

Signal and Data Processing for Machine Olfaction and Chemical Sensing: A Review

Santiago Marco, Agustín Gutiérrez-Gálvez

Departament d'Electrònica, Universitat de Barcelona, Martí i Franquès 1, 08028-Barcelona. Spain.

Institute for Bioengineering of Catalonia, Baldiri i Reixach 4-8, 08028-Barcelona. Spain.

Abstract

Signal and data processing are essential elements in electronic noses as well as in most chemical sensing instruments. The multivariate responses obtained by chemical sensor arrays require signal and data processing to carry out the fundamental tasks of odor identification (classification), concentration estimation (regression) and grouping of similar odors (clustering). In the last decade, important advances have shown that proper processing can improve the robustness of the instruments against diverse perturbations, namely, environmental variables, background changes, drift, etc. This article reviews the advances made in recent years in signal and data processing for machine olfaction and chemical sensing.

1. Introduction

Since the original definition of *electronic nose* by Prof. J.W. Gardner in 1994 [1], it is clear that pattern recognition or, more generally, signal and data processing, is an essential component of Artificial Olfaction instruments. In this regard, every particular implementation of an electronic nose includes a data analysis and interpretation module to obtain the final prediction based on the sensor responses. Data processing can be carried out off-line, after the measurement has been carried out, using a desktop computer with dedicated software, or it can be integrated directly into the instrument for on-line evaluation [2].

Ten years ago, Gutierrez-Osuna [3] reviewed the building blocks of pattern recognition systems for machine olfaction. Since then, the basic methodology for building the data processing component of the instruments is firmly established and some of the terminology used in this paper can be traced to this reference. There have been recent reviews on particular aspects of the data processing methodology [4], [5], [6]. The present paper has the objective of complementing and updating these previous revisions with a focus on proposals made in the last decade. Additionally, it will emphasize the potential of signal and data processing to ameliorate the performance of chemical sensing systems.

It is usually recognized that Artificial Olfaction Systems have not achieved the market penetration that was expected in the mid-90s. We firmly believe that a contributing factor to this issue has been a lack of a rigorous analysis of the underlying pitfalls. While electronic noses are still considered in many application sectors as unreliable instruments, many of the problems that plague e-noses are also present in a variety of spectroscopic techniques (e.g. but not only, Near Infrared Spectroscopy). It is well known that routine application of state of the art chemometric techniques has greatly improved the usability and predictive power of these techniques.

This paper does not intend to review every single paper covering the last decade of research in Machine Olfaction. In particular, the tremendous work devoted to sensor amelioration from the material viewpoint is outside the scope of this work. The last decade has been very fruitful in gaining insight into the various aspects that affect the sensing performance, namely: the effect of microstructure, the size feature of the material, the advantages of nanostructured materials (e.g. nanowires, nanorods, and nanoribbons) and the possibility of functionalizing sensing materials by adding catalysts. Further information on the progress of sensor technology from a materials point of view can be obtained elsewhere [7,8,9,10].

Instead, we will focus on those principled proposals that provide signal and data processing solutions to particular issues in Machine Olfaction. We have structured the contents into two sections: first we shall review the latest trends regarding the main building blocks of a general data processing unit; second we shall select the specific proposals aiming to improve the robustness of machine olfaction systems.

Due to size restrictions, this paper does not cover algorithmic developments regarding odor localization (see instead reviews [11,12]), and the very exciting topic of biologically inspired computation for chemical sensing (see reviews [13,14] and book [15]). Finally, it is not the objective of this paper to provide a tutorial on pattern recognition or chemometrics. For an introduction to some the methods and techniques, as well as the basic terminology that appears throughout this review, there are excellent references for general pattern recognition, chemometrics and learning systems [16,17,18].

2. Basic building blocks

Qualitative and quantitative predictive models in chemical instrumentation include a number of conventional building blocks. Chemical sensor signals have to be preprocessed, informative features have to be extracted and/or selected and finally a classifier is built for label prediction or a regressor for quantitative prediction.

Before addressing the latest proposals for the different building blocks, it is worth reviewing the main characteristics of chemical sensor signals. It is clear that different sensor operational modes will provide diverse signals types. For space reasons, we cannot illustrate all of them here. We will, however, revise those routinely encountered in the literature.

In its simplest configuration, the sensor array is continuously exposed to the target atmosphere and the sensors are not enclosed in any chamber but directly embedded in the gas. In this case, sensor signals follow the time-varying gas composition of the monitored area. For instance, this is a typical configuration in low-cost safety alarms. Alternatively, there could be an active sampling system that switches from a reference condition (typically pure or filtered air) to the gas under analysis. This has been a typical configuration in most commercial and lab-made instruments. In this second configuration, the exposure of the sensor array to a reference condition provides a valuable way to check the stability of the system by observing the time evolution of the sensors baseline. It is well-known that simple baseline correction techniques are an effective way to remove system drift (see section 3). Additionally, and particularly in headspace sampling cases, the sensor transients while the chamber is filled, and curiously enough the transients, when purging the chamber with pure air, provide additional information about the sample analyzed [19]. Figure 1 displays the typical sampling transients from a conductive polymer sensor when exposed to three different vapors. Observe how the waveform is slightly different. The search for the discriminant features in such waveforms has been the focus of intense research in the last decade.

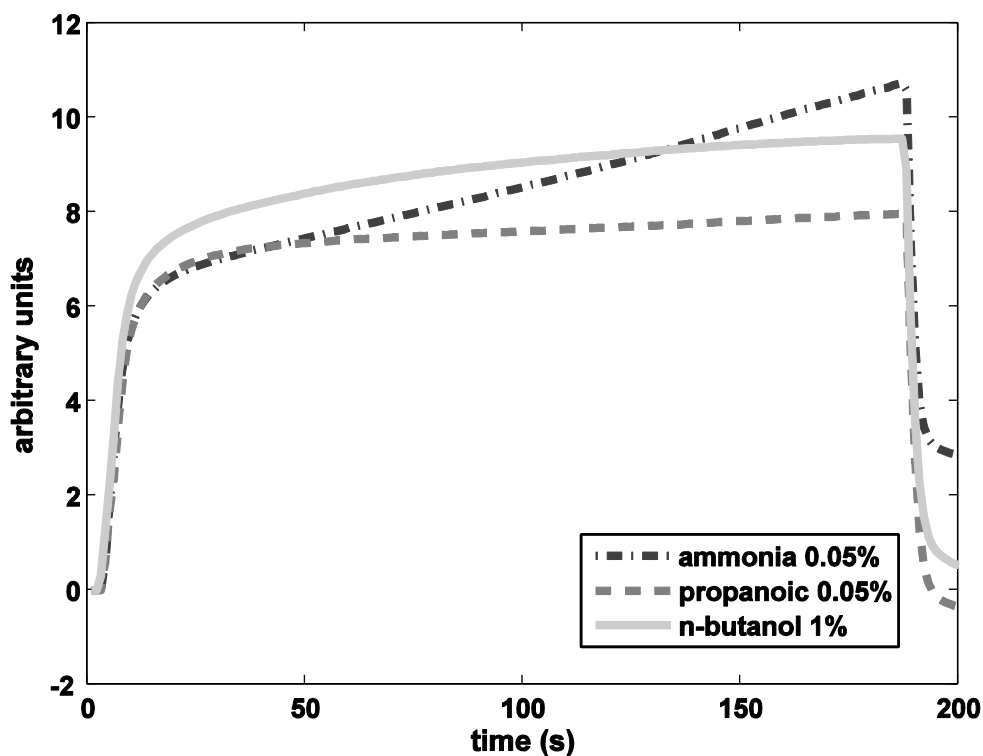


Figure 1. Sampling transients for a conductive polymer sensor when exposed to three different vapors.

On the other hand, if we just take the steady state response we obtain a feature vector with as many dimensions as the cardinality of the sensor array. Figure 2 shows the patterns for a 17 elements sensor array. Observe the high correlation among sensors.

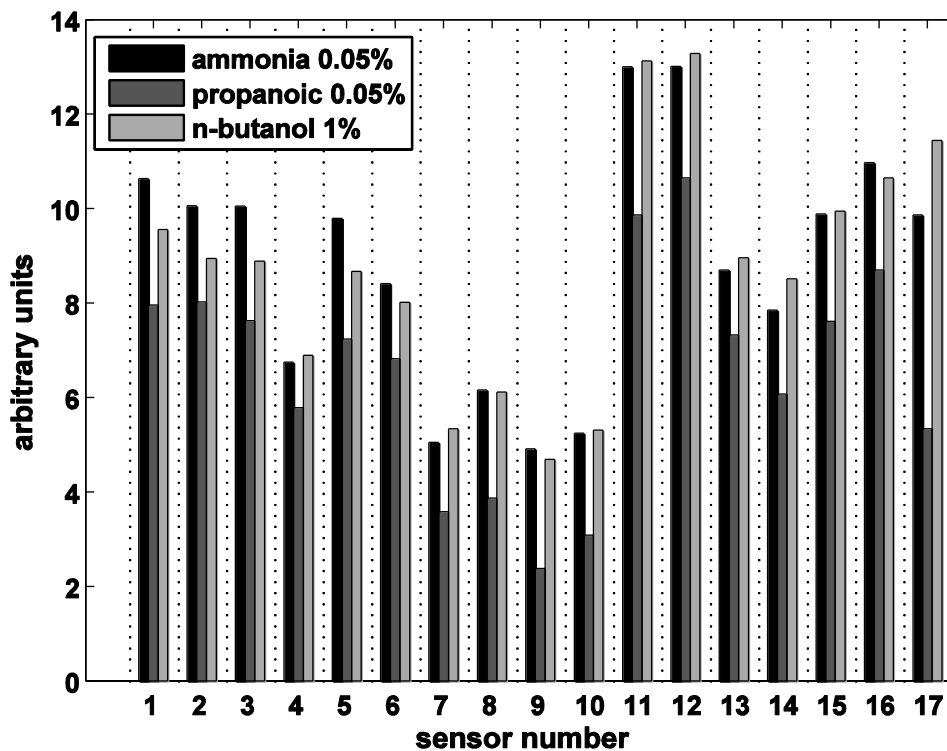


Figure 2. Steady-state patterns for three different vapors in a 17 element sensor array for three different analytes.

In addition to sensor sampling transients, the potential for sensor tuning using temperature modulation in metal oxide sensors has received a great deal of attention. For instance, figure 3 illustrates a FIS SB-31 response under temperature modulation (a sawtooth voltage was applied to the heater element) while monitoring a mixture of CO and CH₄ in air. It is very clear how the changes in the composition over time produce a different pattern of response.

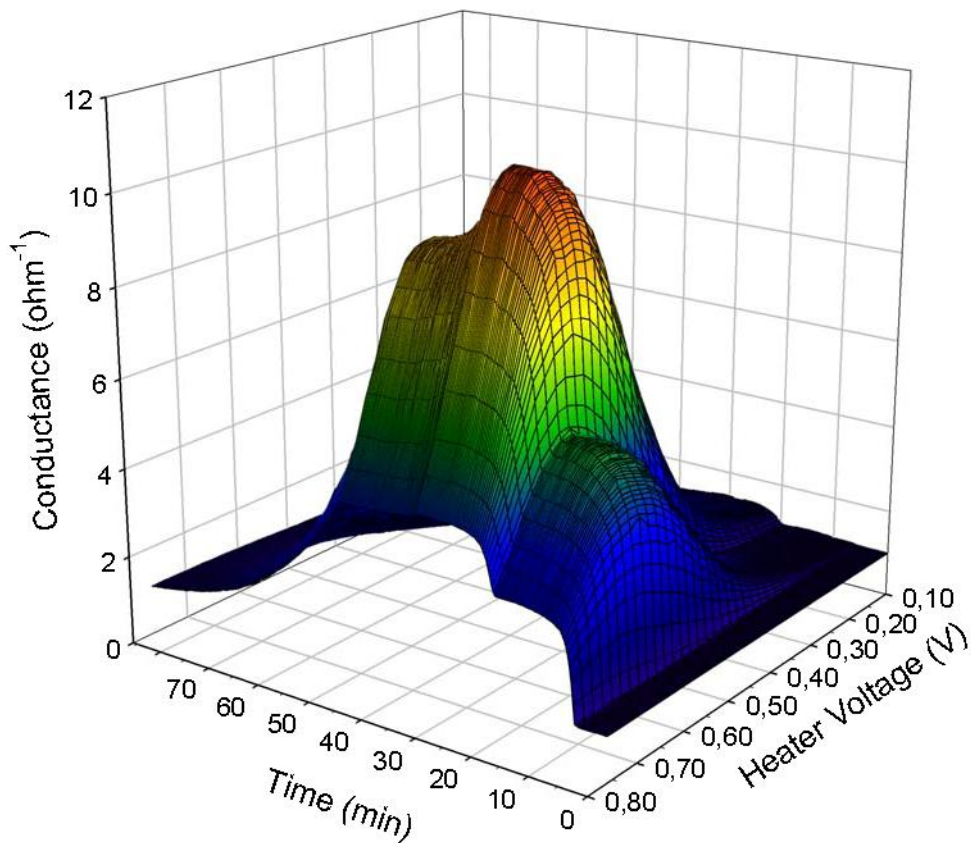


Figure 3. Evolution of the pattern of response of a SB31 sensor from FIS Japan when the heated voltage is scanned due to changes in the gas composition (mixture of CO and CH₄).

One of the problems that affect sensor array instruments is the lack of reproducibility in the sensor signals when technologically identical devices are exposed to a chemical under the exact same conditions. Figure 4 shows the voltage waveforms of 11 temperature modulated Metal Oxide Sensors. It is obvious that the pattern of response is slightly different for both analytes, but the waveform scatter is also remarkable, which poses problems during sensor replacement and calibration transfer scenarios (see section 3).

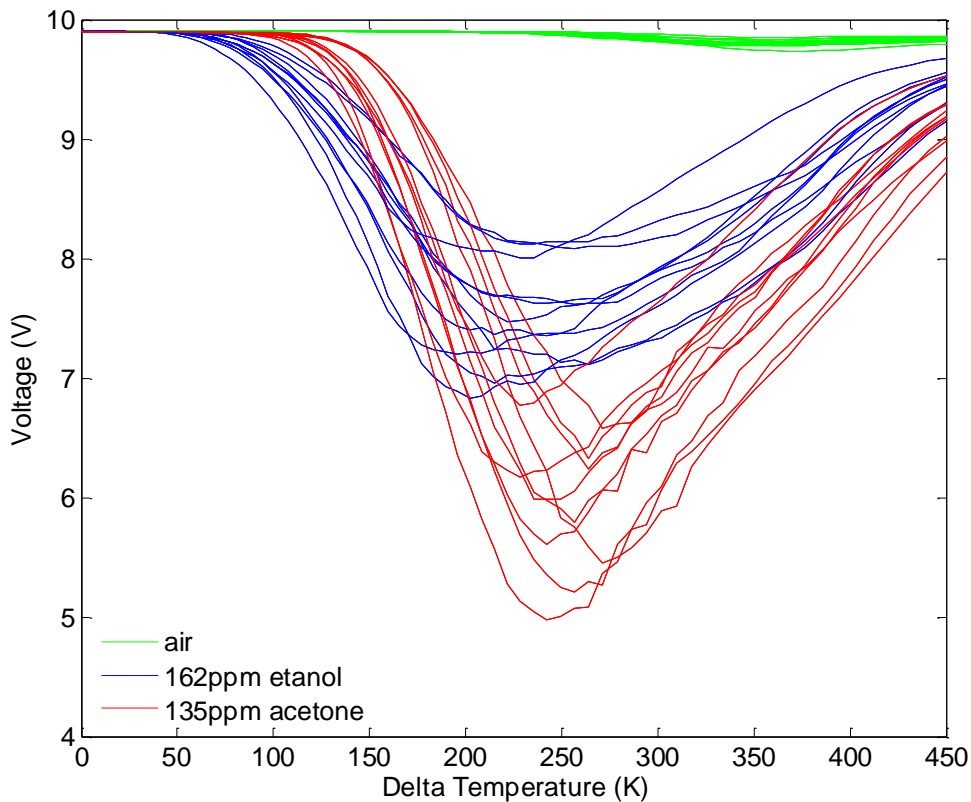


Figure 4. Response of 11 identical Metal-Oxide Sensors in a half-bridge configuration when the heater temperature is scanned. Note the different waveform shape for acetone and ethanol, as well as the waveform scatter in different sensor copies.

2.1 Dimensionality Reduction

2.1.1 Feature extraction

Working with raw, time-sampled sensor responses leads to high dimensional patterns. To avoid the curse of dimensionality, an explicit or implicit feature extraction block is needed. The aim of feature extraction is to obtain informative features from mathematical transformations of the multivariate time responses of chemical sensor arrays. These transformations reduce the dimensionality of the input space, and aim to keep the information relating to the target problem. Beyond the straightforward use of steady state sensor responses, there is no consensus on the best techniques for feature extraction.

Nevertheless, there is convincing evidence that scanning internal sensor operation parameters (such as temperature in Metal-Oxide [20,21] sensor, voltage in voltammetric sensors [22] or mirror position in Fabry-Perot Interferometers [23]), provides a first order response (that is, a vector instead of a scalar) from a single sensor and additional discriminative information. It is

also well-known that the use of headspace sampling transients provides richer sensor responses improving their information content [19,24].

The analysis of first order sensor signals originating from the operational modes above requires a feature extraction stage. Several approaches can be found in the literature: (1) direct use of subspace projection methods such as PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), FFT (Fast Fourier Transform), DWT (Discrete Wavelet Transform), etc., (2) curve fitting methods, (3) waveform descriptors, (4) non-linear subspaces projection (SOM (Self-Organizing Maps), Sammons mapping, etc), (5) clustering in feature space, etc. It is also common practice to concatenate several feature extraction methods. For instance, a typical implementation could extract a feature vector based on waveform descriptors, then PCA to reduce dimensionality and avoid overfitting in a final LDA step-

While PLS (Partial Least Squares) was not originally designed for discrimination purposes, PLS in conjunction with a binary class-coding for the Y block leads to PLS-DA (Partial Least Squares-Discriminant Analysis) that is often found in the sensor feature extraction literature. In addition to the direct use of the class prediction technique, PLS-DA can be thought of as a subspace feature extraction technique, projecting X data onto the subspace spanned by the latent variables. The relationship between PLS-DA and LDA has been explored from a geometrical point of view [25].

Subspace methods have a solid mathematical grounding and are favored by many researchers. With PCA and LDA the most popular techniques, they have their own drawbacks. PCA aims to capture maximum variance in few dimensions, ignoring discriminant information. On the other hand, LDA is a supervised method but assumes unimodal Gaussian classes with different mean and equal class-covariance. Additionally, it is well-known that LDA is prone to overfitting, providing false class separability on the training set for small sample number to dimensionality ratio. LDA has the additional limitation of having a maximum dimensionality given by $C-1$, with C the number of classes. PLS-DA overcomes this limitation, but focuses more on maximizing mean class distances without counterbalancing within class variance.

Lindgren et al. [26] presents a linear feature extraction technique called Asymmetric Class Projection (ACP) that is especially thought to deal with asymmetric distributions of data. ACP aims at obtaining discriminant features between the classes involved, in a similar way to the LDA. However, ACP is able to find better discriminant features than LDA when the class distributions are concentric and have different variance distributions, but in the absence of a symmetry plane. The technique is successfully tested with synthetic data but provides comparable results to LDA for an experimental data set.

A subspace projection method with biomimetic inspiration was proposed by Perera et al. [27]. They presented a dimensionality reduction technique inspired by the receptor convergence present in the olfactory system. To achieve this goal, a projection of the initial features onto the class space is made, keeping the initial sensor variance. Authors suggest their usefulness for those specific cases in which only a small number of high dimensionality samples is available. The method is tested with a set of experimental data coming from temperature modulated gas sensors stimulated with a sinusoidal profile.

In the last ten years the focus has been on analysis of dynamic sensor signals. While the Fast Fourier Transform was already being used by several authors (e.g. Heilig et al. [28]) in the 90s, the trend has been towards exploring more powerful multi-resolution techniques for transient signals analysis. Among them, the Discrete Wavelet Transform for feature extraction has been frequently used since the late 90s [29,30]. When using DWT, usually a posterior analysis is needed to find out at which decomposition level class separability is maximum. It has been frequently found that DWT outperforms FFT [31]. DWT has been used for processing sampling transients, temperature-modulated MOX sensor signals, but also voltammetric sensors [32]. In this last work the authors propose an ANOVA based feature selection method to find the most discriminate wavelet coefficients. In its particular application DWT outperforms PCA.

Acevedo et al. [33] compares classical statistical methods as PCA and LDA to DWT using a probabilistic Support Vector Machine (SVM) as classifier. They conclude that PCA and LDA provide the best results, delivering projection subspaces with lower dimensionalities than those delivered by DWT to obtain similar performances. On the other hand, the authors have commented that the projection matrices for DWT do not require training data.

While orthogonal basis functions such as sinusoidals and wavelets have been used extensively, there has been recently proposed to investigate the use of other orthogonal basis such as Bessel Functions. In a recent paper, Vergara et al. [34] use this orthogonal basis to model signal transients. They claim that these functions are better suited to capture the transient characteristics of the sensor response. They compare the performance with Exponential Moving Average (EMA) [35] feature extractors and steady state responses, resulting in better results in classification. They also test these features to build predictive models using PLS.

When trying to represent sensor signals using basis functions, there can be multiple choices. However, the question remains as to how to select the most appropriate basis for processing a particular class of signals. Different basis functions will highlight different properties of the signals under analysis. Using a purely signal representation criterion, Leone et al. [36] propose an automatic method to select the best signals for the expansion. The proposal is based on the use of overcomplete dictionaries and then the matching pursuit method in order to select the optimal atom of the dictionary. While the procedure provides extremely compact signal representations and then quasi-optimal compression (in fact, optimal compression is given by PCA), this is not a guarantee for informative features.

Some authors have explored the analysis of transients using multi-way methods, taking advantage of the data structure (transient time x samples x sensors). Skov and Bro introduced sensor array trilinear modeling for the first time using PARAFAC2 [37]. This technique is able to handle small time shifts that may appear in sensor waveforms due to instabilities in sampling flows. Padilla et al. [38] used PARAFAC to model sampling transients of headspace analysis of potato chips with different odorant levels. They found that one of the loadings was correlated to the odorant quantity enabling an easy linear calibration model.

Martinelli et al. [39] proposed a feature extraction method for sensor waveforms. Lending inspiration from the canonical variables of analytical mechanics, they use phase space to

represent the dynamics of the sensor responses and extract features for odor discrimination and concentration estimation. The phase space includes, as dimensions, the response of the sensors along with their first time derivatives. The authors propose to use the area spanned by a phase space trajectory as a feature during the sensor adsorption and desorption processes. The classification and concentration results estimations obtained with the phase-space area outperform those obtained with the simple steady state as feature in androstenone and apple ripening detection problems. This feature extraction technique was used later by Falasconi et al. [40] to evaluate coffee ripening. After comparing against R/R_0 and the sensor response integral, the phase-space integral provided the best performance. Martinelli et al. [41] and later on Vergara et al. [42], also exploited the representation potential of the phase space to propose dynamic moments as new features. Dynamic moments are morphological descriptors of phase trajectories previously proposed to describe chaotic system dynamics. This approach was applied to the analysis of transient responses due to changes in concentration and/or thermal modulation of metal oxide gas sensors. To test its performance, two sets of data are prepared. The first consists of samples of headspace of potato crisps. These samples are sorted according to their rancidity state. After data analysis, the use of the proposed approach leads to an improvement of the classification. The second dataset contains the response of temperature-modulated micro-hotplate sensors. In this case, phase-space methods give the opportunity to use novel features that describe sensor trajectories, while adsorption and reaction kinetics are altered by the temperature modulation. The new set of features is used to build a supervised pattern recognition model based on Partial Least Squares (PLS-DA). The results obtained provide an increase in the gas identification rate and also a great reduction in the time needed to do the measurements. Working also in phase space, Zhang et al. [43] propose a new set of heuristic features to describe the trajectory. Their results overcome the use of the phase-space integral. They do not compare with dynamic moments.

To extract features from sensor waveforms in temperature modulated micro-hotplate gas sensor arrays, Vergara et al. [44] introduced the so called energy vector (EV) approach for feature extraction. In this proposal, the feature vector is formed by the $N \cdot (N-1)$ cross-energies between all the sensor waveforms (N being the number of sensors). Cross-energies can be seen as correlations between the sensor waveforms. The rationale behind this proposal is that these correlations are dependent on the sensed gas. In their study, the cross-energy feature vector compares favorably to features extracted by Fast Fourier Transform, but they do not compare with previous proposals by the same authors. The classifier used was PLS-DA.

When working with fast time-varying concentration inputs, either of random nature or step-like, sensors exhibit dynamic responses [45]. In those cases, the predictive power of the sensor array is enhanced, using time-delayed inputs as features. This was explored by Pardo et al. [46] when trying to predict the concentrations of binary mixtures of toluene and octane from using a QMB sensor array. This idea has been further explored by Pardo et al. [47], Zanchettin [48] and De Vito [49]. With a different approach, Yamanaka et al. [50] used linear state-variable dynamic models to increase the speed of the analysis for fast-varying inputs.

For step like concentration inputs, and in order to accelerate the computation of the extracted features, Muezzinoglu et al. [35] proposed the use of an exponential moving average feature that is readily available at the beginning of the sensor transient. They show that is not

necessary to wait for the steady state in order to obtain good results in quantification and identification problems. This confirms earlier results by other groups [51].

Some authors have found that the MOX sensor response dynamics to a heater temperature change depend on the target gas and its concentrations. To estimate the dynamics, linear system identification techniques (e.g. ARMAX) are used [52]. The estimated parameters become the input features for later gas identification and concentration estimation.

Curve fitting has been in use since the early 90s: examples are multi-exponential [24,53,54] and multi-Gaussians models [55]. The fitted model parameters become the extracted features. Carmel et al. [56] presented the use of a Lorentzian model for the sensor responses with good results. Later, Haddad et al. [57] presented a generalization of the Lorentzian model for signals exhibiting more than a single peak. Curve fitting can result in extremely compact features if the selected basis suits the shape of the sensor waveform. However, on many occasions, curve fitting requires non-linear parameter optimization techniques. It is well-known that on many occasions, the feasible parameter space is nonconvex and therefore leads to multiple local minima. If the optimization algorithm gets trapped in diverse local minima due to small changes in the sensor waveform, this could produce a large variability on the selected features and ultimately it could produce unreliable results. This is mostly the case when the basis functions stay highly correlated despite changes in the inner parameters.

E. Llobet et al [58] proposed to use self-organizing maps as feature compressors in a problem of quantitative analysis of binary mixtures using FT-NIR spectrophotometry. He proposes to project the input data (spectra) in the subspace generated by the neurons of a previously trained self-organizing map. From 777 initial wavelengths provided by the spectrometer, they reduce the dimensionality to 40. They found that the linear regression models built in the reduced space provide lower error than in the original space.

During the last ten years, feature extraction proposals have been mostly focused on extracting information from sensor waveforms generated by sampling transients or by external modulation of a tuning parameter. There is a large number of evidence that transient signals in fact carry additional information beyond the steady-state response. In a similar way, temperature modulated metal oxide sensors provide richer signals than isothermal response. Obviously there are infinite ways to describe a sensor transient signal and a large number of descriptive features can be extracted. However, it is important to mention that in the context of chemical sensing, signal representation is not the ultimate goal. In other words, while some sets of features can adequately describe the signal shape, this does not guarantee that those features are especially informative about the predictive task. In any case, any compact representation of the signal dynamics can provide adequate dimensionality reduction and reject noise, and in this way, they can reduce the problems associated with the curse of dimensionality. Other proposals, like the energy vector approach by Vergara, do not describe the signal shape directly but try to extract features that could be affected by changes in the sample composition. Since most of the approaches presented do not optimize any particular figure of merit, a posterior step of Linear Discriminant Analysis, PLS-DA is used in order to find the proper weight to the different extracted features. Alternatively, the option is to optimize a subset of features by feature selection.

2.1.2 Feature Selection and Sensor optimization

Feature selection is a dimensionality reduction approach that consists of selecting a certain number features from all the available ones for subsequent use by a learning algorithm. In instrumental development, feature selection techniques are closely related to system optimization. For instance, when features are grouped by sensor, feature selection or better feature-group selection algorithms can be used to finally choose the final subset of sensors to be included in the instrument. For tunable sensors (e.g. Spectral filter locations in optical sensors, voltages in voltammetric sensors or temperatures in Metal Oxide Sensors), when the features correspond to the sensor responses for specific values of the tuning parameter, the optimization of the instrument operation can also be posed as a feature selection problem. It is important to note that in this formulation the sensor/instrument response is available for every possible feature subset selection during the optimization process. This optimization is usually carried during the design phase. The reader can read more about this topic in the recent review by Vergara and Llobet [59].

Feature selection is a combinatorial problem and an exhaustive search can only be performed in some cases. If all potential combinations are allowed 2^N-1 possible configurations exist, N being the number of features. However, on many occasions the desired number of features (occasionally this is just the number of considered sensors) is decided a priori to be $M < N$, with $\binom{N}{M}$ the number of potential configurations to explore.

At this point it is important to note that when the number of sensors is limited, exhaustive search is an option [60], and computational methods only serve to speed up the search for the optimum configuration. In an example by Pardo et al. [61] from an initial set of 30 features from 6 chemical sensors, they selected the best subset with cardinality 1-5 using exhaustive search. Among the selected features, those based on phase space were selected among the best. The authors found that selecting only 2 features gave better results than the initial set of 30.

An early example of sensor selection in an array of 12 MOX sensors for a discrimination problem was presented by Pardo et al. [62] Sequential Backward Selection (SBS), Sequential Forward Selection (SFS) and Genetic Algorithms (GA) were compared with techniques based on the PCA or LDA scores. The global optimum with just 5 sensors was achieved only by Genetic Algorithms.

Sensor selection techniques, when applied to hybrid sensor arrays, provide further insight into the performances of the different technologies. Pardo et al, [60] after an exhaustive search, found that hybrid sensor arrays (in the sense of arrays comprising sensors of different technologies) gave better performance than homogeneous arrays. The importance of hybrid sensor arrays was already pointed out by Ulmer et al. to reduce sensor correlation [63].

In a recent work, B. Raman [64,65] applied sensor selection techniques to an array of metal oxide microhotplates under temperature modulation. The motivation problem was that of recognizing hazards in the presence of interferences and different levels of relative humidity. While the microhotplates are temperature modulated, their array optimization problem refers to the selection of metal-oxide materials and the number of replicas for every material type. To select the best subset they use a measure of cluster separability derived from Fisher's Linear Discriminant Analysis.

$$J = \frac{\text{trace}(S_B)}{\text{trace}(S_B) + \text{trace}(S_W)} \quad (1)$$

Where S_W and S_B are the within-cluster and between-cluster scatter matrices. The paper does not mention any search strategy, but the figure of merit is used to compare different subset options. Curiously enough, the best configuration found uses just two materials but four replicas of every material. The use of redundancy provides a better limit of detection, and allows discrimination of several analytes from the variability of the background. The best combination is tested for classification performance with leave-one-out, with very good results.

Similarly, in a recent paper Muezzinoglu et al. [66] optimized the active layer temperatures of two metal-oxide sensors TGS26000 and TGS2610 (Figaro, Japan). As a figure of merit to select the best working temperature, they proposed to use the squared Mahalanobis distance among the class means using a weighted average of the two covariance matrices.

$$D^2(C_1, C_2) = (\mu_1 - \mu_2)^T S_{12}^{-1} (\mu_1 - \mu_2) \quad (2)$$

Where μ_i is the i -class center and S_{12} is the weighted average of the two covariance matrices S_1 and S_2 . This approach is generalized to multi-class problems using as a figure of merit:

$$MD^2 = \sum_{i \neq j} D^2(C_i, C_j) \quad (3)$$

This approach is conceptually very similar to the previous one by Raman et al. in the sense that both are based on Gaussian classes and try to compare the distance among classes with the sample scatter for the classes. Although no comparison has been made, results should not differ much. In this paper, the authors do not propose any particular search strategy and they just scan the temperatures with both sensors working at the same temperature. The results of the prediction by this figure of merit are then compared with the classification rate of three classifiers: a kernelized support vector machine, a linear support vector machine and a nearest neighbor classifier, with good agreement

In a series of papers, Vergara [67,68,69] explores the optimization of the operating temperature modulation of metal oxide gas sensors. The basic idea is to excite the sensor with a multisinusoidal signal. The frequencies of the individual sinusoids are optimized for

maximum discrimination among the target odors. In this proposal, sensors are considered as linear systems where the input is the modulation voltage and the output is the sensor response. While the system is strictly non-linear, it can always be approximated in first order by a linear dynamic system in the sense of Volterra series expansion for dynamic non-linear systems [70]. The dynamic response of such a system depends on the sensed gas. It is then possible to estimate the impulse response of the system for every target gas. This is done by the use of white noise inputs and by computing the output-input cross-correlation. The frequency response is then calculated by the Fourier Transform of the impulse response. At this point, frequencies are ranked depending on a figure of merit named resolution power and defined as the between-group variance divided by the intra-group variance. The higher ranking frequencies are selected. Note that this approach does not consider interactions among the selected sinusoidal frequencies and it can be sensitive to spectral variance due to non-parametric spectral estimation proposed. The proposed approach outperforms the identification success with isothermal responses, but it is not compared to other simple modulation schemes (e.g. saw-tooth signals).

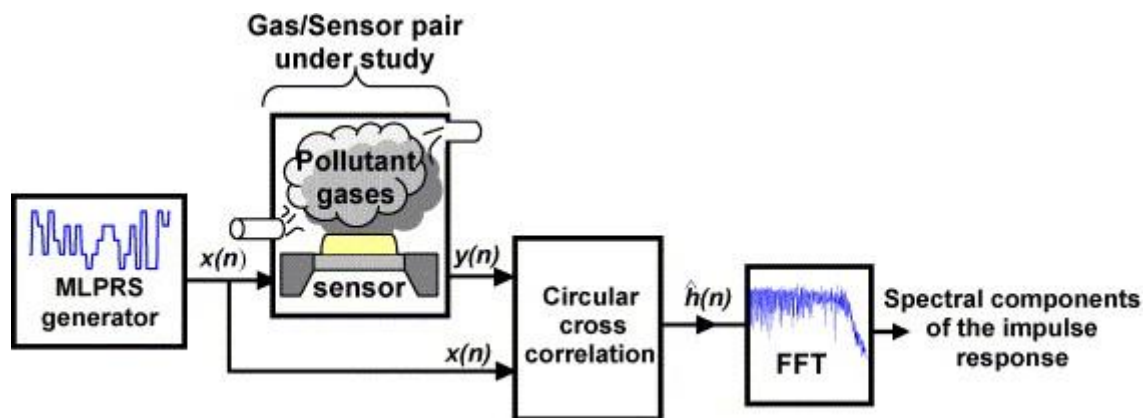


Figure 5. Study of the sensor/gas system using MLPRS (multi-level-pseudo-random-signals). The MLPRS voltage signal $x[n]$ is the input and the transient sensor conductance is the output. An estimate of the impulse response can be calculated by cross-correlation [69] (with permission of Elsevier Science).

Vergara et al. [71] approached the temperature optimization problem from an information theoretic perspective. They studied the use of the Kullback-Leibler divergence as a tool to find an optimum operating temperature for an odor discrimination problem. A symmetrized version of the divergence is used to compute the distance between the probability density distributions of the sensor responses for different odors. The maximum distance of the distributions across temperatures yields the optimum operating temperature. The method is applied to a multi-odor classification problem with 4 sensors that are optimized independently.

A more challenging approach to optimize the operation temperature of the sensors was followed by Gosangi and Gutierrez-Osuna [72]. They proposed a method based on active sensing to determine in real time the optimum sensor temperature profile for a discrimination problem. The method is based on Partially Observable Markov Decision Processes providing an optimized temperature profile that balances the risk of classifying the odorant against the power consumption cost of increasing the sensor temperature. The optimization method is

experimentally validated exposing Metal Oxide Semiconductor sensors to three analytes.

While the bulk of work on sensor optimization has been devoted to temperature selection in Metal-Oxide Sensors, some examples can be found, for instance, in optical sensors. Calaza et al. [23] explored the best sequence of mirror positions in a Fabry-Perot Interferometer using Genetic Algorithms. They conclude that it is not necessary to scan the full set of positions but it is possible to concentrate on the most discriminant ones. Fonollosa et al. [73]. When analyzing the performance of a four-channel non-dispersive infrared micro-spectrometer (NDIR) based on MEMS technology aiming to detect ethylene, ethanol and ammonia (plus a reference channel) they faced the problem of the optimum wavelength window for the spectrometer filters. They analyzed the angular distance between the sensitivity vectors for different optical window central wavelengths in order to avoid overly correlated channels and as a consequence, correlated feature vectors. However, in this paper, the authors do not introduce any automatic procedure for selection, or any specific goal function that combines sensitivity and specificity.

Among the different techniques for feature/sensor selection, Genetic Algorithms (GA) have been the preferred option because of their resilience to get trapped in local optima by many researchers [74,75]. J.W. Gardner et al. [76] specifically propose the enhancement of the performance of electronic noses through the use of genetic algorithms for sensor selection. They propose a GA technique in which the number of features is fixed from the beginning, instead of exploring the entire combinatorial space. Phaisangittisagul and Nagle [77] reported a sensor selection technique based on genetic algorithms to generate sensor subsets and a classifier as subset evaluation criteria (wrapper approach). They applied the optimization to features extracted with discrete wavelet transform (DWT) from the sampling transient response of the sensors. The technique was validated with MOX sensor responses to different types of coffee and soda. In a posterior paper [78], the same authors propose to speed up the technique, selecting the initial state for the genetic algorithm with the maximum margin criterion.

A different approach to speeding up the sensor selection process is to use sensor clustering to constrain the search space of sensor subsets [79,80]. It is usually the case that the sensors provide correlated responses. The principle behind this approach is that it is better to select sensors that are more orthogonal or diverse. Cluster analysis is used to detect groups of correlated sensors. An easy way to reduce the number of sensors is to select a single sensor per cluster, since neighboring sensors in feature space will provide only correlated information. Greater sensor diversity is found by selecting sensors from different clusters. To find the final configuration of the array, a genetic algorithm generates these constrained subsets that are evaluated based on a resolution factor proposed by the authors, based on averaging a Fisher ratio for all the input dimensions.

Llobet et al. [81] have proposed a feature selection method combining genetic algorithms and classification with fuzzy ARTMAP. This method addresses the problem of irrelevant feature selection of the Genetic algorithms by concatenating two GA based search stages. The resulting features of the first search stage are fed to the second GA search stage to further reduce the number of features and discard the irrelevant ones. The method is tested on the

classification of three volatile organic compounds.

Gualdron et al. [82] compared five feature selection techniques (SFS, SBS, GA, Simulated Annealing, stepwise selection) coupled with two neural based classifiers (Fuzzy ARTMAP and Probabilistic Neural Networks). The performance of the method is evaluated on the basis of three figures of merit: classification rate, number of selected features and computational efficiency. Their results show that with any of the two classifiers, SFS is faster, provides a parsimonious model and good classification rates.

Mass spectrometry based e-noses results in high dimensional features vectors. Specifically for these instruments, Llobet et al. [83] propose a three-step feature selection methodology. In the first step, they remove noisy features evaluated using a univariate Fisher ratio. In the second step, collinear features are identified and redundant features removed. In the last step, they apply GA or SA for fine tuning.

A different approach to sensor selection is based on the sensitivity of predictor models towards the different features. Szczówka et al. [84] approached sensor selection using a classical two-layer back-propagation MLP and sensitivity analysis to determine the more relevant sensors. They follow a backward subset selection with the removal of one sensor per iteration to find sensors that provided redundant information. The method was validated with data from MOX sensors estimating the concentration of organic compounds.

A method based on the evaluation of the margin vector in an SVM classifier was proposed by Gualdron et al. [85]. The authors first sorted the features according to the square value of the weights in the margin vector in decreasing value. They select the features with the largest weight and the optimum number of features in cross-validation. They validate the method with a dataset of chemical sensor responses and another of mass spectrometry based e-nose.

Current state of the art regarding feature selection makes heavy use of genetic algorithms, while some papers propose that sequential searches can provide similar performance with lower computational cost. Regarding the objective function, measures based on Fisher ratios and similar formulations (Mahalanobis distance) have been proposed by several groups. Other teams prefer to use a direct wrapper approach and introduce a classifier directly in the optimization loop.

These techniques are of special interest for their application to system optimization. Among the different proposals, we would like to highlight the adaptive optimization proposed by Gutierrez-Osuna as a significant step towards more intelligent chemical sensor machines.

2.2 Classification

Classification of chemical sensor data is, in many cases, plagued with problems. It is quite usual to encounter highly correlated features, scattering due to concentration [86] or drift, multimodality and non-Gaussian data distribution per class. Additionally, and due to high calibration costs, the number of training data is limited. For all these reasons, it is important to

select classifiers that are robust to the problems above. In the past, Multilayer Perceptrons [87] or Self-Organizing Maps [88] have been used to cope with these difficulties. In recent years, there is a trend to use statistical learning based methods. Among them the most widespread technique is Support Vector Machines. Other explored methods are Mixture Models, Bayesian Networks, Random Forests and Shrunken Centroids. Obviously, any published classifier can be tested against sensor array data. However, on many occasions the gain is only marginal compared to classical solutions. An effort is needed to explain the advantages of the latest machine-learning classifiers to the artificial olfaction practitioner.

Support vector machines have been one of the preferred methods for classification in the last years within the chemical sensing community [89] due to its good generalization properties and its robustness against the curse of dimensionality. Pardo & Sberveglieri [90] studied the use of support vector machines (SVMs) to classify electronic nose data, specifically exploring the occurrence of under-fitting and over-fitting effects during this classification. Two binary classification problems of different difficulties were studied. The problems were super-class constructs assembled from measurements made on seven different blends of espresso coffee. They studied the test error in terms of three parameters: 1) the number of principal components of the raw data that are introduced as input to the SVMs; 2) the parameter value for polynomial and radial basis function kernels; 3) the regularization parameter. The authors concluded from this study that the parameter controlling the trade-off between margin size and classification rate of the training data is critical to reduce the test error. They also found that the principal components with small variance (on the easy problem) have an impact on the result. This result is often found in artificial olfaction. On many occasions, first principal components are dominated by unwanted source of variance like drift, environmental perturbations or sampling errors. While this may be interpreted as a lack of robustness, it does not mean that the classifier cannot succeed. On some occasions, the implemented data processing system can effectively reject that unwanted variance and focus on smaller components of variance that have the discriminant information.

Brudzewski et al. used SVMs to classify blends of gasoline [91] and milk [92]. They proposed an extension of the SVM classifier to account for classes not present at the training stage of the SVM. The extension is a classification procedure in two stages. In the first stage, ordinary SVM classifiers (one against all or one against one) are applied and a preliminary winner class is assigned to the observation. In the second stage the Mahalanobis distance $d(\mathbf{x}, \mathbf{m})$ between the observation \mathbf{x} and the mean vector of the winning class \mathbf{m} is computed. If the computed distance exceeds some threshold value, typically determined from the distribution of samples from the particular class, the observation is classified as unknown. This procedure prevents SVM predicting for samples far away from the training data and it will increase system robustness.

Bicego [93] proposed a method with similarity-based representations as a preprocessing stage combined with a SVM classification stage. In the similarity-based paradigm, objects are not described and classified on behalf of explicit features, but by using measures of similarity between pairs of objects. Non-metric similarity measures can very well be used allowing formulation of generic algorithms independent of actual data representations and uniform

feature spaces. In this paper, the author reported that similarity based representations of data often render similarity spaces with very high dimensions. To circumvent this major drawback a new dimensionality reduction approach based on clustering is presented. Data gathered from an electronic nose is used to demonstrate that a similarity-based approach with the suggested dimensionality reduction technique outperforms standard classification approaches, including regular SVMs.

While Support Vector Machines are used for classification purposes, Muezzinoglu et al. [94] proposes a maximum margin formulation for minimizing the empirical regression error and the classification error jointly. The solution yields a pool of support vectors that achieves both goals simultaneously.

Other techniques from the machine learning domain have been also tested with e-nose data. Relevance Vector Machines were proposed to overcome some of the limitations of SVM. They provide a probabilistic output and typically require fewer support vectors. They were tested in e-nose data by Wang et al. [95].

Brahim-Belhouari et al. [96] evaluates the performance of different probability density models for the identification of combustion gases from sensor array data. The compared models were *Gaussian mixture models (GMM)*, *probabilistic PCA mixtures* and *generative topographic mapping* classifiers. The results of this study show that GMM outperforms the other two mixture models provided that PCA is used for pre-processing. Probably this is related to the fact that GMM are very sensitive to the curse of dimensionality. It was also found that the use of mixture models and GMM yielded more accurate classification performance when compared to other classification techniques including K-NN, MLP, and RBF.

Phaisangittisagul et al. [97] presented a classification algorithm to speed up the classification of multiple dataset samples. The algorithm has a different classifier for each dataset and a classifier selection stage that directs each new sample to its corresponding classifier. This is to say that they select the dataset each new sample belongs to. Therefore, only one classifier needs to be computed each time, reducing the computation time. The classifier selection is performed comparing the concatenated response of 15 chemical sensors of a new odor, what they call signature, with the maximum and minimum sensor response values obtained with the training set data. They obtain a maximum and minimum curve for each dataset. The new sample is assigned to the dataset for which the sensor response of the new odor lies within this curve or the distance to them is minimal. The authors validated this algorithm exposing 15 MOX sensors to four different sets of odors: Coffee, Soda, Bacteria, and Rice.

Bayesian Networks were used to combine the classification results for two chemical sensing instruments. Li et al. [98] discriminated between damaged and healthy apples using two instruments: Cyranonose 320 equipped with 32 thin film carbon black polymer sensors and zNose containing one surface acoustic wave sensor. After a PCA feature extraction stage, the sensor responses of both instruments are classified independently by two multi-layer perceptrons with 3 layers. The output layer contains one neuron per class: damaged, healthy. Considering the output value as the probability of a sample belonging to each class, a Bayesian

Network is fed with them to combine the classification of the two MLPs. This stage is only of any use when the classification of both MLPs is different. The results show that error rate decreases from a 20% zNose – 13% Cyranonose to an 11% due to the Bayesian network.

Cho et al [99] uses a Classification and Regression Tree (CART) algorithm [100] to classify Acetone, Nitrotoluene, Nitrobenzene and Octane using a four-component nanosensor array. They obtain very good results (94% classification accuracy) despite the patterns appearing to be highly correlated from the inspection of the scores plot. There is no discussion about the use of CART compared with other pattern classification methods.

Matteo Pardo [101] proposed the use of Random forest and Nearest Shrunken Centroids for the classification of sensor array data. Random forests are ensembles of classification trees. Every tree is constructed using a different bootstrap sample of the data and, in addition, random forests change how the trees are constructed. In standard trees, each node is split using the best split among all variables. In a Random Forest, each node is split using the best among a subset of predictors (features) randomly chosen at that node. Random Forest has only two hyperparameters (the number of variables in the random subset at each node and the number of trees in the forests), and it is usually not very sensitive to their values. On the other hand, Nearest Shrunken Centroids is a modification of the standard Nearest Centroid Classification. The modification consists of shrinking the class centroids toward the overall centroid of the data by an amount called the threshold. It has two advantages, 1) it can make the classifier more accurate by reducing the effect of noisy features and 2) it does automatic feature selection. In particular, if a feature is shrunk to zero for all classes, then it is eliminated from the prediction rule. The paper applies these methods to two datasets originating from food applications. One of them investigates the detection of toxigenic strains in corn and the second the presence of defects in olive oil. The datasets were originated by the commercial EOS-835 e-nose by Sacmi s.c.a.r.l. The system is equipped with five sensors featuring different types of metal oxide sensors. The paper concludes that the performance of Random Forests is similar to that of Support Vector Machines, while the Nearest Shrunken Centroid performs worse. The authors highlight the capabilities of both methods to rank input features, and in fact, results are basically consistent with both techniques.

In the last decade, multilayer perceptrons, simple k-NN classifiers and Linear Discriminants are still used by many teams. There is a shift of thinking towards SVM as the gold standard method for classification purposes. On most occasions, the differences between the different methods are minor. Since in most cases the authors do not report the confidence intervals for the reported classification rates, on many occasions it is not clear if the reported differences are statistically significant.

2.3 Regression

Regression or quantitative prediction models, together with classification, are some of the main final goals of signal processing for chemical sensing. We can define a regressor as a continuous mapping between the sensor response and a certain continuous parameter of the

chemical sample. This parameter is usually concentration, but it could also be another physico-chemical characteristic of the sample. The most common methods to address regression problems are: artificial neural networks (ANN [87]), support vector regression (SVR), and partial least squares regression (PLS).

A number of papers have followed the ANN approach to solve the regression problem. Gulbag & Temuras [102] have compared the performance of artificial feed forward neural networks (FFNNs) with the performance of adaptive neuro-fuzzy inference systems from measurements made on binary gas mixtures with an array of QCM sensors. Six different and well-recognized training algorithms were used to train the FFNNs while the neuro-fuzzy system was trained with a hybrid training method combining back-propagation algorithms with least squares. This comparative study showed that the fastest convergence rates were achieved for FFNNs trained with Quasi-Newton, FFNNs trained with Levenberg-Marquardt, and ANFIS. The mean square error (MSE) was not as good for the ANFIS as it was for the two FFNN approaches considering training data. However, the MSE for the three networks was comparable considering test-data and one of the conclusions were that the generalization performance of the ANFIS structure was better than those of the neural network structures.

B. Mumyakmaz et al. [103] proposed a hybrid algorithm consisting of two neural networks running in parallel. They analyzed sensor responses to ternary mixtures (acetone, chloroform and methanol) from a gas sensor array composed by 6 QCM. One of the neural networks is specifically devoted to the quantification of the chemicals present in the mixture, while the second network identifies the mixture components. The output of this second neural network is a three bits number that codes the components of the mixture. This network provides a kind of filter for the predictions of the first neural network, in the sense that if the second ANN does not detect the presence of a particular chemical in the network, then its predicted concentration is automatically reset to zero. In this way, they correct small residual concentrations always predicted by the first ANN even when the components are not present. This algorithm however is trained only with a number of discrete points in the space of 3D ternary mixtures and it is not clear where validation takes place.

The machine learning formalism known as support vector regression (SVR) was applied in soft sensor applications by Desai et al. [104]. Their paper serves as a good introduction to SVR, emphasizing advantages like strong statistical basis, convergence to global optimum, improved generalization performance and interpretability. It has shown that SVR is an attractive alternative to artificial neural networks for the development of soft sensors.

Some authors used SVR to approach the regression problem. De Vito et al. [49] studied in detail the possibility of using a Time Delay Neural Network and a Time Delay Support Vector Regressor for a scenario of continuous analysis. According to the authors, real time analysis is a very important scenario that has not received proper attention. In real time, rapid variations may occur and responsiveness of control actions is decisive. The regressor system must be capable of correctly taking into account sensor dynamics in order to forecast the real instantaneous concentration, thus enhancing the overall system responsiveness. They explore a scenario with mixtures of NH_3 , NO_2 and humidity. The sensor's response appears to be slow

in following the rapid changes in concentration. They use a time delay of 5 minutes, much smaller than the sensor (or system response). For the Neural Network they explored 5 to 30 hidden neurons, and for the SVR they optimized the C and the hyper-parameter of the kernel function (Gaussian kernels) with a grid procedure. While SVR slightly outperformed the best TD-MLP, the interesting point is that with only a 5 min time window, the system was able to compensate very low system dynamics and give very acceptable predictions even with the presence of a slow drift.

The calibration of a sensor array system is an expensive process requiring access to accurate gas measurement systems. On top of this shortcoming, many chemical sensor techniques suffer from manufacturing variability, making frequent calibration sessions necessary; e.g. for each assembled single sensor array or for each batch of assemblies. To alleviate the trouble and expenses caused by frequent calibrations there is an interest in finding regression techniques that require a minimal number of calibration points while retaining sufficient accuracy. In light of these considerations, Shmilovici et al. [105] compared a special type of support vector regression technique, ν -SVR, with PLSR. The ν -SVR has an explicit ν parameter for controlling the sparseness of the solution and thereby also the number of calibration points needed. By studying data from ternary mixtures of CO, CH₄, C₂H₅OH measured with 12 metal oxide sensors the authors found that ν -SVR indeed generated a smaller mean square error with a smaller number of parameters as compared with PLS. It was however recognized that a current deficiency with the ν -SVR formulation is that each gas requires its own dedicated model. There is absolutely no guarantee that the individual models would pick the same set of support vectors making the number of parameters, and hence the needed calibration point number, to increase with the number of gases to be predicted

Lozano et al. [106] addressed the prediction of wine sensorial descriptors with PLS. They used an array of 16 MOX sensors to analyze 28 wines from different regions, varieties and elaboration processes. All the sensor signals were referenced to the signal from a mixture of ethanol (12% v. and distilled water). Results were good and the validation was made by leave-one-out.

Cho et al. [99] built a zero order Sugeno Fuzzy Inference System on the basis of rules previously learnt by a CART. The tree obtained can be assimilated to the asset of crisp rules. To convert the hard decision rules to fuzzy rules an extended sigmoidal membership function (MF) is defined. This MF has two inner parameters but no clue is given on their optimization. Finally a defuzzication step is carried out to interpolate between the predicted values. As expected, the system interpolates correctly, but unfortunately it seems that the test concentration values are the same ones used for building the tree. Since the algorithm is highly non-linear, it is not clear how it will perform in the presence of mixtures or other interferences.

PLS is still the most commonly used technique for quantitative analysis in chemical sensor arrays, and in most cases is the gold standard to compare other techniques. If the sensors show only mild non-linearities, PLS is usually able to model the relationship by adding additional latent variables. Non-linear PLS (namely, Poly-PLS, Kernel PLS) are also options that have been seldom used for chemical sensors. MLP and SVR are also used in many applications

with good results. Other options, while possible, are not mainstream choices.

3. Robustness of Multivariate Data Processing Models

From our point of view, a more demanding instrumentation technique validation is needed in the field of artificial olfaction for an effective translation of laboratory results to industrial or clinical applications. Technique validation must be made through a robustness study of all aspects of the measurement: sampling, instrumental (sensors, flows) and environmental characteristics.

For a definition of *robustness* (or *ruggedness*) we can refer to the definition given by the WHO (World Health Organization) when referring to analytical techniques in the pharma industry: *“Robustness (or ruggedness) is the ability of the procedure to provide analytical results of acceptable accuracy and precision under a variety of conditions. The results from separate samples are influenced by changes in the operational or environmental conditions. Robustness should be considered during the development phase, and should show the reliability of an analysis when deliberate variations are made in the method parameters”* [107]

A number of inconveniences plague a wider adoption of e-noses for practical uses and are causes of the alleged unreliability, or lack of robustness, of artificial olfaction instruments. Most of the problems arise because, in practical applications, the conditions in which we use the instruments differ from the conditions in which the multivariate calibration model was built. Let us illustrate this point by a number of hypothetical examples:

- (i) **Drift:** While the instruments provide good initial results for many different applications (and this is supported by an enormous number of scientific evidences), these results rapidly deteriorate in time requiring frequent recalibrations. Since calibration of electronic noses is typically time consuming and costly, full recalibrations are not an option. The possible reasons for this fast degradation are numerous, namely: (i) sensors may age due to thermomechanical fatigue, heater drift, or active surface alteration due to high temperature operation, (ii) sensors may get poisoned due to exposure to aggressive chemicals or silicone vapors, (iii) absorption and desorption on tubing and sensor chambers may lead to memory effects, etc [108].
- (ii) **Change in environmental/sampling conditions:** The instrument has been calibrated in laboratory conditions, but in field operation humidity and operation temperature have a large range of variation beyond the calibration range. Changes in the sampling flow will also introduce shifts in dynamic features [109]. The predictions will degrade.
- (iii) **Background change:** The instrument is able to differentiate between two different fungi growing in a variety of cardboard materials. However, in operation, the same instrument is used to differentiate the fungi growing in wallpaper. Predictions

again degrade. Also in this group, we may encounter applications that include a variation of the process (samples) the instrument is monitoring. For instance, in the process a new, previously unknown chemical appears. However, it may be the case that the chemical is already present in the matrix but the mean value changes over time. In general, any change in the mutual statistical distribution of background analytes could produce a degradation of the instrument predictions.

- (iv) **Sensor replacement:** For some unknown reason, one sensor starts providing erratic results and needs replacement. Due to the inherent technological tolerances in the fabrication of chemical sensors of diverse technologies, the new sensor will be different from the initial one. There will be changes in baseline and in sensor profile sensitivity to the compounds of interest. This will lead to a distortion of the original multivariate pdfs (probability density functions) in the feature space dimensions that depend on the response of the replaced sensor.
- (v) **Calibration transfer (a):** In this case, an e-nose company has developed a pattern recognition system for an instrument (the master) but wants to sell it for use in the whole series of apparently identical instruments. Without individual calibration, direct application of the master model to the remaining instruments causes degraded performance. Obviously, and again due to the differences between the instruments, the secondary instruments test set cannot be considered to be extracted from the same probability density function that characterized the master instrument.
- (vi) **Calibration transfer (b):** A Spanish Company develops an application to automatically detect the onset of ripening in apples "Verde Doncella" when stored in climacteric chambers using an electronic nose from a certain vendor. The Spanish company has a delegation in Portugal and wants to transfer the instrument and its calibration model to the local production of Apples in Portugal. Direct application of the model originally developed with Spanish apples caused degraded performance in Portuguese apples of the same variety.
- (vii) **Prior probabilities change:** An e-nose instrument has been trained to differentiate lung cancer subjects from healthy subjects in a certain clinical study where the training set was balanced for both conditions. However, in practical applications the prior probability of the patient's condition is unknown or has seasonal changes.

From a data processing perspective, all these examples can be summarized as a shift in the statistics of the test set compared to the original training set. It is important to realize that this simple statement captures many different problems that appear in the practical application of e-nose instruments, as we have seen in the examples above.

When applying a prediction model, either quantitative or qualitative, it is always prudent to ensure that the new data does not fall outside the statistics of the training data. In some cases, users have a kind of blind faith that the prediction model will magically extrapolate outside the range of the training data. This could be argued for the class of linear instruments modeled

with linear models, but it cannot be assumed in general. It is less known that multivariate data processing predictive models embed other variance onto the characteristics of the background, the instrument and the environmental conditions. Those are implicitly calibrated effects, even if the model does not provide any explicit output for them. Then models could fail or degrade if they are applied in conditions that depart from the implicitly calibrated conditions [110]. Far from the calibration conditions, predictions will be biased. In summary, sources of variation for the new data that did not appear in calibration conditions will invalidate the calibration model.

Despite those important issues that hinder e-nose usability, pattern recognition and chemometrics textbooks usually present the different techniques by ignoring these issues. It is always presumed that the test set and the training set match. In other words, they assume both sets have been extracted i.i.d (independent identically distributed) from the same generative probability density function. In real world conditions, this is not the case. Additionally, these kinds of problems are not usually addressed by the bulk of e-nose literature. While we found a large number of successful applications reported, in the great majority of cases the validation procedure chosen is based in cross-validation (leave-one-out, k-fold, bootstrap, etc). These validation procedures implicitly assume the i.i.d. hypothesis and totally neglect the fact that, in most practical applications, we cannot assume that the training set and the test set have been drawn from the same joint probability distribution.

In fact, a more rigorous validation technique will certainly give more confidence on the robustness of the proposed methods. Due to costly measurement procedures, training samples are usually scarce. Small sample size and high feature vector dimensionality can lead to overfitting in many learning algorithms. Additionally, in most occasions algorithms have free parameters that control the complexity of the learning machines. Those free parameters (number of Principal Components in PCA, latent variables in PLS regression, number of hidden neurons in a Multilayer Perceptron) are typically adjusted in an inner cross-validation loop. We advocate here for the use of a fully independent test set, free of any parameter optimization task. Due to time-varying statistics, it is much better if the test set is future to the training set.

At this point we would like to point out some methodological errors that mostly appear in leave-one-out (LOO) cross-validation procedures. The interpretation of this method depends on what instance is left-out. This instance left out should reflect the expected operational scenario for the instrument and the expected generalization characteristics of the system. For instance, if the LOO procedure leaves out a single measurement, and keeps in the training set replicate measurements of the same sample, we should only expect that the system will generalize if it has replicate measurements in the training set and this is a very limiting condition. In a similar case, if the LOO procedure leaves out a particular wine, but leaves in the training set samples from the same wine variety, we cannot expect that the learning system generalizes well for other varietal wines.

While the test set statistics may be different from the training set, it is clear that there is preservation in the data structure. That is, even if shifted, the test set somehow resembles the training set. To avoid full re-calibration, methods to improve stability should model this distortion in the data structure with a minimum of, and eventually no, additional

measurements. This can be framed under the general concept of transfer learning, that is, how to transfer information from the particular conditions under which a calibration model was built to do prediction in new conditions [111].

As we will see later, many methods rely on dedicated experiments to correct the observed data shift (so-called explicit or extrinsic methods). Ultimately, the use of a calibrant or set of calibrants to retrain the data processing unit may be the most robust method for correcting the degradation of the prediction model. Obviously, a minimum set of calibrants is desired for cost and time reasons, and a number of techniques with diverse levels of complexity are at our disposal to update the calibration model on the basis of the new calibrants. Other techniques try to model undesired variance directly from the calibration set. For this later approach to be successful, it is essential that the calibration conditions contain all the expected sources of variability to be found in the posterior instrument operation phase.

Data shift correction and model update can be carried out using very diverse strategies that we will present in the following subsections. The instruments can be made more robust by signal pre-processing, by selecting robust features, by univariate recalibration, by multivariate strategies or by adaptive methods.

3.1 Sensor array linear response model

To introduce the rationale of these methods, we will present a linearized model of the array response in a quantitative multi-component analysis scenario adapted from Gujral et al. [112]. Let \mathbf{x} be the feature vector of dimension n_s and \mathbf{c} the n_c dimensional concentration vector, where n_c is the number of species in the sample the sensors respond to. On many occasions n_c could be very large and unknown. The linearized sensor array response model is:

$$\mathbf{x}^T = \mathbf{c}^T \mathbf{S} + \mathbf{n}^T \quad (4)$$

Where \mathbf{S} is the $n_c \times n_s$ pure component response matrix and \mathbf{n} is a noise term (typically assumed to be zero mean). It is important to understand that in most cases, the objective is to determine only n_k analytes of interest, where $n_k \ll n_c$. In addition we may have an additive drift term \mathbf{d} (including here also environmental perturbations). We assume that the random vector \mathbf{d} lies in a low-rank subspace. In this sense we can write:

$$\mathbf{x}^T = \mathbf{c}_k^T \mathbf{S}_k + \mathbf{c}_{int}^T \mathbf{S}_{int} + \mathbf{d} + \mathbf{n}^T \quad (5)$$

Where \mathbf{c}_{int} is the concentration vector of the interferent substances and \mathbf{S}_{int} their pure responses. It is a typical scenario that the analytes of interest are systematically controlled during the calibration phase but the values of the interferent concentrations remain largely unknown. It is also commonplace that in test phase there could appear interferent substances that were absent during the calibration phase. In this simple model, the uncontrolled variation of interferent substances will produce drift-like responses.

If linear additive drift is present during calibration, inverse calibration models already provide drift compensation [113]. However, it can easily be the case that the drift term in operation

has components in subspaces orthogonal to the drift subspace seen in calibration phase. Those components will not be compensated by the inverse calibration model.

In order to clarify the purpose of the different techniques for robustness improvement, let us now consider the inverse prediction model using a linear multivariate model (a regression problem)

$$\hat{C} = X\hat{b} + E \quad (6)$$

Where \hat{C} are the predicted variables (typically analyte concentrations, but they could be other sample characteristics), X are the feature vector, \hat{b} is the regression vector and E is the matrix of residuals. This matrix of residuals will have at least three components: (i) errors due to noise in operation phase, (ii) modeling errors related to the estimation of the regression vector b (e.g. due to model inadequacy or noise-drift during calibration), (iii) errors due to unseen drift in calibration phase.

\hat{b} can be found by ordinary least squares (OLS):

$$\hat{b} = (X_{cal}^T X_{cal})^{-1} X_{cal}^T C_{cal} \quad (7)$$

Where X_{cal} are the sensor features in calibration phase and C_{cal} are the reference analyte concentrations in calibration. At this point, is worthwhile to mention that in the presence of collinearities among the sensor features the matrix $X_{cal}^T X_{cal}$ could be ill-conditioned. This is easily checked by computing the condition number. Ill-conditioning in general provides regression vectors with large modules. Solutions for ill-conditioned regression problems can be based on projection methods like PLS or PCR, or shrinkage methods like Ridge Regression. However, ill-conditioning is not always present, and in those cases the OLS solution may be enough.

The changes referred above in its various forms, produce a shift in the test set when compared to the training (calibration) set. Those changes can be thought of as a perturbation δx . This perturbation produces also an error in the predicted values [114].

$$\delta \hat{C} = \delta x^T b \quad (8)$$

$$|\delta \hat{C}| = |\delta x| |b| \cos(\delta x, b)$$

Therefore we can conclude that to minimize the error in prediction, one or more of the three terms in the right hand of equation must be minimized. We will say that a feature vector is robust when the different perturbations described above produce minimum changes in the feature vector δx . A second strategy is to minimize the norm of the regression vector. In this sense, regularized solution like Ridge Regression (RR) will provide in general more robust solutions than plain ordinary least square (OLS). Finally, we can try to build models that are particularly robust against perturbations. In the simple model above those models will provide regression vectors that are orthogonal to the perturbations.

3.2 Examples of dataset shift handling in the chemical sensing literature

An obvious strategy to make models more robust is to design calibration experiments in such a way as to include all expected sources of variance trying to obtain a truly global model (in contrast to models local to particular calibration conditions). For instance, by including changes in operating temperature and humidity [115] under the calibration conditions, or changing backgrounds.

For example, M. Kuske et al. [116], when trying to detect moulds growing indoors, used a variety of building materials to make the model more robust to changes in the background. B. Raman [64,117] uses different background interferences to test for the robustness of the hazard identification despite the presence of varying backgrounds (bleach, house paint and window cleaner). The same group tested the capability of their system to detect traces of acetone in a varying background that simulated inhaled and exhaled breath. The simulated breath was however clean of other impurities beyond acetone [118]. Mumyakmaz et al. [119] when trying to predict toluene with a QMB sensor array, included measurements of toluene in a full range of relative humidity values. Similarly, it is possible to build a sensor model depending on analyte concentration but also on temperature and humidity if sufficient training data is available to cover the full range. For many sensor technologies, this extended model becomes highly non-linear. To cope with this, Aliaghasarghamish et al. [120] proposes to use a Recursive Least Squares Method to train a Fuzzy Inference System that outputs a compensated resistance value that is then inverted to predict the concentration of a single analyte (methanol in this case). Instead of producing a direct model and then inversion, Hajmirzaheydarali [121] builds an inverse model to predict methanol concentration, using a multilayer perceptron that uses temperature, humidity and sensor resistance as inputs. With a similar strategy, one option is to train the classifier/regressor with data that already contains substantial drift [122]. While drifted data remains within the range seen in calibration this could be a good option. However, if the sensors readings depart from this explored area in the input space the generalization capabilities of the pattern recognition system remain unclear.

A very interesting approach is to calibrate the system directly in field conditions. This approach has been followed by De Vito et al. [123,124] when trying to predict atmospheric pollutants in urban scenarios. The presented work leads to very interesting conclusions relevant to dataset shift problems in chemical sensing. In particular, the study aimed at the prediction of Benzene, CO, NO₂ and NO_x. The reference measurements were obtained by a certified pollution test station based on the state-of-the-art spectrometers. The testing campaign lasted for one year (March 2004 to April 2005). The sensor array consisted of 5 MOX sensors plus temperature and humidity sensors. The predictor was based on a Multilayer perceptron with regularization based on Automatic Bayesian Regularization [125]. The scenario was characterized by very high correlations among the target species. De Vito showed that typically a 15 days calibration (360 hourly averaged measurements) was enough to have satisfactory performance for about 6 months. From that time on, prediction quality degraded. Calibration learned during spring time was not good for predicting pollutant levels in winter time. In fact, De Vito interprets that the performance achieved in spring-summer is due to an intrinsic correlation between the pollutant levels. However, the arrival of winter changes the pollutants correlation and

calibration models learned in spring become obsolete. In fact, recalibration at that time greatly improves performance. This is a very nice example of how models could be local to apparently hidden conditions (correlations among the pollutants), and on the other hand, it shows that the path is to calibrate the system in the exact conditions it is expected to operate. In this sense, De Vito raises doubts about the direct application of learned models in other Italian cities.

Pollution monitoring, and other large area monitoring applications, will benefit from advances in MEMS technology and wireless communications. Today, environmental test stations based in spectrometers provide accurate concentration readings for a very limited set of compounds. However, due to their high costs their placement is very sparse. Due to local variations and canyon effects, especially within urban scenarios, pollution levels can vary drastically over short distances. To have a better representation of the spatial distribution of pollutants, low cost mobile chemical sensors will be desirable. Since their readings lack the required stability, Tsujita [126,127] proposed an auto-calibration method where each sensor is recalibrated when in proximity to the reference test stations, or within proximity to a second mobile sensor. Through a simulation study they show that this mechanism can propagate the stability of the reference stations through the network.

In a few selected cases, the physico-chemical origin of the drift is known. For instance, for micro hotplates with polysilicon resistors, it is known that the resistance value and even the temperature coefficient of resistance can drift in time. If the device is operated in constant current or voltage conditions, heater drift will change the operation temperature and in consequence this will be a cause of sensor drift. Rogers and Semancik show how a constant power operation of the micro-heater improves the stability of metal-oxide microhotplate chemical sensors [128].

3.3 Robustness Improvement by Pre-processing

An option to overcome the stability problems mentioned initially is to apply some data pre-processing or feature extraction/selection strategy, to minimize those effects in the ulterior classification or regression models. An example is the use of rank features [129]. Obviously perturbations that do not produce a change in the ordering of the sensor responses do not change the pattern. These methods provide robustness at the expense of reducing the resolution of the system.

There are a number of techniques that provide some robustness improvement and that are routinely used by most researchers: among them, normalization and autoscaling.

Normalization is mostly used for rejection of common mode perturbations and multiplicative effects, typically pattern scattering due to changes in concentration, but also instabilities in sampling and changes in environmental parameters. A variety of normalization rules have been proposed:

- (i) normalization to unit norm:
$$x_i^{new} = x_i / \sqrt{\sum_i x_i^2}$$
- (ii) normalization to unit area:
$$x_i^{new} = x_i / \sum_i x_i$$
- (iii) Standard Normal Variate (SNV): Each pattern is centered and then scaled by its standard deviation. Similarly, Robust Normal Variate (RNV), proposed by Guo et al. [130] is a robust version of SNV less sensitive to extreme values in the pattern.

Solid-State Sensors are very prone to suffer from baseline shifts. Periodic presentation of a reference substance, typically zero-air, allows the evolution of the baseline to be estimated. Baseline variations can be corrected by SNV, or baseline models can be directly subtracted from the sensor response.

B. Raman et al. [117] propose an interesting approach to the automatic detection of the onset/offset of gas presentation. They use temperature programmed metal-oxide microhotplates. They computed the correlation between sensor responses at different temperatures using a moving window approach and they found that the correlation is maximal when there is a fast variation of the gas composition in the sensor chamber. Once this onset has been detected, the sensors are referenced in a univariate manner to the most recently tagged baseline.

A number of authors have proposed the use of linear or exponential models for the evolution of the baseline in time. A more elaborated estimation of the baseline in chemical sensing is proposed by Wenzel et al. [131]. Their proposal uses the Kalman Filter formalism for an on-line estimation of the baseline evolution. In their approach, they use a first order model for the sensor dynamics. Since a number of parameters are unknown they propose a five dimensional state vector that contains: 1) the analyte concentration, 2) the baseline value, 3) the baseline drift rate, 4) the sensor steady-state response and 5) the sensor time-constant. The final state-variable model results in a non-linear model with non-Gaussian variables. To estimate the state-vector they propose to use a bank of Extended Kalman Filters that are linearly combined using a Gaussian Sum Filter. They test this approach with synthetic signals and a limited number of experimental responses from a SAW sensor. It is important to note that the model assumes the periodic presentation of an analyte of unknown concentration and a flush/reference gas. It is very interesting that the model allows the baseline to be estimated even when short flush times prevent the baseline from fully recovering. This is a common problem in chemical gas sensor systems, as in many occasions the baseline recovery is extremely slow, greatly reducing the system throughput. While the approach seems very interesting, more extensive testing is needed to prove its validity.

Pre-processing steps based on filtering can also provide more stable responses. Since drift is typically low frequency, Zuppa [132] et al. and previously Llobet [133] proposed filtering sensor responses using a Discrete Wavelet Transform (DWT). Elimination of the low frequency terms rejects drift and improves class separability. The use of the DWT for de-noising using soft-thresholding initially proposed by Donoho [134] has shown to be very effective in sensor signals by Huang and Leung [135].

If response shifts are caused by a known interferent, one possibility is to measure it with a dedicated sensor and compensate the responses of the main unit. If the responses to the analyte of interest and the measurement of the interferent have different time dynamics, then it may be necessary to estimate a dynamic model due to the effect of the interferent on the main sensor signals. In [2], an adaptive linear filter based on the LMS algorithm (least mean squares) was used to compensate the variations in humidity when a chemical sensor array was working inside an industrial fridge. It is important to remark that in many scenarios the interferents are too many, mostly unknown, or there is not a readily available sensor to measure them.

3.4 Univariate approaches for Stability Improvement

Many researchers advocate the use of univariate approaches [136] to datashift correction for simplicity reasons. In some cases, the stability improvement is directly aimed at compensating environmental factors, namely temperature and humidity. In this sense, Hossein and Ghafarinia [137] explore the dependence of a tin dioxide sensor to measure methanol in a range of temperatures (23-43 °C) and humidities (10-80 r.h.). They propose to use an Artificial Neural Network that uses a time-delay input on the sensor resistance but also on independent measurements of the temperature and humidity, to predict the analyte concentration.

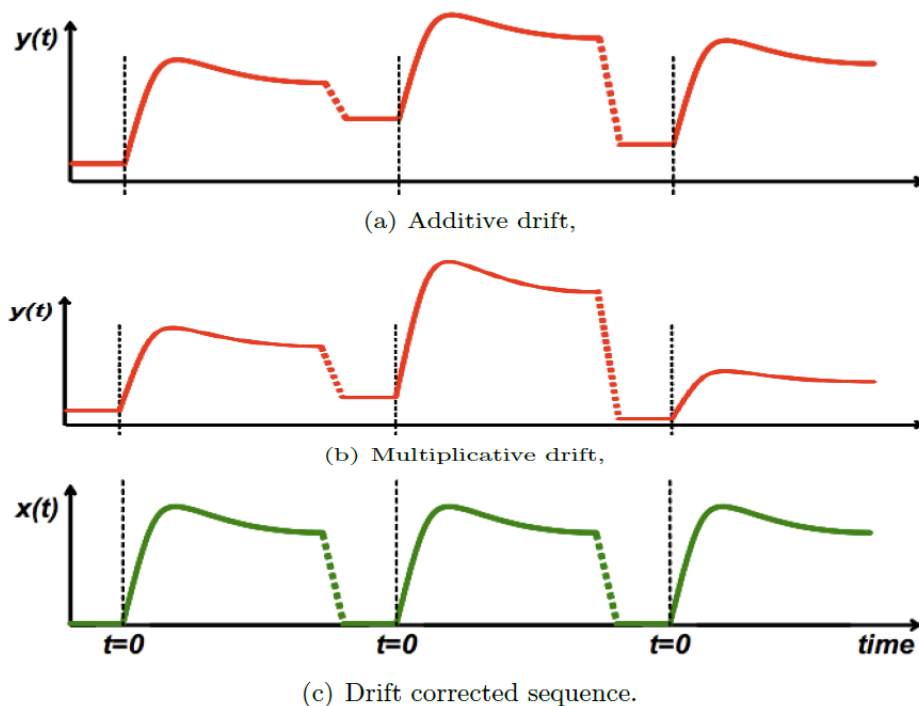


Figure 6. Illustration of additive and multiplicative drifts on a sensor transient response (courtesy of R. Gutierrez-Osuna).

Directed to drift compensation, Sisk and Lewis [138] explore the use of preprocessing steps like baseline correction and normalization, followed by univariate multiplicative recalibration using a single calibrant. In the framework of their experiment, this was enough to retain classification accuracy. The paper from Sisk and Lewis provides a drift analysis based on the

study of a dataset of 15 compositionally distinct carbon black-polymer composite chemoresistors. The analytes under study were five hydrocarbons (cyclohexane, n-hexane, n-heptane, n-octane and isooctane) and three other analytes (tetrahydrofuran, ethanol and ethyl acetate). The dataset contains five runs. In each run, the eight analytes are presented to the sensor array in random order. The dataset contains a number of time discontinuities (from 7 days to 70 days), that is; periods of time where the measurement activities stopped. It was observed that those measurement discontinuities in data collection are a major source of drift and shifts. For instance, sensor responses were clearly affected by a 2-month break, while some sensors showed transient spikes after shorter breaks before attaining a new stable region of responses. An interesting conclusion of their study is that the best correction performance is obtained when the calibrant is chemically similar to problem substances. Additionally, their paper investigates a number of things of practical importance such as the optimum calibration frequency or the trade-off between the sizes of the recalibration set and the test (real use) set. A recommendation of their work is to implement an adaptive recalibration scheme based on the departure of the actual sensor readings from the ranges observed in the calibration phase, instead of fixed calibration periodicity.

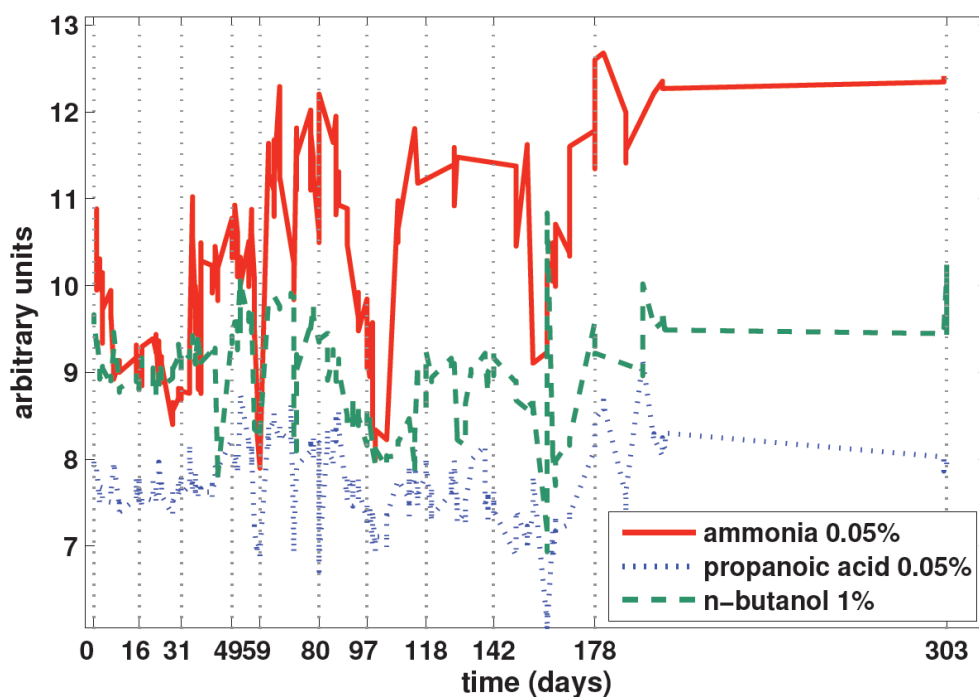


Figure 7. Example of sensor drift: erratic sensor readings in identical measurement conditions (Manchester dataset).

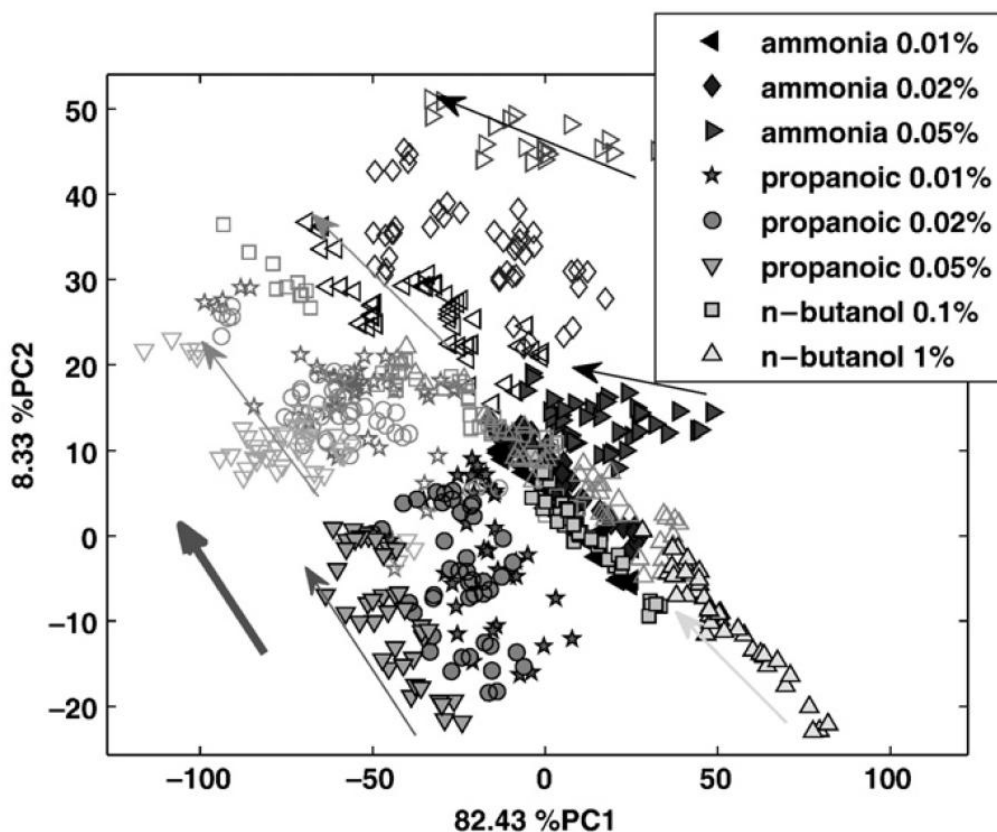


Figure 8. Principal Component Analysis of long term measurements with a conducting polymer sensor array (Manchester dataset). The arrows indicate the direction of the pattern shift for every class.

Similar approaches to univariate shift correction have been proposed by Romain and Nicolas [136] and Tomic et al. [139]. In all those cases, it seems that the drift can be decomposed in an additive term that is compensated by the baseline correction, and a gain drift that is compensated in a multiplicative manner, taking into account the response to the calibrant. Frank et al. [140] use a set of 8 calibrants every day and calculate the average sensor value. The sensor signals to unknown samples are referenced in a multiplicative manner to this average value to the calibrants. Tomic et al. also propose a multiplicative correction method. However, they additionally propose that this correction factor may be fitted with a straight regression line to interpolate the correction factor between calibration factors. Sisk and Lewis [138] note that this procedure does not work properly in extrapolation mode.

An interesting approach to univariate drift correction has been proposed by Huang and Leung [135]. They combine an initial pre-processing step of de-noising using DWT, followed by drift estimation and correction using the Papoulis-Gerchberg Method. At the core of the engine we found the hypothesis that the sensor signals can be adequately described by a Lorentzian model as originally proposed by Carmel et al. [56]. The method uses non-linear model fitting to estimate drift that is additionally forced to be a low-pass signal.

3.5 Multivariate Orthogonal Projection Methods

Orthogonal Projection Methods intend to estimate the subspace that contains the undesired variance. There are many different flavors of the basic idea: Net Analyte Signal (NAS) [141], Orthogonal Signal Correction (OSC) [142], Direct OSC [143], Direct Orthogonalization (DO) [144], orthogonal projection to latent structures (OPLS [145] and O2PLS [146]), Kernel OPLS [147], Projected OSC, Piecewise OSC [148], external parameter orthogonalization (EPO) [149], error removal by orthogonal subtraction (EROS) [150], Orthogonal Discriminant Projection (ODP) [151], Dynamic Orthogonal Projection (DOP) [152], etc. In the field of Artificial Olfaction, proposals by Artursson: Component Correction [153] and R. Gutierrez-Osuna [154], belong to this family of techniques.

The main action of OP methods is a reduction of the data dimensionality. OP techniques search for dimensions of the subspace that describe maximum variance related or unrelated to the information of interest. As mentioned before, the intention is to make the posterior prediction of the model independent of the influence of these non-desired variations in the data and thus improving data and model interpretation. They are maximally effective when the dataset shift occurs in a subspace of small dimensionality. Problems arise when this subspace contains also the discriminant information. In those cases, an optimum trade-off has to be found.

In general, these methods are based on extracting from the multivariate measurement space D , the subspace not related to the components of interest Y . For this purpose, different approaches have been developed and there are review works comparing them [155], [156], [157]. Orthogonal projection methods identify an orthonormal basis P capturing the maximum of systematic variations of the drift-like term:

$$d_{tot} = c_{int}^T S_{int} + d \quad (9)$$

and then project X onto the subspace orthogonal to P .

$$X^* = X [I - P P^T] \quad (10)$$

Alternatively, Generalized Least Square Methods (GLS) [158] estimate the drift subspace and then shrink the regression vector in the direction of the drift subspace.

Different algorithms identify P with different principles. Extrinsic, also named explicit, methods use dedicated experiments to identify P . Those dedicated experiments may contain time delayed experiments to estimate time-drift, variations of the environmental variables, changes in the chemical background, sensor replacement, calibration transfer, etc. Extrinsic methods are for instance CC, EPO, EROS or IIR [159] (Independent Interference reduction). Extrinsic methods have the advantage of not requiring the Y information but the disadvantage of requiring dedicated experiments. On the other hand, intrinsic methods identify P directly from X . For instance doing a PCA analysis of X and then orthogonalizing the scores to Y .

Component Correction originally proposed by Artursson [153] has been the most popular OP method in the Artificial Olfaction Community. In fact, it can be considered as a popular benchmark for multivariate drift correction methods.

Tomic et al. [160] explore drift correction in a commercial electronic nose NST3320 (AppliedSensor, Sweden) for measurements of milk separated in time 2 years. They aimed at using the old model with the new data. They use of uncorrected data produced drastic failure with classification rates extremely low. In the paper, they explore the performance of Multiplicative Drift Correction (MDC), Component Correction (CC) and combinations of both methods. The results reported there-in showed that the results greatly improved with all the methods.

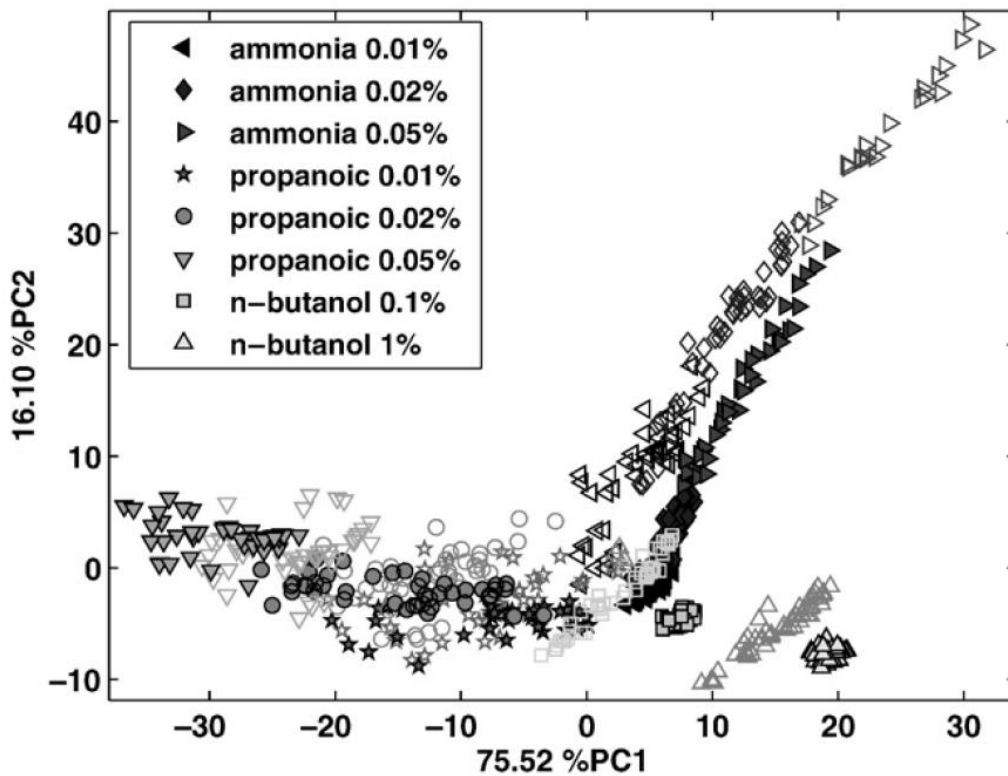


Figure 9. Principal Component Plot of the long term Manchester dataset corrected by Artursson's Component Correction. Calibration set (solid symbols), last test set (non-solid symbols) The reference class (n-butanol 1%) appears extremely compact, while the drift correction is less effective for the other classes.

OSC has been applied as a drift counteraction method by Padilla et al. [161]. In her work Padilla uses a dataset provided by the University of Manchester: 17 Conductive Polymers operating during 10 months for a total of 3415 measurements for three analytes at different concentration levels. Padilla calibrates the system for a classification task at the beginning and performs a validation in the future with a moving window for testing. Without treatment classification performance degrades in time due to uncontrolled random variation in the sensor readings. OSC boost classification performance for about 15% (average). This is clearly seen in the scatter plot where raw class clusters become elongated in the drift direction, while after treatment clusters become more compact with a better class separability (see figure 9). Padilla's studies also show dependence on training set size. Results improve with the size of the training set size but they stabilize after 10 days.

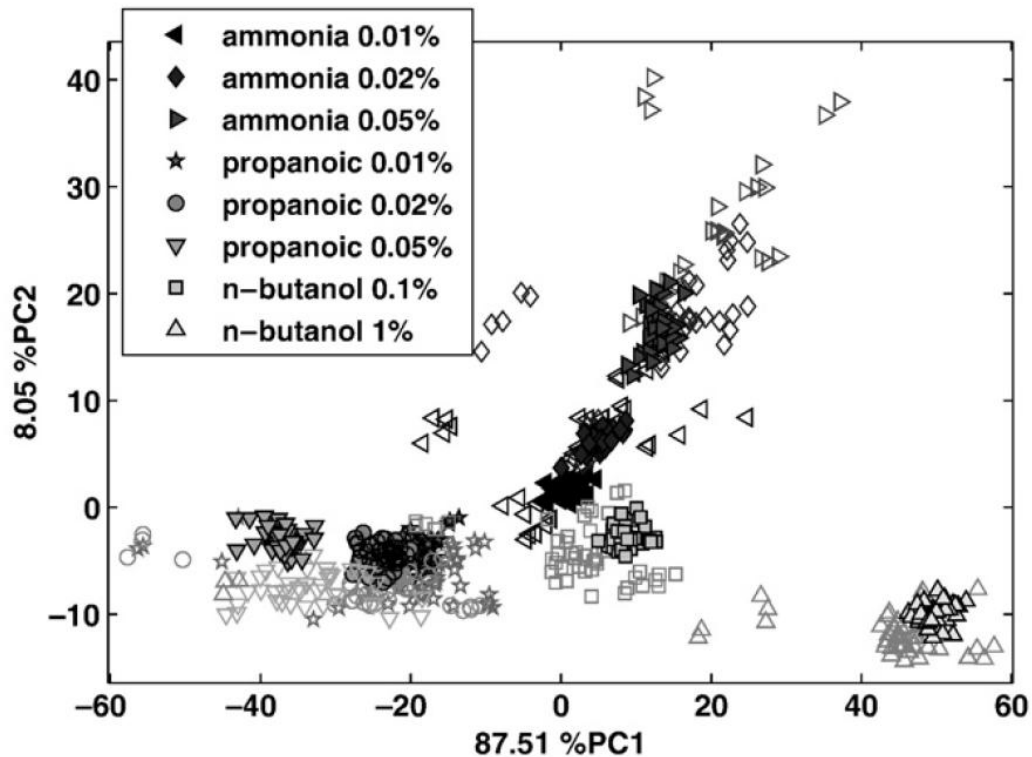


Figure 10. Principal Component scores of the calibration set (solid symbols) and the last set (non-solid symbols) after orthogonal signal correction in the Manchester dataset.

An improved CC method has been presented recently by Zityadinov et al. [162]. In the traditional version of CC, a single reference class is used to find P^* . This extrinsic method has the advantage of simplicity but it has a certain dependence on the chosen class for counteraction (see figure 10). CC assumes that all classes shift in the same direction in the input multivariable space. However, on many occasions it has been found that different classes shift in different directions. For maximum performance, the reference class should present a shift in a subspace that contains, as much as possible, shifts of other classes. Instead, Zityadinov proposes to use Common Principal Component Analysis to use all classes simultaneously to determine the subspace that contains shift directions. The technique is maximally efficient if the experimental design automatically provides data for all classes in the perturbed condition.

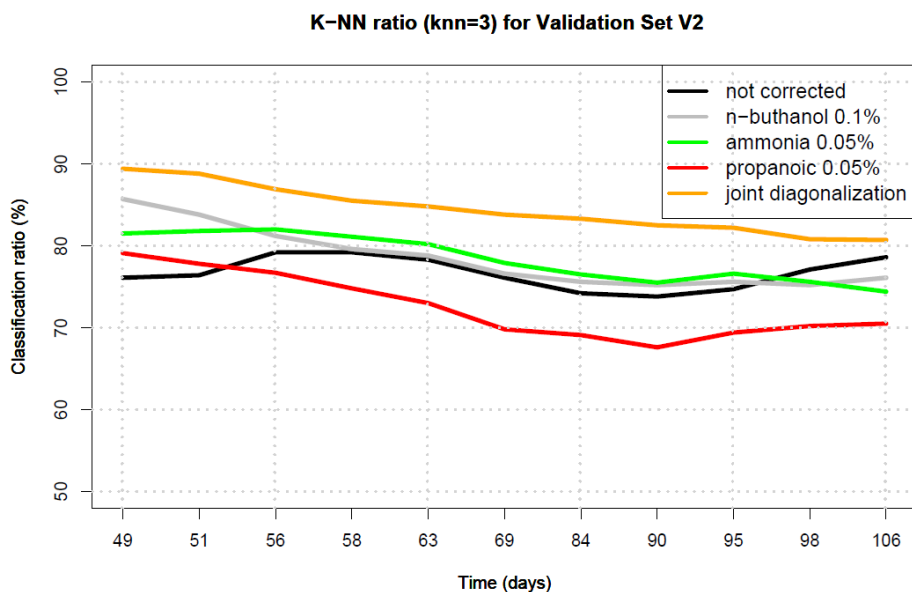


Figure 11. Classification rate vs. the time distance between the training set and the test set in the Manchester dataset: Comparison of Artursson Component Correction (CC) using three different reference classes and Joint Diagonalization using Common Principal Component Analysis. For some classes, CC even degrades the instrument performance. In this example, common PCA outperforms CC.

3.6 Robustness Improvement by Feature Selection

A robust model can be built if we identify features that are more stable in time, or less sensitive to variations in the instrumental or environmental conditions. A rich literature has been developed over the years on Feature Selection for Artificial Olfaction (see section 2.1). Feature Selection selects a subsets of features based on a certain objective function. Proper definition of this objective function can lead to select features which are particularly stable and keep prediction accuracy over time. This application is straightforward if the training data already contains a substantial amount of shift. An example of this approach is the contribution by Lazzerini et al. [163]. In their paper they propose to use a partially supervised Fuzzy clustering algorithm [164] followed by a Fuzzy k-NN classifier. However, instead of using a plain Euclidean distance, they introduce a diagonal distance metric, that is a weighted distance, and in the clustering step they optimize not only the cluster prototypes but also the weights. Their hypothesis is that features that degrade the cluster quality will be discarded, resulting in more compact clusters and better classification results.

3.7 Adaptive Correction Methods

Most Adaptive Correction Methods are based in the intuitive idea of trying to follow clusters as they shift in feature space. They are particularly interesting when the dimensionality of the drift space is large or when the drift occurs in the same subspace that contains the discriminant information.

Adaptive Correction based on Self-Organizing Maps was advocated in the late 90s by Holmberg [165] and Marco [166]. Later, Distanto proposed a modification based on multiple SOMs: each one specializing in a single class or cluster [167].

While the proposals above adapt the classifier to track the drifting patterns, recently Di Carlo et al. [168] have proposed a method to adaptively correct the input patterns to make them closer to the class centroid in the training set. The correction is linear and the coefficients of the matrix adapt automatically using an optimization method: covariance matrix adaptation evolution strategy (CMA-ES). The goal function in this optimization is to minimize the cumulative distance between the patterns and training set centroids over a moving time window. The paper explores several distance metrics. The authors explore the methodology in synthetic and real data with very good results.

Kim et al. have proposed a scheme for drift correction based on Radial Basis Functions (RBF). The network is fully trained using a stochastic optimization algorithm: centers, widths and weights are optimized with calibration data. For datashift correction a dedicated set of experiments is needed. They claim that simple adaptation of the network weights keeping centers and widths constants could be enough. For weight adaptation two schemes are proposed in the framework of the Information Learning Theory (ILT) proposed by Principe [169]. The first one maximizes the similarity between the pdfs in output space using a Euclidean norm [170]. In a second paper they propose a similar idea but using correntropy as an optimization goal [171].

An interesting proposal in adaptive drift correction was introduced by Hui et al. [172]. As in the previous work, they use an adaptive PCA model for event detection. However, the event is defined as a sensor drift onset. As a consequence the feature vector slowly departs from the PCA model. In the field of failure detection [173], it is known that sensor drift is one of the most difficult failure detection problems. This is basically due to the fact that the event shows up slowly in the sensor signals. In fact, their approach is very similar to the multi-scale PCA fault diagnosis methodology of Misra et al. [174], where the event detection is carried out based on PCA models computed on the coefficients of the DWT. The proposed methodology not only detects the onset of drift but also allows identification of which sensor is drifting. At this point, Hui proposed to correct the wavelet coefficient that departs from the PCA model with a model linear in time, and then reconstruct the sensor signals. Their study shows that this approach is much more sensitive than applying an adaptive PCA model directly on the time signals. This improvement is clearly due to the previous multiresolution analysis. In their experimental dataset, they prove that the strategy is able to detect and correct the sequential onset of two drifting sensors.

In the more general context of process monitoring, Angelov et al. [175] proposes the use of Evolving Fuzzy Takagi-Sugeno systems in the presence of drift when the calibration samples are available periodically in batches. In his proposed technique, not only can the parameters of the system be tuned during re-calibration, but the model structure can also evolve: rules may appear and disappear automatically to tune the inference system to the new conditions. His proposal consists more of adapting an existing model, instead of fully recalibrating the model with the newer training samples. For details of the technique the reader is referred to [176].

3.8 Background Suppression

An adaptive approach to event detection, also named novelty detection, was proposed by A. Perera et al. [177]. In their problem application, an oil leakage had to be detected while the sensors suffered severe drift. Their approach is based in Recursive Dynamic Principal Component Analysis using a moving window. In this way, the PCA subspace model captures the normal variability of the sensor signals due to time-varying perturbances and drift. The moving DPCA model is used to calculate the Q^2 statistics for the next sample. This approach provides more sensitivity and faster detection of changes than the related approach based on Evolving Window Factor Analysis (EWFA) [178].

One possible way of approaching the problem of mixture analysis is by removing the effect of the interferent gases to focus on the target analyte. This approach is referred to in the literature as background suppression. Gutierrez-Osuna et al. [179] proposed two signal-processing methods to perform background suppression in the multivariate response of a chemical sensor array. The first method is based on a committee machine of linear discriminant functions, each trained to discriminate a single analyte. The elimination of the background is achieved by reducing the weight of those linear discriminant functions that show high sensitivity to previously detected odors. The second method is based on the proposed approach of TP-LDA (topology preserving linear discriminant functions) that performs an LDA-based projection, ensuring that the natural topological ordering of analytes and mixtures is preserved. Background suppression is then obtained considering the background odor as belonging to the same class as the neutral odor (water) when computing the LDA projection. The same authors proposed [180] a generalization of the second method where the adaptation is not only to single but also to mixed backgrounds.

Following a biologically inspired approach, Gutierrez-Osuna et al. [181] used the KIII, a neural population model of the olfactory system, to remove background odors from sensor array signals. They mimicked odor habituation by depressing the mitral connections to other neural populations within the KIII model. This depression mechanism is triggered locally in each mitral cell and mediates the adaptation of the system to background odors previously presented. Since the performance of the habituation method was hindered by sensor correlation, Gutierrez-Galvez et al. [182] combined habituation with a contrast enhancement mechanism to reduce sensor correlation. Both mechanisms were implemented in the KIII model improving the ability of the system to remove background odors (Figure 12).

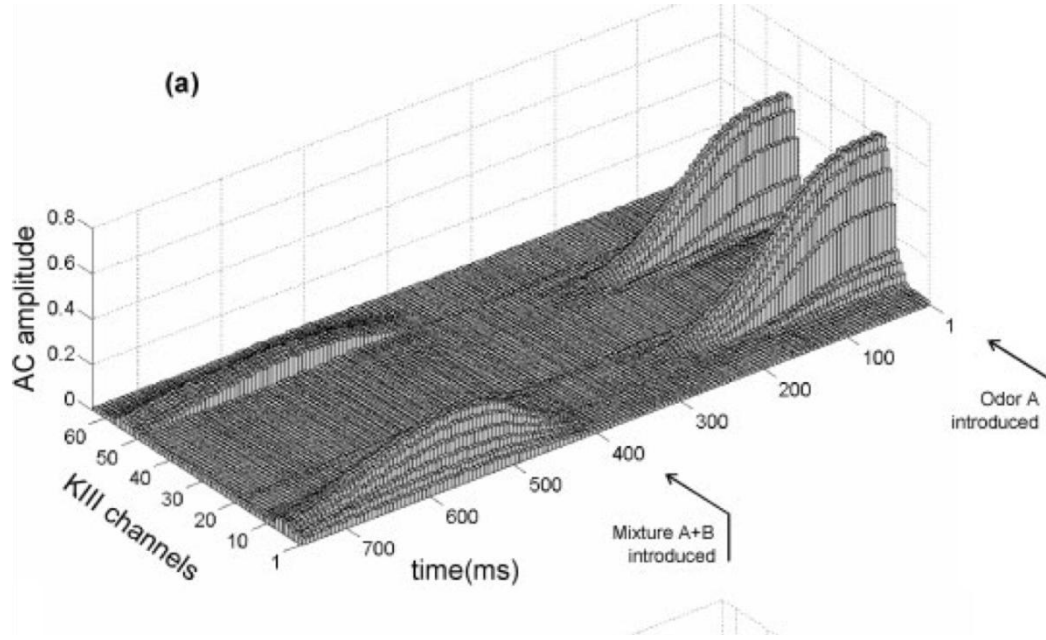


Figure 12. Habituation scenario with gas sensor response patterns. Odor A is introduced at $t = 1$ ms. Following habituation of odor A, the Kill is presented with a 50/50 mixture of odors A and B. (a) Kill model trained with the hebbian/anti-hebbian rule: The system response to mixture A + B is a contrast-enhanced version of pattern B alone. (Time axes are mirrored to help visualize the time evolution of the channels.)

More recently, Feng et al. [183] proposed a method based on wavelet transform to remove the effect of the natural odor of mice in order to detect wound infection. Using the first order Daubechies (db1) wavelet and the thirteen decomposition level, the authors decompose the sensor signals corresponding to healthy and wounded mice. Computing the correlation at each different scale between both signals, they detect and eliminate the components of the background (healthy mice) that are present in the target signal (wounded mice).

3.9 Robustness Improvement by Regularization

It was already mentioned that Ridge-Regression will generally produce solutions which are more robust than ordinary least squares. It is well-known that Ridge Regression [184] provides a regularized solution to ill-conditioned linear prediction problems.

$$\hat{b} = (X^T X + \lambda I)^{-1} X^T C \quad (11)$$

An evolution of this idea is the covariance matrix transforms (CMT) and counter-balanced distortion (CBD) methods proposed by DiFoggio [185]. In CMT, the intention is model desensitization by augmenting the covariance matrix in the direction of the expected artifacts. His solution is:

$$\hat{b} = (X^T X + \lambda W^T W)^{-1} X^T C \quad (12)$$

where W has the same number of columns as X and k rows. Every row contains a potential artifact vector. In this way the regression vector becomes less sensible to changes in X in the directions modeled by the artifact vectors. Those directions of undesired variance can be determined by dedicated experiments. For those cases (for instance due to non-linearity), where the solution cannot be set in the form of a generalized linear regression problem, DiFoggio proposes to add random variations to the training data in the directions of the

expected perturbations or artifacts. This proposal resembles the original proposal by Bishop [186] when he proved that training with noise was equivalent to Tikhonov regularization.

3.10 Robustness Improvement by Blind Source Separation

On many occasions a bilinear model is used to represent the data matrix.

$$X^T = C^T S + N^T \quad (13)$$

Blind Source Separation (BSS) techniques aim at identifying the set of sources (C: analyte signals, but also environmental perturbations or drift), that contribute after linear mixing to the observed signals X, in an unsupervised manner. To do so, some hypothesis about the sources has to be made first. The most well-known method for BSS is Independent Component Analysis (ICA). ICA assumes that the sources are mutually independent in statistical terms. Other techniques assume that the signals are non-negative, which is the case in Non-Negative Matrix Factorization (NMF) [187] and in Multi-Curve Resolution alternating least squares (MCR-ALS) [188], [189].

Di Natale et al. [190] proposed to use ICA to identify which part of the observed signal was due to environmental disturbances. In their experiment, a particular Independent Component was well correlated to temperature and humidity variations. Removal of that component improved the separability of the classes. With a similar idea Kermit and Tomic [191] applied ICA and found that a particular independent component was capturing mostly sensor drift. Removal of that component and ulterior signal reconstruction improved results.

BSS methods can also be used to automatically detect the presence of varying interferences than pollute the signal of interest without prior knowledge of them. In Pomareda et al. [192], when optimizing a luggage scanner based on an Ion Mobility Spectrometer, the application of MCR permitted the detection of the presence of an unknown interferent that was attributed to contaminants in the compressed air used to blow the luggage. In the same work, MCR allows memory effects in the system to be identified, that is, the presence of interferences in the analysis of the luggage under test due to analytes hidden in previously analyzed luggage. In a simulation study Bedoya et al. [193] propose to use blind source separation techniques in an array of two ISFETs measuring Ammonia, in the presence of an unknown concentration of potassium ions. The proposed technique is able to deal with the non-linear mixing effect present in ISFETs.

3. 11 Calibration transfer / Instrument Standardization

As mentioned above, a practical limitation of the use of multivariate data processing models occurs when an existing model developed with a comprehensive set of training samples is applied to a separate instrument. This secondary instrument may be a new instrument, or just the primary instrument used in different conditions, under the effects of aging and drift or because the prediction model has to be updated due to sensor replacement [160].

While some of the methods described above can also be used for calibration transfer [194]. On occasions, researchers into electronic noses have produced dedicated literature on this topic [195], [196].

Calibration transfer can be accomplished in three ways: (i) by transforming the secondary instrument readings to resemble the primary instrument, (ii) by updating the calibration model of the primary instrument, taking into account a reduced set of measurements taken with the secondary instrument, and (iii) by transforming the target values or labels used in the calibration.

In the electronic nose literature the most common approach is the first one. Researchers have been using univariate and multivariate approaches. The univariate approach was originally proposed by Shenk et al. [197]. The multivariate approach is known as Direct Standardization (DS) as it was originally proposed by Wang [198].

Simple univariate calibration transfer methods based in additive corrections or multiplicative corrections have been proposed in mass spectrometry-based e-noses and also in chemical sensor arrays [160]. Multivariate calibration transfer techniques applied to chemical sensor arrays were pioneered by Balaban et al. [199]. They attempted to build a mapping between two instruments with identical sensors: 12 Conducting Polymers. This paper tested univariate regression, multilinear regression and multilayer perceptrons. Best results were obtained by multilinear regression.

Tomic et al [160] describe two approaches to map the responses of the second instrument to the primary instrument. They report on calibration transfer between a master and four slave QMB instruments with 8 sensors. The first approach is just a linear regression of individual sensor responses between both instruments. The second approach is to implement a mapping using multivariate calibration models, in particular Partial Least Squares (PLS). In their application based on QMB sensors, they found both approaches equivalent in terms of performance. Mappings were built using 72 common training samples. While the readings become similar after the standardization, authors do not compare the performance with the direct individual calibration of each instrument on its own.

A similar idea was explored by L. Carmel [200]. In his paper, he builds mappings from e-noses of different technologies: (i) the cyranose 320, (ii) a QMB model of the MOSES and (iii) and a 32 Conductive Polymer Sensor Array. The dataset consisted of 23 pure chemicals. He tested linear regression (PLS), Neural Networks and Tesselation based Linear Interpolation. Reported results showed that the mapping direction was key. For instance, mapping the CP into the QMB model provided better results than the reverse mapping. Neural Networks were found to provide slightly better results than the other techniques. An interesting question posed by Carmel was the ability of the mapping to correctly classify odors not used in the construction of the calibration transfer. In this challenging scenario the results were worse but still remarkable.

Direct Standardization is limited in the sense that the transfer mapping has to be estimated from few transfer standards, and consequently can have problems to span all the variance

describing instrumental differences. As far as we know, those limitations have not been explored yet in the e-nose literature.

3.12 Sensor failure and replacement

Long time operation of chemical sensor arrays in field conditions could produce the failure of one or more sensors. This failure has to be detected and eventually the sensor has to be replaced. Chemical sensor failure detection has received little attention in the artificial olfaction community, even if this is very present in the process control literature.

A first approach to failure detection is based in outlier detection [201]. A clustering model of the training data is built. New readings have to be close to the previously identified clusters; otherwise a fault alarm is given. This approach is however not robust. New values for the analyte concentration not seen in calibration can be considered as fault indicators. A more principled approach has been suggested by Padilla et al. [202]. In a first contribution they explore sensor failure detection, sensor identification and correction. At this point, it is important to note that they explore the detection of changes in sensor sensitivity. In other words, the sensor keeps responding but with a different gain. Those faults are difficult to detect since the sensor signal keeps within the normal range, and the sensor waveform looks normal. However, those faults make the sensor depart from the previous covariance model. Padilla uses the norm of the residuals to the PCA subspace of normal operation, followed by an indicator function that provides the identity of the faulty sensor. Additionally projection to the PCA plane permits correction of the sensor signal and keeps the system in operation without degradation of instrument performance. If two sensors fail simultaneously, the use of structured residuals still allow the faulty sensors to be identified.

In a second paper, Padilla et al. [203] explores the problem of faulty sensor detection in the presence of drift. In this case, the PCA model obtained in the training phase becomes obsolete in time and degrades the ability of the data processing unit to detect sensor failure. In such a scenario, Padilla proposes to use the Multi-Scale PCA procedure of Misra et al. [174]. The analysis shows that the multi-resolution approach easily permits the separation of the abrupt faults from the slow drift and the sensor failure performance remains unaltered despite the presence of drift.

Even if those procedures permit sensor signal reconstruction on the basis of the correlation from other sensors, it is obvious that this faulty sensor eventually has to be replaced. At this time, it is absolutely impractical to fully recalibrate the system. Calibration transfer strategies can be used to avoid extensive recalibration (section 3.11).

4. Discussion and Future Perspectives

Signal and data processing are essential components for machine olfaction. Early proposals on processing chemical sensor signals date back to the late 80s. During the last 20 years of research we have seen this field mature and develop. Initially the main interest was to prove that multivariate signal and data processing could in fact improve the selectivity of, otherwise, rather broadly tuned sensors. Simple baseline corrections, feature extraction based typically on steady state values, and display plots based on PCA projections were the dominant approach 15 years ago. In some cases, computational intelligence solutions based on multilayer perceptrons (MLP), radial basis functions (RBF) or Self-Organizing Maps (SOM) were used.

In the last decade, we have seen several research trends: (i) Focus on feature extraction techniques in particular for sensor waveforms originated from sampling transients or temperature modulation (the use of only steady-state features is only seen in particular instrument configurations), (ii) Focus on array optimization, either by sensor selection, or by excitation optimization and (iii) Focus on robustness improvement. In the last decade research approach to signal and data processing has matured by focusing more on the distinct problems of chemical sensor signals and less on the straight-forward application of computational intelligence classifiers and regressors. The fact is that in most applications the advantage of using a simple (k-NN) classifier compared to a sophisticated one is marginal, while the largest gain can come from proper feature extraction and array optimization.

While ten years ago feature extraction was either purely heuristic or an alternatively direct application of PCA, today we have a rather exhaustive toolbox of techniques for feature extraction from sensor signals (see 2.1.1). On the other hand, we believe that in this work we have made an effort to systematically group current proposals aiming to improve the robustness of chemical sensor arrays through computational techniques (section 3). From the original proposals for drift compensation using either component correction or adaptive SOMs, this field has flourished and again a full set of techniques and methods have been described.

However, a critical word here is in order. While many valuable proposals have been reported in the last ten years of research, it is also true that they are largely unknown to the practitioner of machine olfaction techniques, unless he/she has a special interest in signal and data processing. There are several reasons behind this fact. Our community has not made those solutions available for the community in terms of code availability. In most cases, the use of the reported techniques requires direct coding of the algorithmic proposal. In most cases, this is just not done. Practitioners obviously prefer to use available code from commercial vendors. However, these commercial vendors (e.g. Eigenvector Research, CAMO and others) do serve the larger chemometric community that is focused on conventional analytical instrumentation by virtue of its much larger market presence. In consequence, the practitioner is biased to use techniques coded in available packages. From our point of view, the only way to revert this trend is by making the code available with proper documentation, a practice that is well

established in other domains such as bioinformatics (see e.g. the R-based Bioconductor package: <http://www.bioconductor.org>).

In a similar line of thought, in the near future we will see datasets made public by diverse labs. If those datasets are properly documented, objective tools intercomparison will be possible (this has long been a dream of many researchers in the community). The trend towards data sharing has already occurred in other related fields and it is also encouraged by some archival publications.

Although, due to size restrictions, it has not been covered in this review, a very promising and sound research line is that of biologically inspired computation for chemical sensing (sometimes also referred as neuromorphic olfaction)[13,14,15]. The basic idea behind it is to get inspiration from the biological sense of olfaction, its peculiar architecture and associated neural circuitry to extract the computational principles biology uses to process information provided by the full set of olfactory sensory neurons. The tremendous progress made in the understanding of the operational architecture of biological olfaction promises neuromorphic olfaction a long and hopefully fruitful path of research. Obviously, current knowledge of the olfactory sense is incomplete and many computational mechanisms are either unknown or poorly understood. However, examination of biological olfaction from a computational perspective can lead to breakthroughs in Machine Olfaction. For instance, slowly but steadily the immense Cartesian map (Receptors x Ligands) is being filled [204] and by examination of this empirical data our knowledge of odor coding by biological systems starts to develop. Its comparison with odor coding in machine olfaction [205] is a very interesting path of research and can reduce the difficulties machine olfaction has, to mimic human perception.

Concerning topics reviewed in this paper, we foresee that sensor selection and tuning optimization will keep high interest, and in particular recent proposals regarding automatic adaptation deserve further exploration [5]. Additionally, emphasis on robustness improvements in its various flavors will remain high since those techniques are a must to leave the lab bench and go to field applications.

From an application point of view, integration of chemical sensors into RFID tags [206] or small sensor motes [207] forming part of self-organizing sensor networks will enlarge the potential applications of chemical sensor arrays. Quite probably those deployments will have its own needs regarding signal and data processing.

5. Conclusion

Signal and data Processing is a main component of artificial olfaction systems. While the basic methodology to build qualitative or quantitative prediction models is firmly established, there are still limited literature regarding the use of signal and data processing to improve the robustness of the systems. Despite the limited activity, this review reports a toolbox of techniques that have shown their ability to improve the performance of the instruments. Only better dissemination of the available techniques and their application in different scenarios may prove that they are general solutions in AO systems.

6. Acknowledgements

SM and AGG are from a consolidated Grup de Recerca de la Generalitat de Catalunya and have support from the Departament d'Universitats, Recerca i Societat de la Informació la Generalitat de Catalunya (expedient 2009 SGR 0753).

7. References

- [1] Julian W. Gardner, Philip N. Bartlett, "A brief history of electronic noses", *Sensors and Actuators B: Chemical*, Volume 18, Issues 1-3, March 1994, Pages 210-211,
- [2] A. Perera, T. Sundic, A. Pardo, R. Gutierrez-Osuna, S. Marco, "A portable electronic nose based on embedded PC technology and GNU/Linux: hardware, software and applications," *IEEE Sensor J.*, vol. 2, pp. 235-246, 2002.
- [3] R. Gutierrez-Osuna, "Pattern analysis for machine olfaction: a review," *Sensors Journal, IEEE*, vol.2, no.3, pp. 189- 202, Jun. 2002.
- [4] A. Hierlemann, R. Gutierrez-Osuna, "Higher-order chemical sensing," *Chemical Reviews*, vol. 108, pp. 563-613, 2008.
- [5] R. Gutierrez-Osuna, A. Hierlemann, "Adaptive Microsensor Systems," *Annual Review of Analytical Chemistry*, vol. 3, pp. 255-276, 2010.
- [6] S.M. Scott, D. James, Z. Ali, "Data Analysis for Electronic Nose Systems," *Microchimica Acta*, vol. 156, pp. 183-207, 2006.
- [7] N. Yamazoe, "Toward innovations of gas sensor technology", *Sensors and Actuators B: Chemical*, Vol. 108, pp. 2-14, 2005.
- [8] A. Gurlo, Nanosensors: Does Crystal Shape Matter? *Small*, Vol 6, pp. 2077-2079, (2010).
- [9] M.E. Franke, T.J. Koplin, U. Simon, Metal and Metal Oxide Nanoparticles in Chemiresistors: Does the Nanoscale Matter? *Small*, Vol. 2, pp. 301, 2006.
- [10] E. Comini, C. Baratto, G. Faglia, M. Ferroni, A. Vomiero, G. Sberveglieri, "Quasi-one dimensional metal oxide semiconductors: Preparation, characterization and application as chemical sensors", *Progress in Materials Science*. Vol. 54, pp. 1-67, 2009.
- [11] G. Kowadlo, R.A. Russell, "Robot Odor Localization: A Taxonomy and Survey" *Int. J. of Robotics Res.* Vol. 27, pp. 869-894, 2008.
- [12] K. McGill, S. Taylor, "Robot Algorithms for Localization of Multiple Emission Sources", *ACM Computing Surveys* Vol. 43, article. 15, 2011.
- [13] S. Marco, A. Gutierrez-Gálvez, "Recent Developments in the Application of Biologically Inspired Computation to Chemical Sensing", *AIP Conf. Proc.* Vol. 1137, pp. 151-154, 2009.
- [14] B. Raman, M. Stopfer, S. Semancik, "Mimicking Biological Design and Computing Principles in Artificial Olfaction", *ACS Chemical Neuroscience*, Vol. 2, pp. 487-499, (2011).
- [15] A. Gutierrez-Gálvez, S. Marco (ed). *Biologically Inspired Signal Processing for chemical sensing*, Series on Computational Intelligence, Springer Verlag, 2009.
- [16] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, New York, Springer, 2009
- [17] A.R. Webb, *Statistical Pattern Recognition*, 2nd ed. Sussex, Wiley, 2002.
- [18] S.D. Brown, R. Tauler, B. Walczak, *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Elsevier, 2009.
- [19] E. Llobet, J. Brezmes, X. Vilanova, J.E. Sueiras, X. Correig, "Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state responses of a thick-film tin oxide gas sensor array," *Sensors and Actuators B: Chemical*, vol. 41, no. 1-3, pp. 13-21, Jun., 1997.
- [20] R. Gutierrez-Osuna, A. Gutierrez-Gálvez, N. Powar, "Transient response analysis for temperature-modulated chemoresistors," *Sensors and Actuators B: Chemical*, vol. 93, no. 1-3, pp. 57-66, Aug. 2003.
- [21] A. Ortega, S. Marco, A. Perera, T. Šundic, A. Pardo, J. Samitier, "An intelligent detector based on temperature modulation of a gas sensor with a digital signal processor," *Sensors and Actuators B: Chemical*, vol. 78, no. 1-3, pp. 32-39, Aug. 2001.
- [22] J. Janata, *Principles of Chemical Sensors*, Heidelberg/New York, Springer Science, 2009.
- [23] C. Calaza, E. Meca, S. Marco, M. Moreno, J. Samitier, L. Fonseca, I. Gracia, C. Cane, "Assessment of the final metrological characteristics of a MOEMS-based NDIR spectrometer through system modeling and data processing," *IEEE Sensors Journal*, vol.3, no.5, pp. 587- 594, Oct. 2003.
- [24] J. Samitier, J. M. López-Villegas, S. Marco, L. Cámara, A. Pardo, O. Ruiz, J. R. Morante, "A new method to analyze signal transients in chemical sensors," *Sensors and Actuators B: Chemical*, vol. 18, pp. 308-312, 1994.
- [25] M. Barker, W. Rayens, "Partial Least Squares for Discrimination" *J. of Chemometrics*, Vol. 17, 166-173, 2003.
- [26] D. Lindgren, P. Spångéus, "A novel feature extraction algorithm for asymmetric Classification," *IEEE Sensors Journal*, vol. 4, no. 5, pp. 643-650, 2004.
- [27] A. Perera, T. Yamanaka, A. Gutiérrez-Gálvez, B. Raman, R. Gutiérrez-Osuna, "A dimensionality-reduction technique inspired by receptor convergence in the olfactory system," *Sensors and Actuators B: Chemical*, vol. 116, no. 1-2, pp. 17-22, 2006.
- [28] A. Heilig, N. Bârsan, U. Weimar, M. Schweizer-Berberich, J.W. Gardner, W. Göpel, "Gas identification by modulating the temperatures of SnO₂-based thick film sensors", *Sensors and Actuators B:Chemical*. Vol. 43, pp. 45-51, 1997.

-
- [29] C. Distante, M. Leo, P. Siciliano, K.C. Persaud, "On the study of feature extraction methods for an electronic nose", *Sensors and Actuators B: Chemical*, Vol. 87, pp. 274-288, (2002).
- [30] R. Ionescu, E. Llobet, "Wavelet transform-based fast feature extraction from temperature modulated semiconductor gas sensors", *Sensors and Actuators B: Chemical*, Vol. 81, pp. 289-295, (2002)
- [31] X.J. Huang, Y.K. Choi, K.S. Yun, E. Yoon, "Oscillating behaviour of hazardous gas on tin oxide gas sensors: Fourier and Wavelet transform analysis", *Sensors and Actuators B: Chemical*, Vol. 115, pp. 357-364, 2006.
- [32] K.E. Kramer, S.L. Rose-Pehrsson, M.H. Hammond, D. Tillett, H.H. Streckert, "Detection and classification of gaseous sulfur compounds by solid electrolyte cyclic voltammetry of cermet sensor array", *Analytica Chimica Acta*, Vol. 584, pp 78-88, 2007.
- [33] F. J. Acevedo, S. Maldonado, E. Domínguez, A. Narvaez, F. Lopez, "Probabilistic support vector machines for multi-class alcohol identification," *Sensors and Actuators B: Chemical*, vol. 122, pp. 227-235, 2007.
- [34] A. Vergara, E. Martinelli, R. Huerta, A. D'Amico, C. Di Natale, "Orthogonal decomposition of chemo-sensory cues," *Sensors and Actuators B: Chemical*, vol. 159, no. 1, pp. 126-134, Nov. 2011.
- [35] M. K. Muezzinoglu, A. Vergara, R. Huerta, N. Rulkov, M. I. Rabinovich, A. Selverston, H. D.I. Abarbanel, "Acceleration of chemo-sensory information processing using transient features," *Sensors and Actuators B: Chemical*, vol. 137, no. 2, pp. 507-512, Apr. 2009.
- [36] A. Leone, C. Distante, N. Ancona, K.C. Persaud, E. Stella, P. Siciliano, "A powerful method for feature extraction and compression of electronic nose responses", *Sensors and Actuators B: Chemical*, Vol. 105, pp. 378-392, 2005.
- [37] T. Skov, R. Bro, "A new approach for modeling sensor based data," *Sensors and Actuators B: Chemical*, vol. 106, no. 2, pp. 719-729, May 2005.
- [38] M. Padilla, I. Montoliu, A. Pardo, A. Perera, S. Marco, "Feature extraction on three way enose signals," *Sensors and Actuators B: Chemical*, vol. 116, no. 1-2, pp. 145-150, Jul. 2006.
- [39] E. Martinelli, C. Falconi, A. D'Amico, C. Di Natale, "Feature extraction of chemical sensors in phase space", *Sensors and Actuators B: Chemical*, vol. 95, pp 132-139, 2003.
- [40] M. Falasconi, M. Pardo, G. Sberveglieri, I. Riccò, A. Bresciani, "The novel EOS835 electronic nose and data analysis for evaluating coffee ripening", *Sensors and Actuators B: Chemical*, vol. 110, pp. 73-80, 2005.
- [41] E. Martinelli, G. Pennazza, C. Di Natale, A. D'Amico, "Chemical sensors clustering with dynamic moments approach", *Sensors and Actuators B: Chemical*, Vol. 101, pp. 346-352, 2004.
- [42] A. Vergara, E. Llobet, E. Martinelli, C. Di Natale, A. D'Amico, X. Correig, "Feature extraction of metal oxide gas sensors using dynamic moments," *Sensors and Actuators B: Chemical*, vol. 122, no. 1, pp. 219-226, Mar. 2007.
- [43] S. Zhang, C. Xie, M. Hu, H. Li, Z. Bai, D. Zeng, "An entire feature extraction method of metal oxide gas sensors", *Sensors and Actuators B: Chemical*, Vol. 132, pp. 81-89, 2008.
- [44] A. Vergara, E. Martinelli, E. Llobet, F. Giannini, A. D'Amico, C. Di Natale, "An alternative global feature extraction of temperature modulated micro-hotplate gas sensors array using an energy vector approach," *Sensors and Actuators B: Chemical*, vol. 124, pp. 352-359, 2007.
- [45] S. Marco, A. Pardo, F. A. M. Davide, C. Di Natale, A. D'Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, U. Weimar, W. Göpel, "Different strategies for the identification of gas sensing systems", *Sensors and Actuators B: Chemical*, vol. 34, pp. 213-223, 1996.
- [46] A. Pardo, S. Marco, J. Samitier, "Non-linear inverse dynamic models of gas sensing systems based on chemical sensor arrays for quantitative measurements," *IEEE Trans. On Instrumentation and Measurement*, vol. 47, pp. 644-651, 1998.
- [47] M. Pardo, G. Faglia, S. Sberveglieri, M. Corte, F. Masulli, M. Riani, "A time delay neural network for estimation of gas concentrations in a mixture," *Sensors and Actuators B: Chemical*, vol. 65, pp. 267-269, 2000.
- [48] C. Zanchettin, T.B. Ludermir, "Wavelet filter for noise reduction and signal compression in an artificial nose," *Applied Soft Computing*, vol. 7, no. 1, pp. 246-256, Jan. 2007.
- [49] S. De Vito, A. Castaldo, F. Loffredo, E. Massera, T. Polichetti, I. Nasti, P. Vacca, L. Quercia, G. Di Francia, "Gas Concentration estimation in ternary mixtures with room temperature operating sensor array using tapped delay architectures", *Sensors and Actuators B: Chemical*, vol. 124, pp. 309-316, 2007.
- [50] T. Yamanaka, K. Yoshikawa, T. Nakamoto, "Improvement of odor-recorder capability for recording dynamical change in odor," *Sensors and Actuators B: Chemical*, vol. 99, no. 2-3, May 2004.
- [51] C. Di Natale, S. Marco, F. Davide, A. D'Amico, "Sensor-array calibration time reduction by dynamic modelling," *Sensors and Actuators B: Chemical*, vol. 25, no. 1-3, pp. 578-583, Apr. 1995.
- [52] F. Hosseini-Babaei, S. M. Hosseini-Golgo, "Analyzing the responses of a Thermally Modulated Gas Sensor using a Linear Identification Technique for Gas Diagnosis," *IEEE Sensor J.*, vol. 8, pp. 1837-1847, 2008.
- [53] S. Marco J. Samitier, J.R. Morante, "A novel time-domain method to analyse multicomponent exponential transients", *Measurement Science and Technology*, Vol. 6, pp. 135-142. 1995
- [54] R. Gutierrez-Osuna, H.T. Nagle, S. S. Schiffman, "Transient response analysis of an electronic nose using multi-exponential models," *Sensors and Actuators B: Chemical*, vol. 61, no. 1-3, Dec. 1999.
- [55] S. Wlodek, K. Colbow, F. Consadori, "Signal-shape analysis of a thermally cycled tin-oxide gas sensor," *Sensors and Actuators B: Chemical*, vol. 3, no. 1, pp. 63-68, Jan. 1991.
- [56] L. Carmel, S. Levy, D. Lancet, and D. Harel, "A feature extraction method for chemical sensors in electronic noses," *Sens. Actuators B: Chemical*, vol. 93, pp. 67-76, 2003.

-
- [57] R. Haddad, L. Carmel, D. Harel, "A feature extraction algorithm for multi-peak signals in electronic noses", *Sensors and Actuators B: Chemical*, vol. 120, pp. 467-472, 2006.
- [58] E. Llobet, M. Anaimi, A. Pruñonosa, E. Gras, "On the use of a self organizing map as a feature compressor in the building of calibration models: Application to FTIR-spectrophotometry", *Sensors and Actuators B: Chemical* Vol. 158, pp. 252-258. (2011)
- [59] A. Vergara and E. Llobet, "Sensor Selection and Chemo-Sensory optimization: toward an adaptable chemo-sensory system", *Frontiers in Neuroengineering*, Vol. 4, Article 19., pp. 1-21, 2012
- [60] M. Pardo, L.G. Kwong, G. Sberveglieri, K. Brubaker, J.F. Schneider, W.R. Penrose, J.R. Stetter, "Data analysis for a hybrid sensor array", *Sensors and Actuators B: Chemical*, vol. 106, no. 1, pp. 136-143, Apr. 2005.
- [61] M. Pardo, G. Sberveglieri, "Comparing the performance of different features in sensor arrays", *Sensors and Actuators B: Chemical*, vol. 123, pp. 437-443, 2007.
- [62] A. Pardo, S. Marco, C. Calaza, A. Ortega, A. Perera, T. Sundic, J. Samitier, "Methods for Sensor Selection in Pattern Recognition", *Electronic Noses and Olfaction 2000*, IOP Pub. Pp. 83-88. (2000).
- [63] H. Ulmer, J. Mitrovics, U. Weimar, W. Göpel, "Sensor arrays with one or several transducer principles? The advantage of hybrid modular systems", *Sensors and Actuators B: Chemical*, Vol. 65, pp. 79-81, 2000.
- [64] B. Raman, D.C. Meier,, J.K. Evju, S. Semancik, "Designing and optimizing microsensor array for recognizing chemical hazards in complex environments", *Sensors and Actuators B: Chemical*. Vol. 137, pp. 617-629 (2009).
- [65] B. Raman, D. Meier, S. Semancik, "A Statistical Approach to Materials Evaluation and Selection for Chemical Sensor Arrays", in *Computational Methods for Sensor Material Selection*, ed. M.A. Ryan, A.V. Shevade, C.J. Taylor, M.L. Homer, M. Blanco, J. Stetter, Springer, 2009
- [66] M.K. Muezzinoglu, A. Vergara, R. Huerta and M. I. Rabinovich, "A sensor conditioning principle for odor identification", *Sensors and Actuators B: Chemical*, Vol. 145, pp. 472-476, 2010.
- [67] A. Vergara, E. Llobet, J. Brezmes, X. Vilanova, P. Ivanov, I. Gracia, C. Cane, X. Correig, "Optimized temperatura modulation of micro-hotplate gas sensors through pseudo random binary sequences", *IEEE Sensor J.* Vol. 5., pp. 1369-1378, 2005.
- [68] A. Vergara, E. Llobet, J. Brezmes, P. Ivanov, X. Vilanova, I. Gracia, C. Cane, X. Correig, "Optimised temperatura modulation of metal oxide micro-hotplate gas sensors through multilevel pseudo random sequences", *Sensors and Actuators B: Chemical*, Vol. 124, 353-359. 2005.
- [69] A. Vergara, E. Llobet, J. Brezmes, P. Ivanov, C. Cané, I. Gràcia, X. Vilanova, X. Correig, "Quantitative gas mixture analysis using temperature-modulated micro-hotplate gas sensors: Selection and validation of the optimal modulating frequencies," *Sensors and Actuators B: Chemical*, vol. 123, pp. 1002-1016, 2007.
- [70] F.A.M. Davide, C. Di Natale, A. D'Amico, A. Hierlemann, J. Mitrovics, M. Schweizer, U. Weimar, W. Göpel, S. Marco, A. Pardo, "Dynamic calibration of QMB polymer-coated sensors by Wiener Kernel Estimation", *Sensors and Actuators B: Chemical*, Vol. 26-27, pp. 275-285, 1995.
- [71] A. Vergara, M. K. Muezzinoglu, N. Rulkov, R. Huerta, "Information-theoretic optimization of chemical sensors," *Sensors and Actuators B: Chemical*, vol. 148, no. 1, pp. 298-306, 2010.
- [72] R. Gosangi, R. Gutierrez-Osuna, "Active temperature programming for metal-oxide chemoresistors," *Sensors Journal, IEEE*, vol. 10, no. 6, pp. 1075-1082, 2010.
- [73] J. Fonollosa, B. Halford, L. Fonseca, J. Santander, S. Udina, M. Moreno, J. Hildebrand, J. Wöllenstein, S. Marco, "Ethylene optical spectrometer for apple ripening monitoring in controlled atmosphere store-houses", *Sensors and Actuators B: Chemical*, vol. 136, pp. 546-554, 2009.
- [74] P. Boilot, E.L. Hines, M.A. Gongora, R.S. Folland, "Electronic noses inter-comparison, data fusion and sensor selection in discrimination of standard fruit solutions", *Sensors and Actuators B: Chemical*, Vol.88, pp. 80-88, (2003).
- [75] M. Aleixandre, I. Sagayo, M.C. Horrillo, M.J. Fernández, L. Ares, M. Garcia, J.P. Santos, J. Gutierrez, "Analysis of neural networks and analysis of feature selection with genetic algorithm to discriminate among pollutant gas", *Sensors and Actuators B: Chemical*, Vol. 103, pp. 122-128, 2004.
- [76] J.W. Gardner, P. Boilot, E.L. Hines, "Enhancing electronic nose performance by sensor selection using a new integer-based genetic algorithm approach", *Sensors and Actuators B: Chemical*, vol. 106, pp. 114-121, 2005.
- [77] E. Phaisangittisagul, H.T. Nagle, "Sensor Selection for Machine Olfaction Based on Transient Feature Extraction," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, pp. 369-378, 2008.
- [78] E. Phaisangittisagul, H. T. Nagle, V. Areekul, "Intelligent method for sensor subset selection for machine olfaction," *Sensors and Actuators B: Chemical*, vol. 145, no. 1, pp. 507-515, 2010.
- [79] T. Alizadeh, "Chemiresistor sensors array optimization by using the method of coupled statistical techniques and its application as an electronic nose for some organic vapors recognition," *Sensors and Actuators B: Chemical*, vol. 143, no. 2, pp. 740-749, 2010.
- [80] Z. Xu, X. Shi, S. Lu, "Integrated sensor array optimization with statistical evaluation," *Sensors and Actuators B: Chemical*, vol. 149, no. 1, pp. 239-244, 2010.
- [81] E. Llobet, J. Brezmes, O. Gualdrón, X. Vilanova, X. Correig "Building parsimonious fuzzy ARTMAP models by variable selection with a cascaded genetic algorithm: application to multisensor systems for gas analysis," *Sensors and Actuators B: Chemical*, vol. 99, no. 2-3, pp. 267-272, May 2004.
- [82] O. Gualdrón, E. Llobet, J. Brezmes, X. Vilanova, X. Correig, "Coupling fast variable selection methods to neural network-based classifiers: Application to multisensor systems," *Sensors and Actuators B: Chemical*, vol. 114, no. 1, pp 522-529, Mar. 2006.
- [83] E. Llobet, O. Gualdrón, M. Vinaixa, N. El-Barbri, J. Brezmes, X. Vilanova, B. Bouchikhi, R. Gómez, J.A. Carrasco, X. Correig, "Efficient feature selection for mass spectrometry based electronic nose applications," *Chemometrics and intelligent laboratory systems*, vol. 85, no. 2, pp. 253-261, 2007.

-
- [84] P.M. Szczówka, A. Szczurek, B.W. Licznarski, "On reliability of neural network sensitivity analysis applied for sensor array optimization, *Sensors and Actuators B: Chemical*," vol. 157, no. 1, pp. 298-303, Sep. 2011.
- [85] O. Gualdrón, J. Brezmes, E. Llobet, A. Amari, X. Vilanova, B. Bouchikhi, X. Correig, "Variable selection for support vector machine based multisensor systems," *Sensors and Actuators B: Chemical*, vol. 122, no. 1, pp. 259-268, Mar. 2007.
- [86] A. Ortega, S. Marco, T. Šundić, J. Samitier, "New pattern recognition systems designed for electronic noses," *Sensors and Actuators B: Chemical*, vol. 69, no. 3, pp. 302-307, Oct. 2000.
- [87] M. Pardo, G. Sberveglieri, "Remarks on the use of multilayer perceptrons for the analysis of chemical sensor array data", *IEEE Sensor J.*, Vol. 4, pp. 355-363, 2004.
- [88] F. A.M. Davide, C. Di Natale, A. D'Amico, "Self-organizing multisensor systems for odour classification: internal categorization, adaptation and drift rejection," *Sensors and Actuators B: Chemical*, vol. 18, no. 1-3, pp. 244-258, Mar. 1994.
- [89] C. Distante, N. Ancona, P. Siciliano, "Support vector machines for olfactory signals recognition," *Sensors and Actuators B: Chemical*, vol. 88, no. 1, pp. 30-39, Jan. 2003.
- [90] M. Pardo, and G. Sberveglieri, "Classification of electronic nose data with support vector machines," *Sensors and Actuators B: Chemical*, vol. 107, pp. 730-737, 2005.
- [91] K. Brudzewski, S. Osowski, T. Markiewicz, J. Ulaczyk, "Classification of gasoline with supplement of bio-products by means of an electronic nose and SVM neural network," *Sensors and Actuators B: Chemical*, vol. 113, no. 1, pp. 135-141, 2006.
- [92] K. Brudzewski, S. Osowski, T. Markiewicz, "Classification of milk by means of an electronic nose and SVM neural network," *Sensors and Actuators B: Chemical*, vol. 98, pp. 291-298, 2004.
- [93] M. Bicego, "Odor classification using similarity-based representation," *Sensors and Actuators B: Chemical*, vol. 110, pp. 225-230, 2005.
- [94] M. K. Muezzinoglu, A. Vergara, R. Huerta, " Unified framework for Volatile Organic Compound classification and regression," in *International Joint Conference on Neural Networks*, Barcelona, 2010, pp. 1-7.
- [95] X. Wang, M. Ye, C.J. Duanmu, "Classification of data from electronic nose using relevance vector machines," *Sensors and Actuators B: Chemical*, vol. 140, no. 1, pp. 143-148, Jun. 2009.
- [96] S. Brahim-Belhouari, and A. Bermak, "Gas identification using density models," *Pattern Recognition Letters*, vol. 26, pp. 699-706, 2005.
- [97] E. Phaisangittisagul, H.T. Nagle, "Enhancing multiple classifier system performance for machine olfaction using odor-type signatures," *Sensors and Actuators B: Chemical*, vol. 125, pp. 246-253, 2007.
- [98] C. Li, P. Heinemann, R. Sherry, "Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection," *Sensors and Actuators B: Chemical*, vol. 125, pp. 301-310, 2007.
- [99] J.Cho, X.Li, Z. Gy, P. Kurup, "Recognition of Explosive Precursors Using Nanowire Sensor Array and decision tree learning", *IEEE Sensor Journal*, In press.
- [100] L. Breiman, J.H. Friedman, R.a. Olshen, C.J. Stone, *Classification and Regression Trees*, Wadsworth Inc. Belmont, CA, 1984
- [101] M. Pardo, G. Sberveglieri, "Random Forests and Nearest Shrunken Centroids for the classification of Sensor Array Data," *Sensors and Actuators B: Chemical*, vol. 131, pp. 93-99, 2008.
- [102] A. Gulbag, and F. Temurtas, "A study on quantitative classification of binary gas mixtures using neural networks and adaptive neuro-fuzzy inference systems," *Sensors and Actuators B: Chemical*, vol. 131, no. 1, pp. 196-204, 2008.
- [103] B. Mumyalmaz, A. Özmen, M. A. Ebeoglu, C. Yasaltin, "Predicting gas concentrations of ternary gas mixtures for a predefined 3D sample space," *Sensors and Actuators B: Chemical*, vol. 128, pp. 594-602, 2008.
- [104] K. Desai, Y. Badhe, S. T. Sanjeev, D. K. Bhaskar, "Soft-sensor development for fedbatch bioreactors using support vector regression," *Biochemical Engineering Journal*, vol. 27, pp. 225-239, 2006.
- [105] A. Shmilovici, G. Bakir, S. Marco, and A. Perera, "Finding the best calibration points for a gas sensor array with support vector regression," in *IEEE International Conference on Intelligent Systems*, Varna, Bulgaria, 2004, pp. 174-177.
- [106] J. Lozano, J.P. Santos, T. Arroyo, M. Aznar, J.M. Cabellos, M. Gil, M.C. Horrillo, "Correlating e-nose responses to wine sensorial descriptors and gas chromatography-mass spectrometry profiles using least squares regression analysis," *Sensors and Actuators B: Chemical*, vol. 127, pp. 267-276, 2007.
- [107] World Health Organization, "Quality assurance of pharmaceuticals: A Compendium of guidelines and related materials", 2007.
- [108] G. Korotcenkov, B.K. Cho, "Instability of metal-oxide based conductometric gas sensors and approaches for stability improvement (short survey)," *Sensors and Actuators B: Chemical*, vol. 156, pp. 527-538, 2011.
- [109] H. Knobloch, C. Turner, A. Spooner, m. Chambers, "Methodological Variation in headspace analysis of liquid samples using electronic Nose," *Sensors and Actuators B: Chemical*, vol. 139, pp. 353-360, 2009.
- [110] S.D. Brown, "Transfer of Multivariate Calibration Models" in *Comprehensive Chemometrics*, S. D. Brown, R. Tauler, and B. Walczak Eds., Elsevier, 2009, pp. 345-378.
- [111] J. Quiñero-Candela, M. Sugiyama, A. Schwaighofer, N.D. Lawrence, *Dataset Shift in Machine Learning*, Cambridge: MIT Press, 2009.
- [112] P. Gujral, M. Amrhein, B.M. Wise, D. Bonvin, "Framework for explicit drift correction in multivariate calibration models," *J. Chemometrics*, vol. 24, pp. 534-543, 2010.
- [113] P. Gujral, M. Amrhein, D. Bonvin, "Drift correction in multivariate calibration models using on-line reference measurements," *Anal. Chim. Acta*, vol. 642, pp. 27-32, 2009.
- [114] M. Zeaiter, J.-M. Roger, V. Bellon-Maurel, "Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods," *TrAC Trends in Analytical Chemistry*, vol. 24, no. 5, pp. 437-445, May 2005.
- [115] N. Nimsuk, T. Nakamoto, "Study on the odor classification in dynamical concentration robust against humidity and temperature changes," *Sensors and Actuators B: Chemical*, vol. 134, no. 1, pp. 252-257, Aug. 2008.
- [116] M. Kuske, M. Padilla, A.C. Romain, J. Nicolas, R. Rubio, S. Marco, "Detection of diverse mould species growing on building materials by gas sensor arrays and pattern recognition," *Sensors and Actuators B: Chemical*, vol. 119, no. 1, pp. 33-40, Nov. 2006.

-
- [117] B. Raman, R. Shenoy, D.C. Meier, K.D. Benkstein, S. Semancik, "A MEMS-Based Approach that uses Temperature Dependent Sensing Responses to Recognize Chemical Targets in Untrained Backgrounds", *Proc. Of the IEEE Sensors Conference*, pp. 1262-1266, 2010.
- [118] K.D. Benkstein, B. Raman, C.B. Montgomery, C.J. Martinez, S. Semancik, "Microsensors in Dynamic Backgrounds: Toward Real-Time Breath Monitoring", *IEEE Sensor J.* Vol. 10, pp. 137-144, 2010.
- [119] B. Mumykmaz, A. Özmen, M. A. Ebeoğlu, C. Taştaltın, I. Gürol, "A study on the development of a compensation method for humidity effect in QCM sensor responses," *Sensors and Actuators, B: Chemical*, vol. 147, no. 1, pp. 277-282, 2010.
- [120] M. Aliaghasarghamish, S. Ebrahimi, "Recursive least squares fuzzy modeling of chemoresistive gas sensors for drift compensation," in *Innovations in International Symposium on Intelligent Systems and Applications*, Istanbul, 2011, pp.1-5.
- [121] M. Hajmirzaheydarali, A. H. Zare, "A gas sensor system coupled to an artificial neural network capable of self-calibration against ambient humidity and temperature fluctuations," in *19th Iranian Conference on Electrical Engineering*, Tehran, 2011.
- [122] S. Brahim-Belhouari, A. Bermak, M. Shi, and P. C. H. Chan, "Fast and robust gas identification system using an integrated gas sensor technology and Gaussian mixture models," *IEEE Sensors J.*, vol. 5, no. 6, pp. 1433-1444, Dec. 2005.
- [123] S. De Vito, E. Massera, M. Piga, L. Martinotto, G. Di Francia, "On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario," *Sensors and Actuators B: Chemical*, vol. 129, no. 2, pp. 750-757, Feb. 2008.
- [124] S. De Vito, Marco Piga, L. Martinotto, G. Di Francia, "CO, NO₂ and NO_x urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization," *Sensors and Actuators B: Chemical*, vol. 143, no. 1, pp. 182-191, Dec. 2009.
- [125] D.J.C. MacKay, "A Practical Bayesian Framework for Backpropagation Networks," *Neural Computation*, vol. 4, pp. 448-472, 1992.
- [126] W. Tsujita, H. Ishida, T. Moriizumi, "Auto-calibration of dynamic gas sensor network: influence of static sensors," *Solid-State Sensors, Actuators and Microsystems*, vol.2, pp. 1951- 1952, 2005.
- [127] Tsujita, W.; Ishida, H.; Moriizumi, T., "Dynamic gas sensor network for air pollution monitoring and its auto-calibration," in *Proc. of IEEE Sensors*, Vienna, 2004, vol.1, pp. 56- 59.
- [128] P.H. Rogers, S. Semancik, "Feedback-enabled discrimination enhancement for temperature-programmed chemiresistive sensors", *Sensors and Actuators B:Chemical*, Vol. 158, pp. 111-116, 2011.
- [129] D. Bednarczyk and S. DeWeerth, "Smart chemical sensing arrays using tin oxide sensors and analog winner-take-all signal processing," *Sensors and Actuators B: Chemical*, vol. 27, no. 1-3, pp. 271-274, 1995.
- [130] Q. Guo, W. Wu, D.L Massart, "The robust normal variate transform for pattern recognition with near-infrared data," *Analytica Chimica Acta*, vol. 382, no. 1-2, pp. 87-103, Feb. 1999.
- [131] M. J. Wenzel, A. Mensah-Brown, F. Josse, E. E. Yaz, "Online Drift Compensation for Chemical Sensors Using Estimation Theory," *Sensors Journal, IEEE*, vol.11, no.1, pp.225-232, Jan. 2011.
- [132] M. Zuppa, C. Distante, P. Siciliano, K.C. Parsaud, "Drift Counteraction with multiple self-organizing mpas for an electronic nose," *Sensors and Actuators B: Chemical*, vol. 98, pp. 305-317, 2004.
- [133] E. Llobet, J. Brezmes, R. Ionescu, X. Vilanova, S. Al-Khalifa, J.W. Gardner, N. Bârsan, X. Correig, "Wavelet transform and fuzzy ARTMAP-based pattern recognition for fast gas identification using a micro-hotplate gas sensor," *Sensors and Actuators B: Chemical*, vol. 83, no. 1-3, pp. 238-244, Mar. 2002.
- [134] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613-627, 1995.
- [135] D. Huang, H. Leung, "Reconstruction of Drifting Sensor Responses Based on Papoulis-Gerchberg Method", *IEEE Sensor J.*, vol. 9, pp. 595-604., 2009.
- [136] A.C. Romain, J. Nicolas, "Long term stability of metal oxide based gas sensors for e-nose environmental applications: An overview," *Sensors and Actuators B: Chemical*, vol. 146, pp. 502-506, 2010.
- [137] F. Hossein-Babaei, V. Ghafarinia, "Compensation for the drift-like terms caused by environmental fluctuations in the responses of chemoresistive gas sensors," *Sensors and Actuators B: Chemical*, vol. 143, no. 2, pp. 641-648, Jan. 2010.
- [138] B. C. Sisk, N. S. Lewis, "Comparison of analytical methods and calibration methods for correction of detector response drift in arrays of carbon black-polymer composite vapor detectors," *Sensors and Actuators B: Chemical*, vol. 104, no. 2, pp. 249-268, Jan. 2005.
- [139] H. John-Erik, T. Oliver, K. Knut, "A calibration method for handling the temporal drift of solid state gas-sensors," *Analytica Chimica Acta*, vol. 407, no. 1-2, pp. 23-39, Feb. 2000.
- [140] M. Frank, H. Ulmer, J. Ruiz, P. Visani, U. Weimar, "Complementary analytical measurements based upon gas chromatography-mass spectrometry, sensor system and human sensory panel: a case study dealing with packaging materials," *Analytica Chimica Acta*, vol. 431, no. 1, pp. 11-29, Mar. 2001.
- [141] A. Lorber, K. Faber, and B. Kowalski, "Net analyte signal calculation in multivariate calibration," *Analytical Chemistry*, vol. 68, pp. 1620-1626, Jan 1997.
- [142] S. Wold, H. Antti, F. Lindgren, and J. Öhman, "Orthogonal signal correction of near infrared spectra," *Chemometrics and intelligent laboratory systems*, vol. 44, pp. 175-185, Jan 1998.
- [143] J. Westerhuis, S. de Jong, and A. Smilde, "Direct orthogonal signal correction," *Chemometrics and intelligent laboratory systems*, vol. 56, pp. 13-25, Jan 2001.
- [144] C. Andersson, "Direct orthogonalization," *Chemometrics and intelligent laboratory systems*, vol. 47, no. 1, pp. 51-63, 1999.
- [145] J. Trygg and S. Wold, "Orthogonal projections to latent structures (o-pls)," *Journal of Chemometrics*, vol. 16, pp. 119-128, Jan 2002.
- [146] J. Trygg, "O2-pls for qualitative and quantitative analysis in multivariate calibration," *Journal of Chemometrics*, vol. 16, no. 6, pp. 283-293, 2002.
- [147] M. Rantalainen, M. Bylesjö, O. Cloarec, J. K. Nicholson, E. Holmes, and J. Trygg, "Kernel-based orthogonal projections to latent structures (k-o-pls)," *Journal of Chemometrics*, vol. 21, pp. 376 - 385, Jan 2007.

-
- [148] R. N. Feudale, H. Tan, and S. D. Brown, "Improved piecewise orthogonal signal correction algorithm," *Applied Spectroscopy*, vol. 57, no. 10, pp. 1201–1206, 2003.
- [149] J. M. Roger, F. Chauchard, V. Bellon-Maurel, "EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits," *Chemom. Intell. Lab.Syst.*, vol. 66, pp. 191–204, 2003.
- [150] Y. Zhu, T. Fearn, D. Samuel, A. Dhar, O. Hameed, S. G. Bown, L. B. Lovat, "Error removal by orthogonal subtraction (EROS): a customised pretreatment for spectroscopic data," *J. Chemom.*, vol. 22, pp. 130–134, 2008.
- [151] B. Li, C. Wang, and D.-S. Huang, "Supervised feature extraction based on orthogonal discriminant projection," *Neurocomputing*, vol. 73, pp. 191–196, Oct 2009.
- [152] M. Zeaiter, J.M. Bellon-Maurel, "Dynamic Orthogonal Projection: A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations", *Chemometrics and Intelligent Laboratory Systems*, vol. 80, pp. 227–235, 2006.
- [153] T. Artursson, T. Eklöv, I. Lundström, P. Martensson, M. Sjöström, M. Holmberg, "Drift correction for gas sensors using multivariate methods," *J.Chemom.*, vol. 14, pp. 711–723, 2000.
- [154] R. Gutierrez-Osuna, "Drift reduction for metal-oxide sensor arrays using canonical correlation regression and partial least squares", in *Electronic Noses and Olfaction*, pp. 147-152, IOP Pub. 2000.
- [155] H. Goicoechea and A. Olivieri, "A comparison of orthogonal signal correction and net analyte preprocessing methods. theoretical and experimental study," *Chemometrics and intelligent laboratory systems*, vol. 56, pp. 73–81, Jan 2001.
- [156] O. Svensson, T. Kourti, and J. F. MacGregor, "An investigation of orthogonal signal correction algorithms and their characteristics," *Journal of Chemometrics*, vol. 16, pp. 176 – 188, Jan 2002.
- [157] J. Gabrielsson and J. Trygg, "Recent developments in multivariate calibration," *Critical Reviews in Analytical Chemistry*, vol. 36, no. 3, pp. 243–255, 2006.
- [158] H. Martens, M. Høy, B. M. Wise, R. Bro, B. P. Brockhoff, "Pre-whitening of data by covariance-weighted pre-processing," *J. Chemom.*, vol. 17, pp. 153–165, 2003.
- [159] P. W. Hansen, "Pre-processing method minimizing the need for reference analyses," *J. Chemom.*, vol. 15, pp. 123–131, 2001.
- [160] O. Tomic, T. Eklov, K. Kvaal, J. E. Haugen, "Recalibration of a gas-sensor array system related to sensor replacement," *Analytica Chimica Acta*, vol. 512, pp. 199-206, 2004.
- [161] M. Padilla, A. Perera, I. Montoliu, A. Chaudry, K. Persaud, S. Marco, "Drift Compensation of gas sensor array data by Orthogonal Signal Correction", *Chemometrics and Intelligent Laboratory Systems*, vol 100, pp. 28-35, 2010.
- [162] A. Ziyatdinov, S. Marco, A. Chaudry, K. Persaud, P. Caminal, A. Perera, "Drift compensation of gas sensor array data by common principal component analysis," *Sensors and Actuators B: Chemical*, vol. 146, no. 2, pp. 460-465, Apr. 2010.
- [163] B. Lazzerini, F. Marcelloni, "Counteracting drift of olfactory sensors by appropriately selecting features", *Electronic Letters*, vol. 36, pp. 509-510, 2000.
- [164] F. Marcelloni, "Recognition of olfactory signals based on supervised fuzzy C-means and k-NN algorithms," *Pattern Recognition Letters*, vol. 22, no. 9, pp. 1007-1019, Jul. 2001.
- [165] M. Holmberg, F. A. M. Davide, C. DiNatale, A. D'Amico, F. Wlnquist, I. Lundstrom, "Drift Counteraction in odour recognition applications: lifelong calibration method", *Sensors and Actuators B. Chemical*, vol. 42, pp. 185-194, 1997.
- [166] S. Marco, A. Ortega, A. Pardo, J. Samitier, "Gas identification with tin oxide sensor array and self-organizing maps: adaptive correction of sensor drifts," *IEEE Transactions on Instrumentation and Measurement*, vol. 47, pp. 316-321, 1998.
- [167] C. Distante, P. Siciliano, K.C. Persaud, "Dynamic Cluster Recognition with multiple self-organizing maps" *Pattern Analysis and Applications*, vol. 5, pp. 306-315, 2002.
- [168] S. Si Carlo, M. Falasconi, E. Sanchez, A. Scionti, G. Squillero, A. Tonda, "Increasing pattern recognition accuracy for chemical sensing by evolutionary based drift compensation," *Pattern Recognition Letters*, vol. 32, pp. 1594-1603, 2011.
- [169] J.C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*, New York, Springer, 2010.
- [170] N. Kim, H.G. Byun, K.C. Persaud, "Novel Signal Processing Techniques based on PDF Information for Sensor Drift Compensation" *Sensor Letters*, vol 9, pp. 439-443, 2011.
- [171] N.Kim, H.G. Byun, K.C. Persaud, "A New Compensation Method for Sensor-Drift Effect based on the Cross-Correntropy Concept" *Sensor Letters*, vol. 9, pp. 710-713, 2011.
- [172] H. Ding, H. Ge, J. Liu, "High performance of gas identification by wavelet transform-based fast feature extraction from temperature modulated semiconductor gas sensors," *Sensors and Actuators B: Chemical*, vol. 107, no. 2, pp. 749-755, Jun. 2005.
- [173] L.H. Chiang, E.L. Russell, R.D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*, Series: Advanced Textbooks in Control and Signal Processing, Springer, 2001.
- [174] M. Misra, H.H. Yue, S.J. Qin, C. Ling, "Multivariate Process Monitoring and fault diagnosis by multi-scale PCA," *Computers and Chemical Engineering*, vol. 26, pp. 1281-1293, 2002.
- [175] P. Angelov, A. Kordon, Z. Xiaowei, "Evolving fuzzy inferential sensors for process industry," in *3rd International Workshop on Genetic and Evolving Systems*, Witten-Bommerhol, 2008, pp.41-46.
- [176] P. Angelov, D. Filev, "An approach to on-line identification of evolving Takagi-Sugeno models," *IEEE Trans. on Systems, Man and Cybernetics*, part B, vol.34, no1, pp. 484-498, 2004.
- [177] A. Perera, N. Papamichail, N. Barsan, U. Weimar, S. Marco, "On-line novelty detection by recursive dynamic principal component analysis and gas sensor arrays under drift conditions," *IEEE Sesnor J.*, vol. 6, pp. 770-783, 2006.
- [178] H.R. Keller, D.L. Masart, "Peak purity control in liquid chromatography with photo-diode array detection by a fixed size moving window evolving factor analysis," *Analytica Chimica Acta*, vol. 246, pp. 379-390, 1991.
- [179] R. Gutierrez-Osuna, N. Powar, "Odor mixtures and chemosensory adaptation in gas sensor arrays," *International Journal on Artificial Intelligence Tools*, vol. 12, no. 1, pp. 1-16, 2003.
- [180] R. Gutierrez-Osuna, B. Raman, "Cancellation of chemical backgrounds with generalized Fisher's linear discriminants," in *Proc. of IEEE Sensors*, Vienna, vol. 3, pp. 1381-1384, 2004.

-
- [181] R. Gutierrez-Osuna, A. Gutierrez-Galvez, "Habituation in the KILL olfactory model with chemical sensor arrays," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1565–1568, 2003.
- [182] A. Gutierrez-Galvez, R. Gutierrez-Osuna, "Contrast enhancement and background suppression of chemosensor array patterns with the KILL model," *Int. J. Intell. Syst.*, vol. 21, no. 9, pp. 937-953, 2006.
- [183] J. Feng, F. Tian, J. Yan, Q. He, Y. Shen, L. Pan, "A background elimination method based on wavelet transform in wound infection detection by electronic nose," *Sensors and Actuators B: Chemical*, vol. 157, no. 2, pp. 395-400, 2011.
- [184] A. E. Hoerl, E. Arthur, R. W. Kennard, R. W. Hoerl, "Practical Use of Ridge Regression: A Challenge Met," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 34, no. 2, pp. 114–120, 1985.
- [185] R. DiFoggio, "Desensitizing models using covariance matrix transforms or counter-balanced distortions", *J. Chemometrics*, vol. 19, pp. 203-215, 2005.
- [186] C.M. Bishop, "Training with noise is equivalent to Tikhonov Regularization", *Neural Computation*, vol. 7, pp. 108-116, 1995.
- [187] A. Cichoki, R. Zdunek, A.H. Phan, S.I. Amari, *Non-negative Matrix Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Chichester: Wiley, 2009.
- [188] A. De Juan, R. Tauler, "Chemometrics applied to unravel multicomponent processes and mixtures: revisiting latest trends in multivariate resolution," *Anal. Chim. Acta*, vol. 500, no. 1-2, pp. 195–210, 2003.
- [189] I. Montoliu, R. Tauler, M. Padilla, A. Pardo, S. Marco, "Multivariate curve resolution applied to temperature-modulated metal oxide gas sensors," *Sensors and Actuators B: Chemical*, vol. 145, no. 1, pp. 464-473, Mar. 2010.
- [190] Di Natale, E. Martinelli, A. D'Amico, "Counteraction of environmental disturbances of electronic nose data by independent component analysis," *Sensors and Actuators B: Chemical*, vol. 82, pp. 158-165, 2002.
- [191] M. Kermit, O. Tomic, "Independent Component Analysis applied on gas sensor array measurement data," *IEEE Sensor J.*, vol. 3, pp. 218-228, 2003.
- [192] V. Pomareda, D. Calvo, A. Pardo, S. Marco, "Hard modeling Multivariate Curve Resolution using LASSO: Application to Ion Mobility Spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 2, pp. 318-332, Dec. 2010.
- [193] G. Bedoya, C. Jutten, S. Bermejo, J. Cabestany, "Improving semiconductor-based chemical sensor arrays using advanced algorithms for blind source separation," in *Proceedings of the ISA/IEEE Sensors for Industry Conference.*, New Orleans, 2004, pp. 149-154.
- [194] R. Feudale, N. Woody, H. Tan, A. Myles, S. Brown, and J. Ferré, "Transfer of multivariate calibration models: a review," *Chemometrics and intelligent laboratory systems*, vol. 64, no. 2, pp. 181–192, 2002.
- [195] L. Pillonel, J.O. Bosset, R. Tabacchi, "Data transferability between two MS-based electronic nose using processed cheeses and evaporated milk as reference materials," *European Food Research and Technology*, vol. 214, pp. 160-162, 2002.
- [196] J. L. Pérez Pavón, M. Nogal Sánchez, C. García Pinto, M. E. Fernández Laespada, B. Moreno Cordero, A. Guerrero Peña, "Strategies for qualitative and quantitative analyses with mass spectrometry-based electronic noses," *TrAC Trends in Analytical Chemistry*, vol. 25, no. 3, pp. 257-266, Mar. 2006.
- [197] J. Shenk, M. O. Westerhaus, W. C. Templeton, "Calibration Transfer between Near-Infrared Reflectance Spectrophotometers," *Crop Sci.*, vol. 25, pp. 159–161, 1985.
- [198] Y. Wang, D. J. Veltkamp, B. R. Kowalski, "Multivariate Instrument Standardization," *Anal. Chem.*, vol. 63, pp. 2750–2756, 1991.
- [199] M. O. Balaban, F. Korel, A. Z. Odabasi, G. Folkes, "Transportability of data between electronic noses: mathematical methods," *Sensors and Actuators B: Chemical*, vol. 71, pp. 203-211, 2000.
- [200] O. Shaham, L. Carmel, D. Harel, "On mappings between electronic noses," *Sensors and Actuators B: Chemical*, vol. 106, pp. 76-82, 2005.
- [201] S. J. Wellington, J. K. Atkinson, R. P. Sion, "Sensor validation and fusion using the Nadaraya-Watson statistical estimator," in *Proc. of the Fifth International Conference on Information Fusion*, Annapolis, Maryland, 2002, vol.1, pp. 321-326.
- [202] M. Padilla, A. Perera, I. Montoliu, A. Chaudry, K.C. Persaud, S. Marco, "Poisoning fault diagnosis in chemical gas sensor arrays using multivariate signal processing and structured residuals generation", in *IEEE Int. Symp. On Intelligent Signal Processing*, Alcalá de Henares, 2007, pp. 555-560.
- [203] M. Padilla, A. Perra, I. Montoliu, A. Chaudry, K. Persaud, S. Marco, "Fault detection, identification, and reconstruction of faulty chemical gas sensors under drift conditions, using Principal Component Analysis and Multiscale-PCA", in *IEEE Int. Joint Conference on Neural Networks*, Barcelona, 2010, pp. 1-7.
- [204] H. Saito, Q. Chi, H. Zhuang, H. Matsunami, J.D. Mainland, "Odor coding by a mammalian receptor repertoire", *Science Signaling*, Vol. 2, p. ra9, 2009.
- [205] A.Z. Berna, A.R. Anderson, S.C. Trowell, "Bio-benchmarking of electronic nose sensors" *Plos ONE*, 4 (7): e6406.
- [206] E. Abad, S. Zampolli, S. Marco, A. Scorzoni, B. Mazzolari, A. Juarros, D. Gómez, I. Elmi, G.C. Cardinal, J.M. Gómez, F. Palacio, M. Cicioni, A. Mondini, T. Becker, I. Sayhan, "Flexible tag microlab development: Gas sensors integration in RFID flexible tags for food logistics", *Sensors and Actuators B: Chemical*, Vol. 127, pp. 2-7, 2007.
- [207] D. Diamond, S. Coyle, S. Scarmagnani, J. Hayes, "Wireless Sensor Networks and Chemo/Biosensing", *Chem. Rev.* Vol. 108, pp. 652-679, 2008.