f , in terms of functional form and parameters. Statistics provides the tools to estimate, from an experimental data set, the parameters of the function f , in order to approximate the measured experimental data. The classical approach is the Least Squares Method. It is important to reflect about the validity of least squares in chemical sensor array data. This method is based on the assumption that variables are normally distributed and that the quantity \vec{e} of equation 6 is a normally distributed zero-mean variable. The assumption of zero-mean means that all the variables to which the sensor responses are sensitive, except those related to the classification of the samples, may fluctuate but do not bias the measurement. This may not hold for sensor array data where, except for sensor noise, the contributions to \vec{e} (sensor drift, sample temperature variations, sample dilution changes, . . .) are not zero-mean quantities. Nevertheless, solutions based on least squares can be used to establish classification models, but it is important to be aware of the fundamental limitations of the methods. Before discussing practical solutions, it is necessary to detail the general framework. In pattern classification it is very important to estimate the expected error rate after the classifier model has been assessed on a calibration data set. The expected error rate determines the efficiency of the model by giving the probability of misclassifying future samples. Error rate estimation on the same set used for calibration induces large, optimistically biased estimates of performance. This effect is called “over-fitting”. The importance of over-fitting grows with the order of the regression function. In particular, it is important when highly non-linear functions are used (e.g. in neural networks). In these cases the model, while almost perfect when estimating the data on which the model is trained, fails completely when generalizing other data. Over-fitting may be more important for classification scopes, where the samples are extracted from sets that may not be well defined, so that the assignment of samples to classes may be affected by errors due to the vagueness of the clas-

sification scheme. In this situation the possibility of generating models which are not able to predict unknown samples with sufficient accuracy is high. The straightforward solution for error rate estimation is to split the data set into two independent sets, and use one for calibration and the other to test the classifier and estimate the error rate. This method cannot be used when available sample sizes are small. Moreover, how to split samples is a non-trivial problem because the division should be done while keeping the distributions of the two sets as close as possible in order to avoid biasing performance evaluation. A more reliable model validation is achieved using the “leave-one-out” technique [34]. “Leave-one-out” repeats the model building n times for n measures, each time leaving one measure out for testing and using the rest for training. The average test error rate over n trials is the estimated error rate. In case of small data sets, the bootstrap method has been proven to be more efficient than “leave-one-out”. Both “leave-one-out” and bootstrap are kinds of re-sampling methods. The bootstrap method generates new samples (called “bootstrap samples”) by drawing, with replacement, a number N of samples from the original samples [35]. Different methods to generate the bootstrap samples are available. A comparison of the four most popular is discussed in [36], where the efficiency of the methods is compared on a classification problem.

7. Linear Discrimination

The simplest way to estimate a supervised model is to consider that the descriptor of each class may be represented as a linear combination of the sensor responses. Considering N sensors and M classes the expressions can be written as:

$$\{c_1 = \sum_{j=1}^N k_{1j} \cdot s_j + e \quad \cdots \quad c_M = \sum_{j=1}^N k_{Mj} \cdot s_j + e\}. \quad (7)$$

Geometrically, this means sectioning the sensor space with straight lines, each bisecting the space. The result is a partitioning of the space into volumes, each defining one class. Considering a set P of experimental data, the previous set of equations can be written in a compact matrix form as:

$$c_{M \times P} = k_{M \times N} \cdot s_{N \times P} + E_{N \times P}. \quad (8)$$

The matrix $k_{M \times N}$ containing the model parameters can then be directly estimated using the Gauss-Markov theorem to find the least squares solution of generic linear problems written in matrix form [37]:

$$K_{M \times N} = c_{M \times P} \cdot s_{P \times N}^+. \quad (9)$$

Here the matrix $s_{P \times N}^+$ is the generalized inverse, or pseudo-inverse, of the matrix $s_{N \times P}$. The operation of pseudo-inversion generalizes the inversion of square matrices to rectangular matrices. This solution is often called Multiple Linear Regression (MLR). Once the model is assessed, it allows assigning any unknown samples to one class. Due to the presence of the above mentioned error matrix ($E_{N \times P}$), the model provides a numerical estimation of the “one-of-many” encoding of class assignment. In practice, something of different from 0 and 1 is obtained.

The estimated class assignment vector is called “classification score” and the sample is assigned to the class represented by the component with the higher value. This gives the possibility of evaluating a sort of goodness of the classification by considering either the ratio between the first and second values of the components of the estimated \vec{c} or the difference between the highest value and 1 (target value). The components of the matrix $K_{M \times N}$ define the importance of each sensor in the classification of each class. This information can also be used, as the loadings of PCA, to design and optimize the sensor array composition. The pseudo-inversion, like the inversion of square matrices, is influenced by the partial correlation among the sensors. Chemometrics offers methods to solve problems with colinear sensors, such as Principal Component Regression (PCR) and Partial Least Squares (PLS) [15].

Although designed for quantitative analysis, PLS may be applied to solve classification problems. In this case, PLS offers not only a more robust solution of the classification problem, but by plotting the latent variables it is possible to graphically represent the class separation. It is worth mentioning that the PLS latent variables are strictly related to but different from the PCA factors. Geometrically, PCA components are rotated in order to maximize their correlation with the components of the matrix of classes. In linear discrimination only classes separable by straight lines may be correctly classified. Classification improves if non-linear boundaries between classes are used. Figure 5 shows an example where a parabolic function achieves the separation where a linear boundary fails. A simple non-linear discriminant analysis can be obtained by a simple modification of the method previously discussed. As an example, let us consider the following quadratic form:

$$\{c_1 = \sum_{j=1}^N k_{1j} \cdot s_j + \sum_{j=1}^N h_{1j} \cdot s_j^2 + e \cdots c_M = \sum_{j=1}^N k_{Mj} \cdot s_j + \sum_{j=1}^N h_{Mj} \cdot s_j^2 + e\}. \quad (10)$$

This system may be written in the compact form of equation 5 by defining a suitable sensor matrix. Considering p measures, let us define a matrix T from the sensor responses, by:

$$T_{P \times 2N} = \begin{array}{cccc} \lrcorner & s_{11} & s_{1N} & s_{11}^2 & s_{1N}^2 & \llcorner \\ & \vdots & & \ddots & & \\ \llcorner & s_{p1} & s_{pN} & s_{p1}^2 & s_{pN}^2 & \lrcorner \end{array} \quad (11)$$

In the same way, the parameters k_{ij} and h_{ij} are joined to form a unique parameter matrix H . With these definitions a linear problem may be written like that of equation 5. The matrix H can then be estimated either by direct pseudo-inversion or by PLS. It is worth noting that increasing the order of the function increases the colinearity of the sensor matrix and the use of a chemometrics methodology (e.g. PLS) becomes more advantageous. Increasing the order of the discriminant function may solve highly complex classes distribution. The extreme solution is to use a method where the choice of the function is not required. Neural networks offer the possibility of solving the classification problem disregarding the functional form. It is well known that optimized neural networks may reproduce any kind of non-linear function. Neural networks derive all the knowledge from the experimental data, so that increasing the size of the calibration data-set increases the accuracy of the neural network based classifier.

8. Application to the investigation of Chemical Sensors properties

As an example of the use of array methodology to study chemical sensor properties let us consider the thirteen molecular structures reported in Figure 5. To investigate the sensing properties of these molecules we studied the behaviour of the response of thickness shear mode resonators (TSMR) sensors, each coated with a molecular film, to different concentration of various volatile compounds (VOC). Analyte compounds were chosen in order to have different expected interaction mechanisms. The sensors were exposed to the following VOC: n-pentane, methanol, benzene, triethylamine and acetic acid.

Data have been analyzed from a multivariate point of view. In this way the cooperative effects of the different materials is studied and the characteristics of each sensor are easily compared with those of the other sensors. PLS was used as a regression method for calculating the capability of the set of sensors to discriminate between the volatile compounds. Volatile compounds were checked at different concentrations in order to

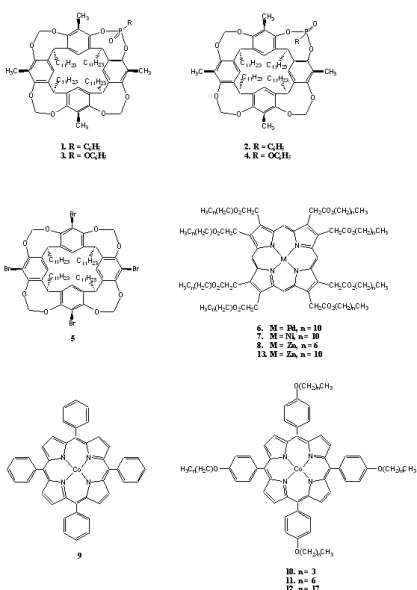


Figure 5.

evaluate the response of sensors in a wide concentration range. Nevertheless, the concentration variation tends to shadow the reaction of sensors with analytes, since the sensor response contains both qualitative (sensor analyte interaction) and quantitative (analyte concentration) information. In order to remove the quantitative information, data have been normalized using the linear normalization discussed in section 3.

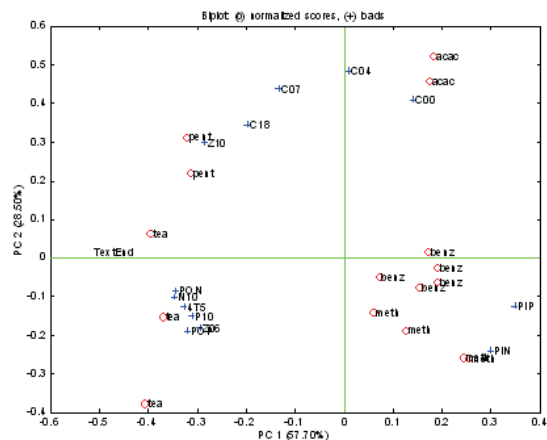


Figure 6.

Figure 6 shows the score plot of the first two latent variables of a PLS model aimed at discriminating the five substances. PLS is a data analysis method where the sensor signals are decomposed in latent variables obtained as linear combinations of the original variables, namely the sensors. The latent variables are chosen with the aim of maximizing the correlation between them and the scope of the regression. In the case treated here, the regression scope was the classification of the five substances. The original matrix decomposition gives place to the scores, the coordinates of the original sensors signals in the latent variable basis, and the loadings, the coordinates of the original variable axis (the sensors) in the new latent variable basis. These quantities can be plotted together in a so-called bi-plot. In a bi-plot the correlation between sensors and data points can be investigated and some conclusions about the importance of some sensors in the detection of some particular species can be argued.

In Figure 6 it is possible to see that the individual species are quite distinct. In particular, methanol and benzene are close while triethylamine and pentane are displayed on the opposite side of the plot. Finally, the acetic acid lies in an orthogonal direction with respect to the others, which indicates that the interaction mechanism for acetic acid is completely different from the others.

The position of the sensors is also of great interest. As argued cavitands *1* and *3* lie in the same direction as the alcohol, in accordance with their selectivity towards alcohols. On the other hand, the co-linearity with benzene is not expected. This is a hint that in these cavitands the hydrogen bond interaction is improved but at the same time other kinds of interactions have a comparable magnitude.

Studying the loadings in Figure 6, the role of the length of the alkyl chain in porphyrins is also evident. *9* reveals in the plot a good affinity towards acetic acid, while the addition of alkyl chains to the CoTPP skeleton increases the importance of sorption interaction. This is clearly visible in Figure 6 where the sensors shift from *9* up to *12*, covering the path leading from a high sensitivity towards acetic acid to a dominant sensitivity towards pentane, for which only sorption interaction is assumed to be present.

Other sensors are mostly grouped towards the triethylamine. In the case of porphyrins *6-8*, *13* the coordinated metal is no longer able to drive the selectivity pattern and the presence of the peripheral alkyl chains completely shadows the coordination interactions. This result can explain the failure to observe the coordination interaction in the sensing mechanism of the metal complexes of the closely related alkyl

chains functionalized phthalocyanines reported in the past by Göpel and coworkers [22].

In order to better investigate the relationship between sensor response and interaction mechanism it is useful to consider the way in which each volatile compound is expected to interact when in contact with a solid phase. These interactions can be modelled using the linear sorption energy relationship approach (LSER) [23].

According to this method and under the hypothesis of weak solubility interactions, the logarithm of the partition coefficient of a sorbent layer with respect to a certain volatile species is the linear combination of five terms expressing the intensity of five basic interaction mechanisms. They are: polarizability, polarity, two terms describe the hydrogen bonding considering the analyte acting as an acid and a base respectively, and finally the solubility terms, a combination of dispersion and cavity interactions.

The relation can then be written as:

$$\log K = K_0 + r \cdot R + s \cdot \pi + a \cdot \alpha_2^H + b \cdot \beta_2^H + I \cdot \log L^{16}. \quad (12)$$

Here K is the layer partition coefficient, R , π , α_2^H , β_2^H , and $\log L^{16}$ are the coefficients of the volatile compounds and r , s , a , b , and l are the coefficients of the absorbing material.

LSER coefficients for various analytes are available in [23]. Figure 7 shows the values of the five parameters for the five volatile compounds considered here.

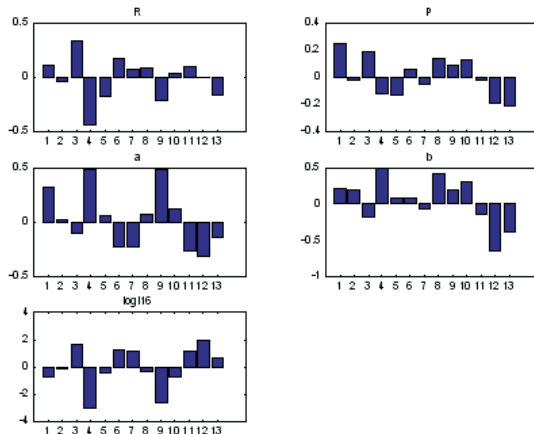


Figure 7.

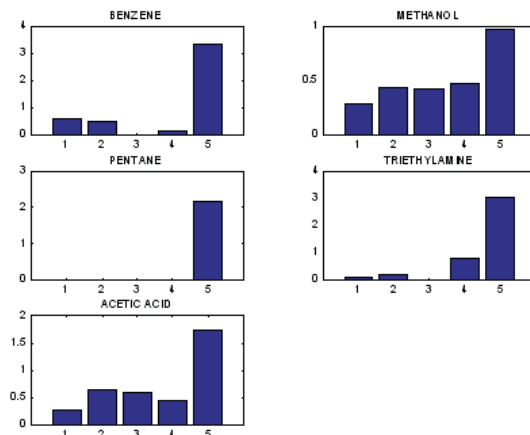


Figure 8.

As can be seen, all the compounds have a strong solubility interaction. Nonetheless Figure 6 shows that they are also well separated, so the more subtle difference between the other four interaction terms are important for their discrimination. It is worth noting that pentane interacts only via solubility, benzene has a very small hydrogen bonding term, while hydrogen bonding is present in methanol and acetic acid.

For a TSMR sensor, the partition coefficient turns out to be proportional to the overall sensitivity of the sensor ($S = \frac{\Delta f}{c}$) [24]. For some sensors a non-linear behaviour between sensor response and concentration is expected; in particular, for those cavitands functionalized to improve hydrogen bond interactions (labelled as 1 and 3 in Figure 8). In these cases, being limited by the number of interaction sites, the characteristic is expected to be steeper at low concentrations and then to reduce the slope at higher concentrations when the interaction sites may be considered completely occupied. These behaviours have been recorded several times once an absorbing layer passes some specific interaction sites [25],[26]. In order to take account of this fact and also to consider sensor sensitivity at its best, sensitivities and partition coefficients have been calculated at small concentration ranges where their values are higher.

The calculation of the partition coefficients for each sensor and for each volatile compound can then be used to estimate the value of the parameters (r, s, a, b, and l) characterizing the way in which each sensing layer may interact with volatile compounds.

Equation 12 can be written in matrix form as:

$$K_{n \times m} = A_{n \times m} \cdot S_{n \times m} \quad (13)$$

where K is the partition coefficient matrix, A is the matrix of the volatile compound LSER parameters, and S is the matrix containing the sensor LSER parameters. n is the number of volatile compounds and m is the number of sensors (5 and 13, respectively, in this study).

K and A being known, the solution of equation 13 allows the determination of the LSER parameters characterizing the sensors studied here. Equation 12 was solved with the least squares method. In Figure 8 the LSER parameters for each sensor are shown.

It is interesting to compare the parameters exhibited by the cavitands with and without the functional group improving hydrogen bonding. The effect of the functionalization is expected to increase the parameter a ; this effectively happens for 1 and 2. The stronger effect of polarization is instead observed for the polarization term. This explains the position of the benzene data in the plot of Figure 8. Another interesting effect is observed for compounds 9-12, where alkyl chains of increasing length are introduced at the peripheral positions of the macrocycle; the decrease of the hydrogen bond (parameters a , b) and polar (parameter R) interactions are observed while at the same time the solution interaction grows continuously from negative to positive values. These effects are best observed in Figure 9 where a , b , P , and l are plotted versus the length of the alkyl chain. As the chain length increases the relative importance of the chain with respect to the porphyrin grows and the interaction with the porphyrin becomes quantitatively negligible.

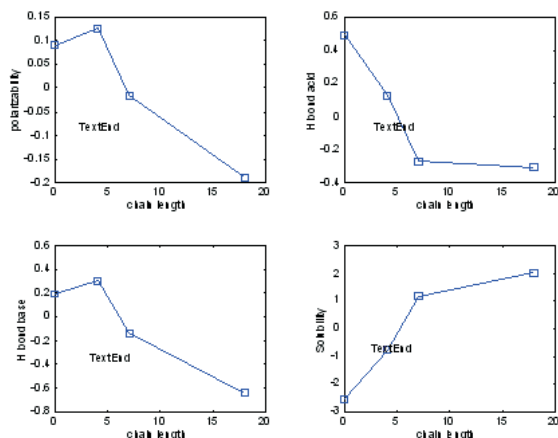


Figure 9.

This effect is complete in the case of porphyrins *6-8,13*, which behave in the same way, independent of metal present in the inner core of the macrocycle. In conclusion, the introduction of these alkyl chains from one side gives a higher porosity to the molecular film and, as a consequence, both a speed up of the response and an increase of the sensor response. On the other hand there is a significant decrease of the importance of the selective interactions in the sensing mechanism of the organic material and, as a consequence, a lowering of the selectivity of the sensor.

A good balance of these two opposite requirements is necessary in order to develop selective sensors.

9. Conclusions

Pattern recognition applied to several disciplines and practical problems produced a huge number of algorithms and techniques that are, in principle, applicable to the classification of chemical sensor array data. Several of these techniques were actually utilized and frequently appeared in the literature. On the other hand, there is not yet an analysis approach based on natural olfaction paradigms. This makes the field almost totally dependent on the developments achieved in other fields. A comparison among the various techniques has also been attempted by some authors, but since the variables determining electronic nose performance are numerous (choice of sensors, samples, sampling systems, . . .) the results achieved are scarcely indicative of a general direction. Nonetheless some of these techniques became a classic of this field (such as scaling and normalization, PCA and discriminant analysis) and are often used without a deep understanding of the hypotheses on which they are based. The potentiality of these methods is not yet fully exploited. As for the other components of chemical sensor systems (e.g. sampling systems) a rethinking of the assumptions, implementations, and interpretation of the methods and solutions adopted is highly advised for a meaningful improvement in the field.

References

- [1] K. Persaud, G. Dodd, *Nature*
- [2] T. Pearce; *Biosystems*, 1997 41 pp. 43-67
- [3] C. Di Natale, F. Davide, A. D'Amico; *Sensors and Actuators B* 1995 23 pp 111
- [4] B.R. Kowalski and S. Wold; in *Handbook of statistics Vol. 2*, P.R. Krishnaiah and L.N. Kanal eds., North Holland Publ. (Amsterdam, The Netherlands), 1982 pp. 673 – 697
- [5] Hierlemann A., Schweizer M., Weimar U., Göpel W.; in *Sensors update Vol. 2*, W. Göpel, J. Hesse, H. Baltés (eds.), VCH (Weinheim, Germany) 1995

- [6] E.L. Hines, E. Llobet, J.W. Gardner; IEE Proc. –Circuits Devices Syst. 1999 146 pp. 297-310
- [7] P.C. Jurs, G.A. Bakken, H.E. McClelland; Chemical Review 2000 100 pp. 2649-2678
- [8] T. Eklöv, P. Mårtensson, I. Lundström; Analytica Chimica Acta 1997 353 pp. 291-300
- [9] E. Martinelli, C. Falconi, A. D’Amico, C. Di Natale; Sensors and Actuators B 95 2003 132-139
- [10] Martinelli E., Pennazza G., Di Natale C., D’Amico A.; Sensors And Actuators B 101 2004 346-352
- [11] C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D’Amico, A. Legin, L. Lvova, A. Rudnitskaya, Y. Vlasov; Sensors and Actuators B 2000 64 pp. 15-21
- [12] G. Horner, C. Hierold; Sensors and Actuators B, 1990 2 p.p. 173-184
- [13] C. Di Natale, R. Paolesse, A. Macagnano, V.I. Troitsky, T.S. Berzina, A. D’Amico; Analytica Chimica Acta 1999 384 pp. 249-259
- [14] K. Fukunaga; Introduction to statistical pattern recognition, Academic Press, New York (NY, USA) 1992 H. Hotelling; J. Educat. Psych. 1933 24 pp. 498
- [15] D.L. Massart, B.G. Vandegiste, S.N. Deming, Y. Michotte, L. Kaufmann; Data handling in science and technology vol.2: Chemometrics: a textbook, Elsevier (Amsterdam, The Netherlands) 1988
- [16] R. Grossberg, I. Procaccia; Physica, 1983 9D pp.189-208
- [17] E. Oja; Neural Networks, 1992 5 pp. 927-935
- [18] J.F. Cardoso; Proc. of IEEE, 1998 9 pp. 2009-2025
- [19] C. Di Natale, E. Martinelli, A. D’Amico; Sensors and Actuators B 82 2002 158-165
- [20] O.M. Kvalheim, K. Oygard, O. Grahl-Nielsen; Analytica Chimica Acta, 1983 150 pp. 145
- [21] T. Hastie, W. Stuetzle; Journal of the American Statistical Association, 1989 84 pp. 502-516 M.A. Kramer; AIChE Journal, 1991, 37 pp. 233-243
- [22] M. Pardo, G. Sberveglieri, S. Gardini, E. Dalcanale; Sensors and Actuators B, 2000 69 pp. 359-365
- [23] H. Kramer, M. Matthewes; IRE Trans. Inf. Theory, 1952 IT-2 pp.41
- [24] T. Kohonen; Self Organising Map, 1995 Springer Verlag, Berlin Germany
- [25] F. Davide, C. Di Natale, A. D’Amico; Sens. and Act. B, 1994 18 pp. 244 4.
- [26] C. Di Natale, F. Davide, A. D’Amico, A. Hierleman, M. Schweizer, J. Mitrovics, U. Weimar, W. Göpel; Sens. and Act. B 1995 25 pp. 808
- [27] G. Kraus, A. Hierleman, G. Gauglitz, W. Göpel: Technical Digest of Transducers ‘95 Conference, Stockholm (Sweden) 25-29 Jun. 1995, pp. 1675-678
- [28] M.A. Kraaijvels, J. Mao, A.K. Kain; Proc. Of 11th Int. Conf. On Pattern Recognition, 1992 IEEE Comp. Soc. Press, Los Alamitos (CA, USA) pp. 41
- [29] C. Di Natale, A. Macagnano, A. D’Amico, F. Davide; Meas. Sci. and Techn. 1997 8 pp. 1-8

- [30] C. Di Natale, J.A.J. Brunink, F. Bungaro, F. Davide, A. D'Amico, R. Paolesse, T. Boschi, M. Faccio, G. Ferri; *Measurement Science and Technology* 7 1996 1103-1114
- [31] J.W. Sammons; *IEEE Trans. Comp.*, 1969 C-18, pp. 401
- [32] J.E.Dennis, R.B. Schnabel; *Numerical methods for unconstrained optimization and non-linear equations*, Prentice Hall Series in computational mathematics, New York, (NY, USA), 1983
- [33] R.A. Johnson and D.W. Wichern; *Applied multivariate statistical analysis*, Prentice Hall, Englewood Cliffs (NJ, USA), 1982
- [34] P.A. Lachenbruch, R.M. Mickey; *Technometrics* 1968 10 pp. 1-11
- [35] B. Efron; *Annual Statistics*, 1979 7 pp. 1-26
- [36] Y. Hamamoto, S. Uchimura, S. Tomita; *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1997 19 pp. 73-79
- [37] S.L. Campbell, C.D. Meyer; *Generalized inverses of linear transformations*, Pitman (London, UK) 1979
- [38] A.J. Maren (editor); *Handbook of neural computing applications*, J. Wiley and sons, (London, UK) 1991
- [39] D.E. Rumelhart, J.L. McClelland; *Parallel Distributed Processing: Explorations in the microstructure of cognition Vol.1: Learning internal representations by error propagation*, MIT Press, Cambridge (MA, USA) 1986
- [40] J. Hertz, A. Krogh, R.G. Palmer; *Introduction to the theory of Neural Computation Vol.1*, Addison Wesley, New York (NY,USA) 1991
- [41] J.W. Gardner, E.L. Hines and M. Wilkinson; *Measurement Science and Techn.*, 1990 1 pp. 446-451
- [42] *Neural Computing: a technology handbook for Professional II/plus© users*, NeuralWare Inc., Pittsburgh (USA) 1993
- [43] C.E. Martin, S.K. Rogers, D.W. Ruck; *Proc. IEEE Int. Conf. Neural Network*, 1994 pp. 305-308
- [44] N. Ueda, R. Nakano; *Proc. IEEE Int. Conf. Neural Network*, 1995 vol. 1 pp. 101-105