Sensors and Actuators B 231 (2016) 276–284

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

journal homepage: www.elsevier.com/locate/snb

Calibration transfer in temperature modulated gas sensor arrays

L. Fernandez[a,b,*], S. Guney[c], A. Gutierrez-Galvez[a,b], S. Marco[a]

[a] Institute of Bioengineering of Catalonia (IBEC), Barcelona, Spain
[b] University of Barcelona, Barcelona, Spain
[c] Baskent University, Ankara, Turkey

ARTICLE INFO

Article history:
Received 27 July 2015
Received in revised form 11 February 2016
Accepted 26 February 2016
Available online 10 March 2016

Keywords:
Calibration transfer
Gas sensor array
MOX
Temperature modulation

ABSTRACT

Shifts in working temperature are an important issue that prevents the successful transfer of calibration models from one chemical instrument to another. This effect is of special relevance when working with gas sensor arrays modulated in temperature. In this paper, we study the use of multivariate techniques to transfer the calibration model from a temperature modulated gas sensor array to another when a global change of temperature occurs. To do so, we built 12 identical master sensor arrays composed of three different types of commercial Figaro sensors and acquired a dataset of sensor responses to three pure substances (ethanol, acetone and butanone) dosed at 7 concentrations. The master arrays are then shifted in temperature (from −50 to 50 °C, ΔT = 10 °C) and considered as slave arrays. Data correction is performed for an increasing number of transfer samples with 4 different calibration transfer techniques: Direct Standardization, Piece-wise Direct Standardization, Orthogonal Signal Correction and Generalized Least Squares Weighting. In order to evaluate the performance of the calibration transfer, we compare the Root Mean Square Error of Prediction (RMSEP) of master and slave arrays, for each instrument correction. Best results are obtained from Piece-wise Direct standardization, which exhibits the lower RMSEP values after correction for the smaller number of transfer samples.

1. Introduction

Shifts in working temperature prevent direct calibration transfer between chemical measuring instruments [1]. That is to say that calibration models built for instrument I_1 working at a temperature T_1 experience an important degradation on prediction when applied to data samples of instrument I_2 at T_2 ($T_2 \neq T_1$). This is a matter of the utmost importance for temperature modulated metal oxide gas sensor arrays [2], where tolerances in heater resistances values, variations on the working flow conditions, and environmental fluctuations can give rise to a global shift ΔT of the sensor nominal temperature profile, and therefore of the sensor response waveform. A naïve approach to overcome invalid calibration transfer is to create independent calibration models for each of the arrays. However, this is an impractical solution, since it implies costly and labor-intensive measurement periods. A preferable methodology is the use of instrument standardization techniques [3] to correct the temperature shift in sensor arrays as compared to a reference array (from now on we will refer to these arrays as slave and master arrays respectively) calibrated for a complete set of experimental conditions and a proper temperature profile. The calibration transfer then relies on the measurement of only a small subset of experimental points in the slave array (herein called transfer samples).

According to Marco and Gutierrez-Galvez [4], calibration transfer can be realized following three different strategies: (i) by transforming the slave instrument readings to keep the calibration model of the master instrument still valid on the slave instrument, (ii) by modifying the target labels of the samples from the slave instrument so as to match those obtained from the master instrument, and (iii) by forcing master and slave readings to become more similar before creating the calibration model. Direct Standardization (DS) and Piecewise Direct Standardization (PDS) are the more popular methods to standardize slave instrument response [5,6]. With respect to the second strategy, the most frequently used method is univariate Multiplicative Signal Correction (MSC) [7]. Finally, Component Correction (CC), Orthogonal Signal Correction (OSC) and Generalized Least Squares Weighting (GLSW) are commonly used to remove instrument-to-instrument variability [8–10].

A large number of studies on instrument standardization have been applied to Near Infrared Spectroscopy (NIRS). However, there is a noticeable lack of studies about the standardization of gas sen-

* Corresponding author at: Institute of Bioengineering of Catalonia (IBEC), Barcelona, Spain. Tel.: +34 93 4034804; fax: +34 93 4021148.
E-mail addresses: lfernandez@el.ub.es, lfernrom@gmail.com (L. Fernandez).

http://dx.doi.org/10.1016/j.snb.2016.02.131
0925-4005/

sor arrays, with the exception of three important contributions. Balaban et al. [11] built a calibration model to identify the age of milk samples with a chemical sensor array of 12 conductive polymers. They transferred this model to a different array with the same sensors. To do so, they transformed the slave array response into master array readings by applying three different types of corrections: Univariate Regression, Multivariate Regression (MLR) and Multilayer Perceptrons (MLP). These calibration transfer methods were evaluated comparing the classification rates of the master and the transformed master arrays. Multivariate regression showed the best performance in standardizing the instruments. In a similar study, Tomic et al. [12] aimed at compensating the effect of sensor replacement in a hybrid sensor array composed of 12 MOS (metal-oxide semiconductor) sensors and 5 MOSFET (metal-oxide semiconductor field-effect transistor). The problem to solve was to distinguish between milk in good condition from off-flavor milk. They acquired twice the complete set of measurements, before and after the sensor replacement. Then they modeled the data of the old sensor array, which was selected as the master instrument. Measurements obtained from the new array were adapted to be used in the master classification model with two different techniques: Component Correction (CC) and Multiplicative Drift Correction (MDC). The later was shown to be slightly more efficient in rectifying the slave instrument response using the classification rate obtained for the test as a figure of merit. In a more recent paper, Carmel et al. [13] showed the possibility of building mappings between two different sensor technology arrays, a 32 conducting polymer array (CP) and an 8 sensor quartz microbalance module (QMB), which were exposed to a set of 23 pure chemicals. The authors built a PCA model for each instrument and tried to classify test samples according to the distance to the centroid of the nearest class. After that, they transformed the projected data from one sensor array to the other in both directions. To perform this task, they investigated three different approaches: Multivariate Regression (MLR, PCR, PLS), Neural Networks (NN) and Tesselation-based Linear Interpolation (TLI). Again, the classification rate was the figure of merit used to compare master and the standardized slave instruments. Their results showed that the performance of the different standardization methods was dependent on the mapping direction, obtaining the best results for the conversion from CP to QMB using NN, and applying TLI in the reverse mapping. In all these previous works the complete set of training samples used to create the data models was transferred from the master to the slave instrument.

Beyond these valuable contributions, we have identified three important open questions for calibration transfer in e-noses. (i) E-nose arrays can tune their operational parameters so as to enhance their sensitivity to different compounds [14]. Therefore, instrument dissimilarities due to tolerances on the operational parameters must be corrected accordingly. (ii) In order to make an efficient calibration transfer, a limited subset of experiments should be run in the slave instruments. To the best of our knowledge, no systematic study comparing the performance of different calibration transfer techniques with respect to the number of transfer samples is found in the literature for e-noses. (iii) Continuous calibration models (regressors) provide a more sensitive measure of the calibration transfer performance than discrete calibration models (classifiers). However, in the literature you can only find classification models transferred from one instrument to another.

In this paper, we address these three open questions with the following study. We have explored the calibration transfer problem for temperature modulated metal oxide sensor arrays when a global shift of temperature occurs (i). In an exhaustive study that includes 132 master-slave instrument combinations, we will evaluate the quality of the calibration transfer obtained from several instrument standardization techniques. We will compare master and slave errors (RMSEP) for different temperature shifts and sizes of the transfer sample set (ii) and on concentration prediction (iii).

2. Theory

In this paper, we follow two of the three different strategies proposed in the literature for calibration transfer [Marco and Gutierrez-Galvez [4]]. The first one consists in transforming the sensor responses of the slave instrument so they resemble those of the master instrument. In this way, we can directly use the calibration model built on the sensor responses of the master instrument with the transformed slave sensor responses. In this strategy, we work on the space of responses of the master instrument. To transform the sensor responses of the slave instrument, we used Direct Standardization (DS) and Piece-wise Direct Standardization (PDS). The second strategy consists of transforming not only the sensor responses of the slave instrument but also those of the master instrument to a joint master-slave space. Thus, the calibration model is built in this joint space. The sensor response transformation methods used in this strategy are Generalized Least Squares Weighting (GLSW) and Orthogonal Signal Correction (OSC). Fig. 1 illustrates both strategies.

In addition to this, we realized a sample subset selection to sort out the samples used to study the performance of the calibration transfer in terms of the number of samples considered from the slave instrument. We test two different approaches: select samples before or after creating the calibration model of the master instrument. Next, we describe the main features of the different calibration transfer techniques used in this paper, as well as the two methodologies used to perform sample subset selection.

2.1. Calibration transfer techniques

The purpose of calibration transfer is to correct instrumental differences so that the readings of the slave instrument (X_S) become similar to the readings of the master instrument (X_M). Each of the calibration transfer techniques employed in this work has been trained to perform this task using a subset of samples of the training set of master and slave instruments. These samples are usually called transfer samples S. This notation is employed in the description of the following four calibration transfer techniques.

2.1.1. Direct Standardization (DS)

Direct Standardization [15] is a calibration transfer technique that relates the readings of the slave instrument to those of the master according to the following linear transformation:

$$\bar{S}_M = \bar{S}_S \times F \tag{1}$$

where $\bar{S}_M$ and $\bar{S}_S$ are the mean-centered response matrices of transfer samples of master and slave instruments and F the slave-to-master transformation matrix, which is estimated as the product $\bar{S}_M$ and the pseudo-inverse of $\bar{S}_S$:

$$F = \bar{S}_S^{+} \times \bar{S}_M \tag{2}$$

In this way, new samples from the slave instrument X_S can be projected onto the master instrument response space X_M:

$$X_M^T = X_S^T \times F \tag{3}$$

2.1.2. Piece-wise Direct Standardization (PDS)

The DS method has the limitation of not properly transform the responses from slave to master instruments when the number of variables per sample is greater than the number of samples. Thus, the transformation matrix F (Eq. (2)) becomes underdetermined [7]. Piece-wise Direct Standardization [16] avoids this problem using local PLS models. It creates local linear models f_j that relate the

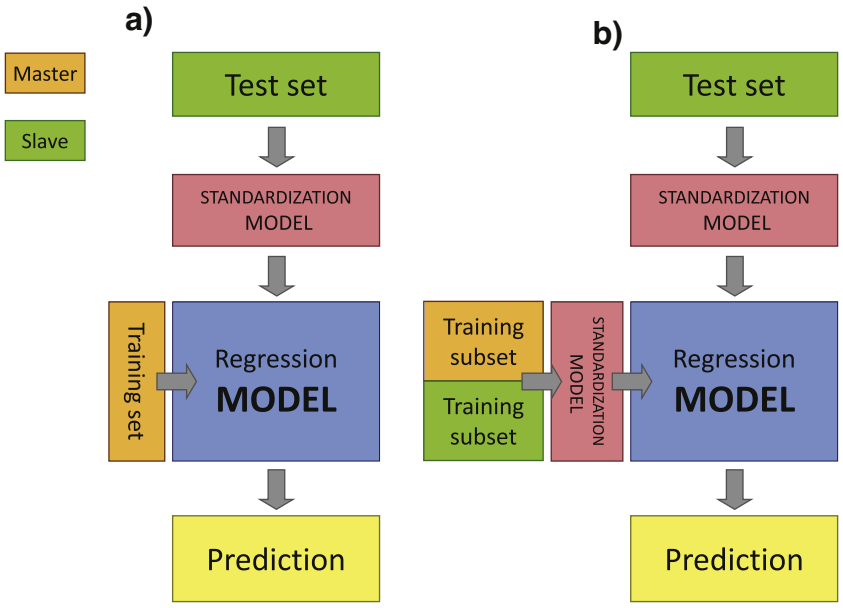

Fig. 1. Block diagram of the calibration transfer process (a) to transform the responses of the slave instrument so as to work on the space of responses of the master instrument (DS and PDS), and (b) to transform the responses of the master and slave instrument in order to work on a joint master-slave space (OSC, GLSW).

response of the master instrument variables within a window of size w centered at the *j-th* variable to the *j-th* variable on the slave array. The resulting transformation matrix for the method *F* has a diagonal structure:

$$F = diag\left(f_1^T \ldots f_2^T \ldots f_k^T\right) \tag{4}$$

where *k* is the number of variables on both instruments. The projection of data from the master onto the slave instrument is performed following Eq. (3).

2.1.3. Orthogonal Signal Correction (OSC)

Orthogonal Signal Correction [17] aims to remove the sources of variance of the slave instrument that are orthogonal to the master array. The OSC algorithm starts calculating the scores vector t_1 of the first Principal component of the slave array matrix of transfer samples, S_S. That vector is then orthogonalized against the master instrument response matrix of transfer samples S_M, giving raise to t_1'

$$t_1' = \left(1 - S_M \times \left(S_M^T \times S_M\right)^{-1} \times S_M^T\right) \times t_1 \tag{5}$$

After that, the weights w_1 of the product $S_S w_1$ are calculated for the maximum projection onto the orthogonal scores vector t_1':

$$w_1 = \bar{S}_S^+ \times t_1' \tag{6}$$

$S_S{}^+$ being the pseudo-inverse of S_S. The scores vector t_1 is then updated:

$$t_1 = S_S \times w_1 \tag{7}$$

Next, the algorithm returns to Eq. (5), where the determination of the orthogonal score vector is repeated until convergence. At this point, the loading vector corresponding to the first orthogonal score is computed as:

$$p_1 = S_S^T \times t_1 \left(t_1^T t_1\right)^{-1} \tag{8}$$

and the first OSC component can be removed from the original S_S matrix obtaining the deflated data matrix $S_{S,1}$:

$$S_{S,1} = S_{S,1} - t_1 p_1^T \tag{9}$$

Finally, the complete process can be repeated until the N-th Orthogonal Signal Component as follows:

$$S_{S,N} = S_S - \sum_{i=1}^{i=N} t_i p_1^T \tag{10}$$

2.1.4. Generalized Least Squares Weighting (GLSW)

Generalized Least Squares Weighting [18] method identifies and down-weights the instrument channels (features) responsible for the major sources of variance between master and slave instruments. To build the filter, the covariance matrix *C* from the difference between the mean-centered master and slave matrix of transfer samples is computed:

$$C = \left(\bar{S}_M - \bar{S}_S\right)^T \left(\bar{S}_M - \bar{S}_S\right) \tag{11}$$

Next, *C* is factorized as the product of three matrices through singular value decomposition (SVD):

$$C = VD^2V^T \tag{12}$$

where V and D are, respectively, the eigenvector and the singular value matrices. After that, the S matrix is weighted in the following way:

$$W = \sqrt{\frac{D^2}{\alpha} + I} \tag{13}$$

Being W the matrix of the weighted eigenvalues, α the weighting parameter and I the identity matrix. The parameter α controls the degree of dissimilarity allowed to the instruments. While high values of α increase the down weighting, lower values of α reduce its effect. The filtering matrix G is then calculated using the inverse of the weighted eigenvalues:

$$G = VW^{-1}V^T \tag{14}$$

2.2. Sample subset selection

Sample subset selection can be conducted in two manners: (a) by looking for the non-characteristic samples (with respect the

multivariate mean) on the master instrument matrix $\bar{X}_M$ through the calculus of the leverage matrix H [19]:

$$H = \bar{X}_M \bar{X}_M^T \quad (15)$$

And (b) by seeking for the most influential samples of the master's instrument calibration model, approaching H as the leverage matrix for the inverse calibration model $\bar{X}_{M^+}$, also mean-centered:

$$H = \bar{X}_M \bar{X}_M^+ \quad (16)$$

In both cases, the maximum diagonal element of H corresponds to the most relevant sample in the training set. Once the first sample is obtained, the rest of the dataset is orthogonalized against it, a new leverage matrix H is created, and the next most influential sample can be selected. Table 1 shows the first 12 samples selected using both methods.

3. Methods

3.1. Experimental

To perform this study, we used a set of three different types of Figaro metal oxide semiconductor sensors (TGS-2600, TGS-2610, TGS-2620) replicated 12 times each. In all experiments, one group of three different sensors was used as a master instrument to find a calibration model and the rest treated as slave arrays to study the calibration transfer. The read out of the sensors is performed through a load resistor (R_L = 6.1 KΩ) in a half bridge configuration. We modulated the sensor temperatures with a ramp profile ranging from ambient temperature to 495 °C ± 5 °C [20] in a period of 90 s. The 36 sensors were exposed during 900 s to 3 analytes (ethanol, acetone, butanone) at 7 different concentrations (0, 20, 40, 60, 80, 100, 120), giving rise to 21 different experiments. Detailed information on the odor delivery system and the estimation of sample concentration can be found in our previous work [21]. After each measurement block, the sensor chamber was cleaned in synthetic air over a period of 1800 s. Using this set of experiments, we built calibration models of the master instruments for the prediction of ethanol, acetone and butanone concentration. We acquired a different number of repetitions per experiment for training (7) and testing (3) the calibration models. Experiments with concentration levels of 0, 40, 80 and 120 ppm were acquired and used as a training set (3 pure analytes × 4 concentrations × 7 repetitions = 84 samples). Similarly, experiments with concentrations of 20, 60, 100 ppm were acquired and employed for testing the calibration models (3 pure analytes × 3 concentrations × 3 repetitions = 27 samples). The selected temperature window used for the calibration of the master instruments was [200–300] °C.

3.2. Calibration model

We have approached the calibration of our instruments as a regression problem to provide more sensitivity when transferring the calibration model to another instrument. In particular, we have used partial least squares regression (PLSR). We note that the PLSR model of the master instrument provides simultaneously a prediction for the concentration of ethanol, acetone and butanone of gas samples. We employed the set of training samples to generate the calibration models, whose level of complexity (i.e., the number of latent variables) was set through a cross-validation stage based on the Leave One Block Out (LOBO) approach. More specifically, each block of samples used for cross-validation belonged to one experiment type of the training set. Therefore, we employed 12 blocks of experiments, with 7 samples each. Basically, the LOBO method computes the Root Mean Square Error in Cross-Validation ($RMSECV_M$) as the average RMSE obtained from predicting each of the different blocks of experiments using a PLSR model built from the complementary blocks of experiments:

$$RMSECV_M = \frac{1}{C}\sum_{k=1}^{C}\sqrt{\frac{\sum_{i=1}^{N_V}\sum_{j=1}^{M}\left(\tilde{y}_{i,j,k} - y_{i,j,k}\right)^2}{N_V \times M}} \quad (17)$$

where $\tilde{y}_{i,j,k}$ and $v_{i,j,k}$ are, respectively, the observed and the predicted concentration values for the i-th sample, the j-th pure substance and the k-th data partition, N_V is the number of samples for testing each partition of the validation set (7), M the number of pure analytes present in the dataset (3) and C the number of blocks of experiments of the training set (12). The number of latent variables of the calibration model was determined calculating the $RMSECV_M$ (lv) for an increasing number of latent variables (lv from 1 to 10). When the current $RMSECV_M$ (lv = r) did not reduce the previous $RMSECV_M$ (lv = r − 1) value more than a 1%, the selected number of latent variables was determined lv = r − 1.

The measure of the model's performance fitting the test data for the master array was the Root Mean Squared Error of Prediction ($RMSEP_M$):

$$RMSEP_M = \sqrt{\frac{\sum_{i=1}^{N_T}\sum_{j=1}^{M}\left(\tilde{y}_{i,j} - y_{i,j}\right)^2}{N_T \times M}} \quad (18)$$

where $\tilde{y}_{i,j}$ and $y_{i,j}$ were, respectively, the observed and the predicted concentration values for the sample i-th sample, the j-th pure analyte, N_T is the number of samples of test set (27), M the number of pure analytes present in the dataset (3). The RMSEP was also used as a measure of goodness of fit for the transformed slave readings ($RMSEP_S$).

3.3. Calibration transfer

In this study, we have evaluated the ability of four techniques (DS, PDS, OSC, GLSW) to counteract the effect of temperature shift on calibration transfer. A series of experiments were conducted where the temperature of the slaves was shifted according to the following temperature values: ΔT = 0 °C, ±10 °C, ±20 °C, ±30 °C, ±40 °C, ±50 °C. Fig. 2 shows the dramatic change on MOX sensor waveforms due to temperature shifting (ΔT = −50 °C), for a temperature modulated TGS-2620 sensor exposed to the 3 test set ethanol concentrations. The effect of the number of transfer samples (from 1 up to 12) on the calibration transfer quality was studied, giving rise to a total of 17424 different calibration models transferred (12 masters × 11 slaves × 11 temperature shifts × 12 transfer samples) per instrument standardization technique. An example of the calibration transfer process is shown in figure (Fig. 3a–c) using Direct Standardization, for a temperature shifting of ΔT = −50 °C and 12 transfer samples. These figures show the scores plot of a PCA model for the master array (3a), the uncorrected slave array (3b) and corrected slave array (3c). Calibration transfer allows placing test samples back to its original position or nearby.

3.3.1. Calibration transfer models

We optimized the 4 calibration transfer methods minimizing the difference between the master and the corrected slave array readings. This procedure optimized the parameters of the different calibration transfer algorithms selecting them among a set of possible values. The range of parameter values depended on the particular technique employed. The window size w was selected from a list of 1–31 channels for PDS. The maximum value for w was limited to 31 so that allowed temperature shift correction between instruments without mixing features from different sensors of the master array. The range of the weighting parameter α in GLSW was selected taking into account different possible degrees of the dis-

Table 1
First 12 transfer samples of the calibration dataset selected using methods 1 and 2.

Transfer Samples	Method 1				Method 2			
	Sample Replicate	Concentration (ppm)			Sample Replicate	Concentration (ppm)		
		Eth	Acet	But		Eth	Acet	But
1	9	0	120	0	10	120	0	0
2	10	0	0	0	2	0	120	0
3	6	120	0	0	5	0	0	120
4	1	0	0	120	7	0	0	40
5	8	40	0	0	10	80	0	0
6	10	80	0	0	10	0	40	0
7	6	0	0	40	3	0	0	80
8	9	0	0	120	2	40	0	0
9	2	0	120	0	7	0	80	0
10	7	40	0	0	7	0	0	120
11	5	0	0	120	4	120	0	0
12	5	0	0	40	8	0	80	0

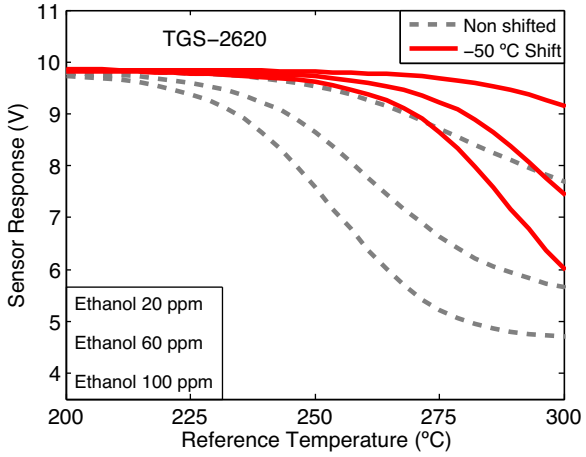

Fig. 2. Response of a TGS 2620 sensor unit to 20, 60, 100 ppm of ethanol within a nominal temperature window of 200–300 °C for a) no temperature shift (gray-dashed corves) and b) for a temperature shift of ΔT = −50 °C (red corves). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

similarity between master and slave instruments. Large values for α (close to 1) are needed for correcting instrumental dissimilarities whose impact in the overall variance of the data is comparable to the useful variance of the measurement. In practice, we selected α from the collection of values 1, 0.5, 0.01, 0.05, 0.001, 0.005. Finally, the number of Orthogonal Components, ncomp, in OSC was set in the range from 2 to 12. The reason for that was that the maximum value of this parameter was limited by number of transfer samples used to perform the calibration transfer between master and slave instruments (from 2 to 12). The validation started performing data correction for each technique and set of parameter values. The calibration model of the master instrument was then applied on the transformed training set of the slave instrument and the predictions of both instruments were compared. The comparison was performed through the calculation of the Root Mean Squared Error of Calibration ($RMSEC_{M\text{-}S}$):

$$RMSEC_{M-S} = \sqrt{\frac{\sum_{i=1}^{N_c}\sum_{j=1}^{M}\left(y_{i,j}^{M} - y_{i,j}^{S}\right)}{N_C M}} \quad (19)$$

where $y^{M}_{i,j}$ and $y^{S}_{i,j}$ are the predicted concentration values of the master and slave instruments for the sample *i*-th sample and the *j*-th pure substances, N_C (84) is the number of training samples and *M* the number of substances present in the dataset (3). The set

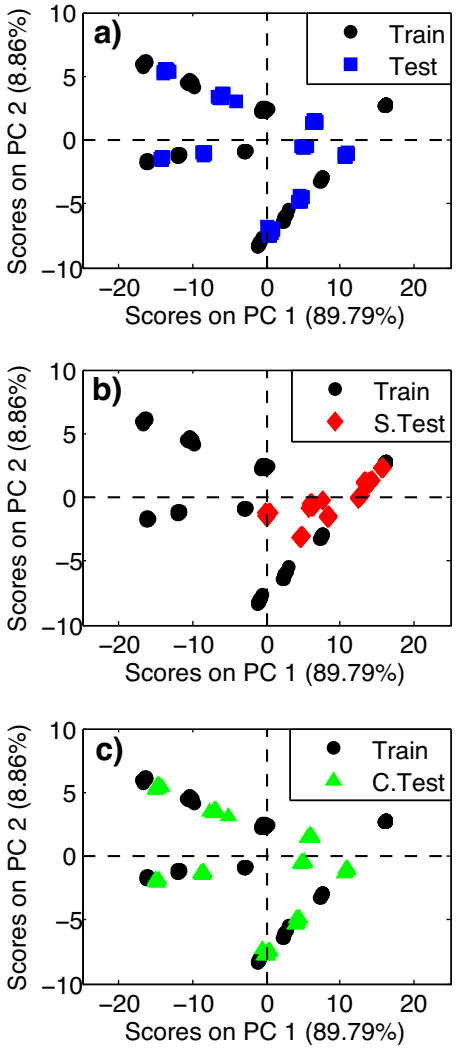

Fig. 3. (a–c) PCA plot of the sensor response for the training (black circles) and test sets with interleaved concentrations for: a) the master experiments (blue squares), (b) the uncorrected slaves (red diamonds) and (c) the corrected slaves after performing a Direct Standardization (green triangles), (ΔT = −50 °C). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

of parameter values whose $RMSEC_{M\text{-}S}$ was not able to be reduced in more than 1% by any other set was selected to build the calibration transfer model, for each calibration transfer algorithm. The

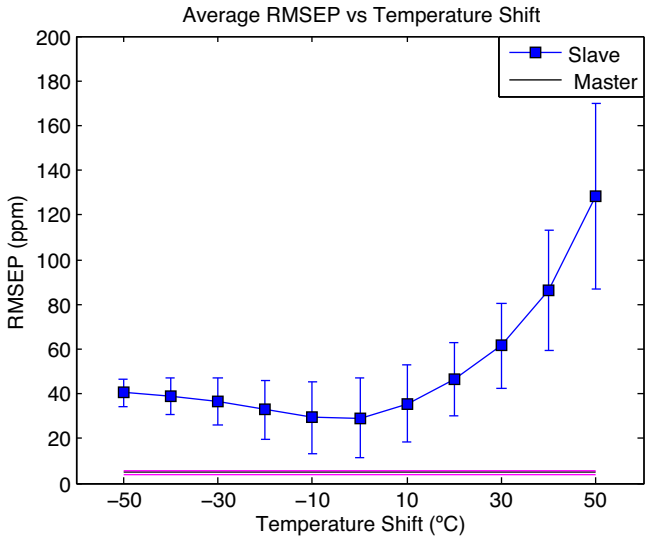

Fig. 4. Averaged $RMSEP_S$ as function of the temperature shift for the non-corrected slave instruments. Note as the worst predictions are biased toward positive temperature shifts.

RMSE was also employed as a measure of goodness of fit for the transformed slave readings of the test set ($RMSEP_{M\text{-}S}$).

4. Results

To gain some insight on the effect of temperature shift on calibration model transfer, we will show first results of the master calibration model applied directly on the slave without correction. This will provide a baseline performance from where to improve. Then, we will present the results of the slave RMSEP for an increasing number of transfer samples and also as temperature shifts varies in the range of [−50,50] °C. Finally, the results for the $RMSEP_{M\text{-}S}$ across all temperature shifts and number of transfer samples will provide a comprehensive picture of the performance of the different calibration transfer techniques.

In this study, each of the array replicates was used both as master instrument for the other replicates or as slave array to be corrected by another master array. When acting as master instruments, the array replicates produced similar calibration models in terms of complexity and model performance. Most of the array replicates built a 4 latent variable PLSR model (9 out of 12) whereas the remaining (3) needed 5 latent variables to achieve the specifications set for cross-validation. The average $RMSEP_M$ for the set of master instruments was (4.7 ± 1.1) ppm.

The direct application of the master calibration model in the slave arrays led, as anticipated, to high $RMSEP_S$. Fig. 4 shows the average prediction error of uncorrected slave arrays ($RMSEP_S$) along temperature shift, for all possible master-slave combinations. The $RMSEP_S$ was substantially higher than the $RMSEP_M$. The minimum difference between instruments was found when no temperature shift was produced ($RMSEP_{S|\Delta T=0} = 29.1 \pm 18.9$ ppm). As can be expected, the $RMSEP_S$ increased as the temperature shift between instruments increased. Though, this effect was not symmetric: shifts toward higher temperatures exhibited a greater penalty on the $RMSEP_S$ than shifts in the opposite direction. Comparing the most extreme temperature shifts in both directions we found that the error of prediction at $\Delta T = +50$ °C was $RMSEP_{S|\Delta T=+50\,^\circ C} = 128.2 \pm 41.4$ ppm, whereas at $\Delta T = -50$ °C was $RMSEP_{S|\Delta T=-50\,^\circ C} = 40.6 \pm 6.1$ ppm.

After data correction, the $RMSEP_S$ of the slave arrays was considerably reduced. The degree of error reduction depended on the amount of transfer samples and the shift of temperature. As a general trend, the $RMSEP_S$ decreased gradually until saturation as the number of transfer samples increased, for any temperature shift and calibration transfer technique. The influence of the transfer sample subset size on the quality of the calibration transfer is illustrated in Fig. 5(a–d). The figure shows the average $RMSEP_S$ of the corrected slave instruments of the different calibration transfer techniques, for an increasing number of transfer samples and a fixed temperature shift of $\Delta T = -20$ °C. DS and PDS obtained the lowest $RMSEP_S$ levels (6.3 ± 2.1 ppm, and 6.1 ± 1.4 ppm, respectively) although PDS needed a fewer number of samples to reach error saturation (five instead of eleven). OSC and GLSW showed higher $RMSEP_S$ values (around 8 ppm, for both techniques) and slower transitions to saturation. Concerning the influence of temperature shift, we found that the lowest $RMSEP_S$ were biased toward negative shifts, for any number of transfer samples and calibration transfer technique. However, PDS demonstrated to be the most robust technique against this direction-dependent effect. An example of this behavior can be seen on Fig. 6(a–d), where we show the average $RMSEP_S$ of the corrected slave arrays using the four instrument standardization methods, for the complete set of the temperature shifts, fixing to 5 the number of transfer samples. The minimum $RMSEP_S$ value for DS and PDS is obtained for a temperature shift of $\Delta T = -30$ °C (9.4 ± 4.0 ppm, and 6.2 ± 1.6 ppm, respectively). On the other hand, OSC and GLSW presented their minimum $RMSEP_S$ value for $\Delta T = 0$ °C (8.7 ± 2.8 ppm for OSC and 9.1 ± 3.3 ppm for GLSW).

Fig. 7(a–d) shows the color maps plots for the average $RMSEP_{M\text{-}S}$ of the transformed slaves, for each number of transfer samples, temperature shift and calibration transfer technique. Dark/light tones denote good/bad performances in correcting instrument dissimilarities. To enhance the contrast of plots, the prediction errors below 5 ppm and above 20 ppm were saturated, respectively, to black and white colors. To compare the quality of the different instrument standardization methods we evaluated the percentage of slave arrays with errors of prediction below 5 ppm in Fig. 7(a–d). According to this criterion, DS was able to correct properly around the 23% of these slave arrays. However, DS needed at least 5 transfer samples to obtain prediction errors below 20 ppm, and tended to present better corrections for slave arrays biased toward negative temperature shifts. Regarding OSC and GLSW, they exhibited a similar behavior in the sense that they experienced difficulties to correct the effect of temperature shifting. Note that both techniques needed at least 8 transfer samples to reduce the prediction error below 5 ppm, for the nearest temperature shift ($\Delta T = -10$ °C). In any case, none of them properly corrected more than a 15% of the slave arrays. Again, PDS presented the best performance, since the technique could cope better with the RMSEP error contribution due to the temperature shift direction. For instance, PDS only needed 5 transfer samples to correct slave arrays in the range of temperature shifts that goes from $\Delta T = -20$ °C to $\Delta T = 20$ °C. Consequently, provided the highest number of properly slave corrections (60% of the slave arrays).

5. Discussion

The reason why PDS performed better corrections than DS is that PDS creates local corrective models for each of the channels of the slave array, whereas DS generates a single global model, less flexible and more complex. This seems to be so also for OSC and GLSW. In addition to this, PDS detected which channels of the master array (within a window) were more correlated to the particular channel on the slave array, down-weighting the contribution to the correction of the non-important channels. As a consequence, the number of transfer samples between master and slave arrays needed to achieve the same error level tended to be lower for PDS.

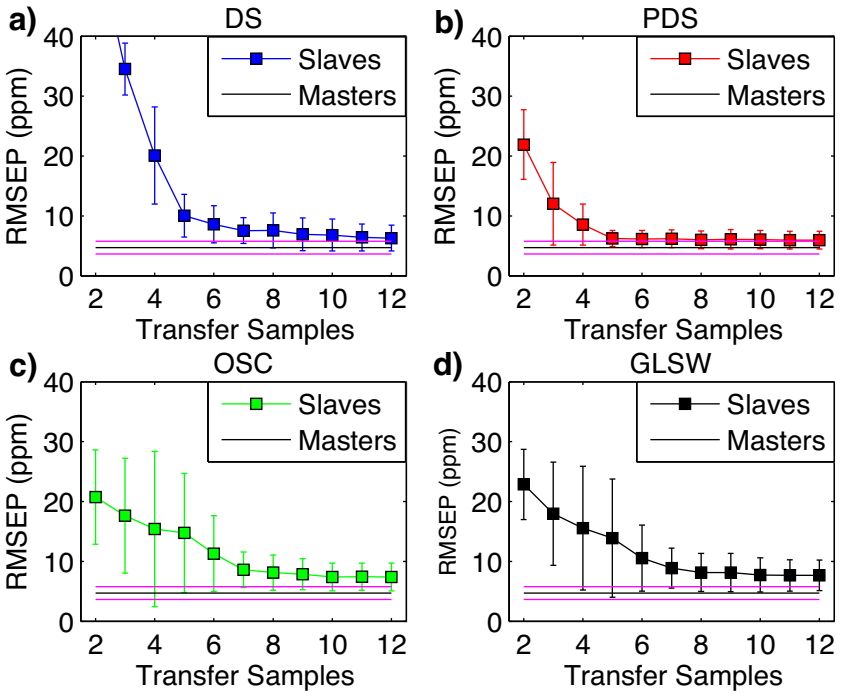

Fig. 5. (a–d) Average $RMSEP_S$ of the corrected slave instruments as function of the number of transfer samples, for a fixed temperature shift of $\Delta T = -20\,^{\circ}C$. Data correction was performed using a) Direct Standardization (blue-dotted line), b) Piece-wise Direct Standardization red-dotted line), (c) Orthogonal Signal Correction (green-dotted line), and (d) Generalized Least Squares Weighting (black-dotted line). The averaged $RMSEP_M$ is included in each of the plots with comparative purposes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

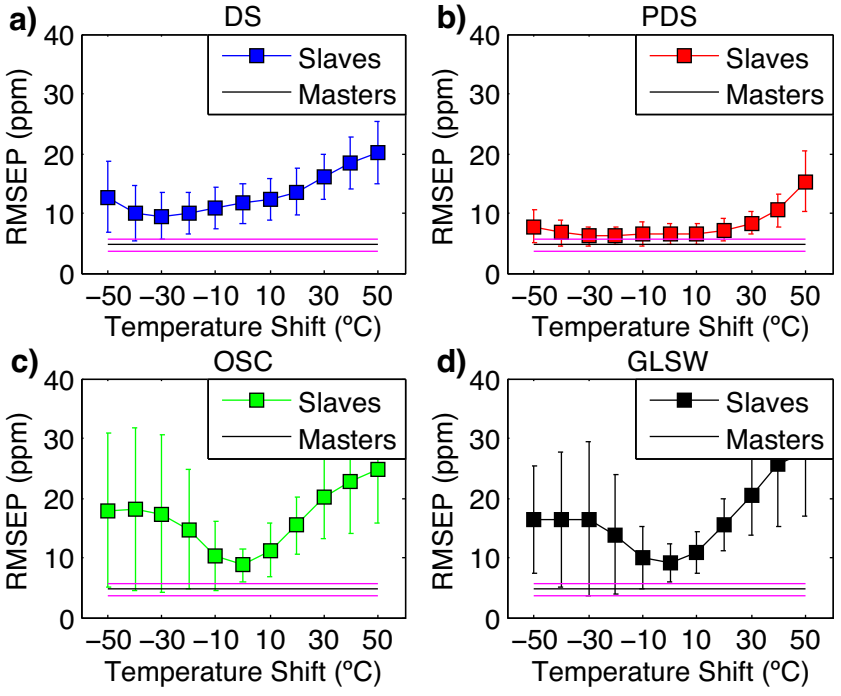

Fig. 6. (a–d) Average $RMSEP_S$ of the corrected slave instruments as function of the temperature shift, for a number of 5 transfer samples. Data correction was performed using (a) Direct Standardization (blue-dotted line), (b) Piece-wise Direct Standardization red-dotted line), (c) Orthogonal Signal Correction (green-dotted line), and (d) Generalized Least Squares Weighting (black-dotted line). The average $RMSEP_M$ is included in each of the plots with comparative purposes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

That suggests that the piece-wised extensions of OSC and GLSW may outperform the results obtained from the global versions of the algorithms, although this discussion is beyond the scope of this paper.

The performance of a calibration model with a high degree of complexity is directly related with the availability of a large number of samples. Effectively, as we know from Fig. 5(a–d), an increment on the number of transfer samples provides, up to a point, an enhancement of the corrected $RMSEP_S$. This improve-

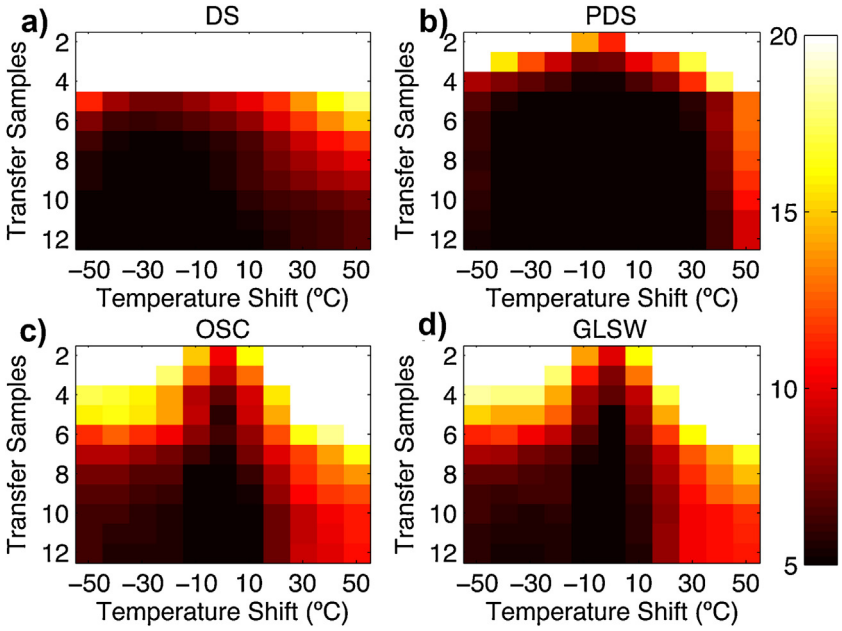

Fig. 7. (a–d) Average $RMSEP_{M\text{-}S}$ of the corrected slave instruments as function of the temperature shift, for each temperature shift and number of transfer samples. Data correction was performed using (a) Direct Standardization, (b) Piece-wise Direct Standardization, (c) Orthogonal Signal Correction, and (d) Generalized Least Squares Weighting.

Table 2a
Median, first quartile and third quartile of the optimized set of parameters used compensate for a temperature shift of −20 °C in the slave arrays, for a different number of transfer sample and calibration transfer technique.

Transfer Samples	PDS (w)			OSC (ncomp)			GLSW (α)		
	Q1	Median	Q3	Q1	Median	Q3	Q1	Median	Q3
2	1	9	11	2	2	2	0.1	0.1	0.1
3	3	9	12	3	3	3	0.1	0.1	0.1
4	5	9	17	2	4	4	0.01	0.1	0.1
5	8	12	19	2	5	5	0.01	0.05	0.1
6	7	13	21	2	6	6	0.01	0.05	0.05
7	7	15	23	3	5	7	0.01	0.01	0.05
8	9	15	23	3	5	8	0.005	0.01	0.05
9	9	15	27	3	6	9	0.005	0.01	0.01
10	11	17	27	3	6	8	0.005	0.01	0.01
11	13	21	27	3	6	8	0.005	0.01	0.01
12	13	21	27	3	6	8	0.003	0.01	0.01

Table 2b
Median, first quartile and third quartile of the optimized set of parameters used to correct the readings slave arrays, for the different temperature shifts and calibration transfer technique and fixing to 5 the number of transfer samples.

Temp. Shift (°C)	PDS (w)			OSC (ncomp)			GLSW (α)		
	Q1	Median	Q3	Q1	Median	Q3	Q1	Median	Q3
−50	17	19	29	5	5	5	0.01	0.05	0.1
−40	13	21	31	5	5	5	0.01	0.05	0.1
−30	9	21	27	5	5	5	0.01	0.05	0.1
−20	8	13	22	3	5	5	0.01	0.05	0.1
−10	7	11	20	2	4	5	0.01	0.05	0.1
0	5	9	13	2	3	5	0.01	0.01	0.05
10	5	7	13	3	5	5	0.01	0.05	0.1
20	7	9	11	2.5	5	5	0.01	0.1	0.1
30	11	13	15	3	5	5	0.01	0.1	0.1
40	11	13	16	5	5	5	0.01	0.1	0.1
50	15	13	19	5	5	5	0.01	0.1	0.1

ment is reflected on the structure of the calibration transfer models (Table 2a), where the parameters that govern the sample transformations are gradually modified until reaching saturation. The reason for error saturation on the corrected slave arrays can be deduced from the selection of the calibration transfer sample subset, shown in Table 1. Basically, for a certain number of selected samples we start to find samples that belong to a previously acquired category (substance and concentration). In consequence, no new information is added to the transfer models and the error of prediction for the corrected slave arrays cannot decrease significantly. The transition to error saturation is faster when the option for selecting the transfer samples is Method 2. That occurs because it includes a representative of each of the categories present on the training set (with the exception of the air samples) before adding sample replicates, while Method 1 discards three sample categories. In reference to the calibration transfer models, those methods that performed data correction before to build the calibration model (GLSW and OSC) exhibited their best results employing the sample subset Method 1, whereas those methods that applied data correction after the creation of the calibration model (DS and PDS) showed their best performance for the sampling subset Method 2.

A special comment deserves the asymmetry in sensor response with respect to temperature shift. Revisiting the results of Fig. 6(a–d) we observe that an increase on the temperature shift forces an increment on the model's specifications (see Table 2b). Interestingly, the asymmetry showed by the $RMSEP_S$ for opposite temperature shift positions is also present on the parameter values of all the calibration transfer techniques. This is in agreement with the results obtained in Fig. 4 for a direct calibration transfer between instruments shifted in temperature, where the higher prediction errors were found toward positive temperature shifts. The asymmetry on the error due to temperature shifting was produced because the uncorrected slave array response tends to saturate to the highest voltage level (10 V) for any substance and concentration, as review in Fig. 2. Projecting the response of test set samples of a slave array shifted toward negatives increments of temperature ($\Delta T = -50$ °C) on a PCA model built from the training set of a master array (Fig. 3b) we see that these samples approximate to the master array response to air. Taking that result as a reference we can estimate the lower bound for the uncorrected slave array substituting

the slave array samples by air measurements of the master array. That gives rise to a lower error bound around the 68 ppm. Toward positive temperature shifts, no saturation on the uncorrected slave array response is produced, so the test samples tend to spread on the PCA space and the error is continuously increasing.

6. Conclusions

In the present study, we showed that the effect of temperature shifts between homologous MOX sensor arrays leads to invalid calibration transfers featured with low predictive performance and direction-dependent error magnitudes. To overcome instrument dissimilarities the use of calibration transfer techniques is required. Among the four different calibration techniques used in this paper, the Piece-wise Direct Standardization procedure showed the best performance in reducing the slave array prediction error for any temperature shift direction and using fewer transfer samples. The main advantage of the PDS method lied in its ability to correct individually each one of the slave instruments channels through the use of multivariate local models. This results in a calibration transfer model with less complexity and more flexibility.

References

[1] J. Lin, Near-IR calibration transfer between different temperatures, Appl. Spectrosc. 52 (1998) 1591–1596.
[2] A.P. Lee, B.J. Reedy, Temperature modulation in semiconductor gas sensing, Sens. Actuators B: Chem. 60 (1999) 35–42.
[3] Y.D. Wang, D.J. Veltkamp, B.R. Kowalski, Multivariate instrument standardization, Anal. Chem. 63 (1991) 2750–2756.
[4] S. Marco, A. Gutiérrez-Gálvez, Signal and data processing for machine olfaction and chemical sensing: a review, IEEE Sens. J. 12 (2012) 3189–3214.
[5] B. Walczak1, E. Bouveresse, D.L. Massart, Standardization of near-infrared spectra in the wavelet domain, Chemometr. Intell. Lab. Syst. 36 (1997) 41–51.
[6] E. Bouveresse, D.L. Massart, Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration, Chemometr. Intell. Lab. Syst. 32 (1996) 201–213.
[7] R.N. Feudale, N.A. Woodya, H. Tana, A.J. Mylesa, S.D. Brown, J. Ferré, Transfer of multivariate calibration models: a review, Chemometr. Intell. Lab. Syst. 64 (2002) 181–192.
[8] M. Padilla, A. Perera, I. Montoliu, A. Chaudhry, K. Persaud, S. Marco, Drift compensation of gas sensor array data by orthogonal signal correction, Chemometr. Intell. Lab. Syst. 100 (2010) 28–35.
[9] J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, S. Wold, An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra, Chemometr. Intell. Lab. Syst. 44 (1998) 229–244.
[10] Q. Fu, J. Wang, G. Lin, H. Suo, C. Zhao, Short wave near-Infrared spectrometer for alcohol determination and temperature correction, J. Anal. Methods Chem. 2012 (2012) 7, pages.
[11] M.O. Balaban, F. Korel, A.Z. Odabasi, G. Folkes, Transportability of data between electronic noses: Mathematical methods, Sens. Actuators B: Chem. 71 (2000) 203–211.
[12] O. Tomic, T. Eklov, K. Kvaal, J.E. Haugen, Recalibration of a gas-sensor array system related to sensor replacement, Anal. Chim. Acta 512 (2004) 199–206.
[13] O. Shaham, L. Carmel, D. Harel, On mappings between electronic noses, Sens. Actuators B: Chem. 106 (2005) 76–82.
[14] A. Hierlemann, R. Gutierrez-Osuna, Higher-Order Chemical Sensing, Chem. Rev. 108 (2008) 563–613.
[15] R. Tauler, B. Walczak, S.D. Brown, Comprehensive Chemometrics, Elsevier, 2009.
[16] Y.D. Wang, M.J. Lysaght, B.R. Kowalski, Improvement of multivariate calibration through instrument standardization, Anal. Chem. 64 (1992) 562–564.
[17] T. Fearn, On orthogonal signal correction, Chemometr. Intell. Lab. Syst. 50 (2000) 47–52.
[18] H. Martens, M. Høy, B.M. Wise, R. Bro, P.B. Brockhoff, Pre-whitening of data by covariance-weighted pre-processing, J. Chemometr. 17 (2003) 153–165.
[19] M.A. Sharaf, D.L. Illman, B.R. Kowalsky, Chemometrics, John Wiley& Sons, New York, 1986, pp. 239.
[20] A.P. Lee, B.J. Reedy, Application of radiometric temperature methods to semiconductor gas sensors, Sens. Actuators B: Chem. 69 (2000) 37–45.
[21] J. Fonollosa, L. Fernández, R. Huerta, A. Gutiérrez-Gálvez, S. Marco, Temperature optimization of metal oxide sensor arrays using Mutual Information, Sens. Actuators B: Chem. 187 (2013) 331–339.

Biographies

Luis Fernandez is a Ph.D. student at the Department of Electronics of the University of Barcelona. He received a B.S. in Physics (2005) and a B.S. in Electrical Engineering (2011) from the University of Barcelona. His current research topic is bio-inspired large sensor arrays based on metal oxide sensors.

Selda Guney received the M.Sc. and Ph.D. degree in electrical and electronics engineering from Karadeniz Technical University of Trabzon, Turkey in 2007 and 2013 respectively. Currently, she is an Assistant Professor with the Department of Electrical and Electronics, Baskent University. Her research interests include, control systems, pattern recognition and signal processing of gas sensor arrays.

Agustin Gutierrez-Galvez received the B.E. degree in physics and electrical engineering from the University of Barcelona, Catalonia, Spain, in 1995 and 2000, respectively. He received the Ph.D. degree in computer science from Texas A&M University, College Station, in 2005. He was a JSPS Post-Doctoral Fellow with the Tokyo Institute of Technology, Tokyo, Japan, in 2006, and came back to the University of Barcelona with a Marie Curie Fellowship. Currently, he is an Assistant Professor with the Department of Electronics, University of Barcelona. His current research interests include biologically inspired processing for gas sensor arrays, computational models of the olfactory systems, pattern recognition, and dynamical systems.

Santiago Marco received the Degree in applied physics and the Ph.D. degree in microsystem technology from the University of Barcelona, Catalonia, Spain, in 1988 and 1993, respectively. He held a Human Capital Mobility Grant for a post-doctoral position at the Department of Electronic Engineering, University of Rome "Tor Vergata," Rome, Italy, in 1994. Since 1995, he has been an Associate Professor with the Department of Electronics, University of Barcelona. In 2004, he had a sabbatical leave at EADS-Corporate Research, Munich, Germany, where he was involved in ion mobility spectrometry. He has recently been appointed leader of the Artificial Olfaction Laboratory, Institute of Bioengineering of Catalonia, Barcelona, Spain. His current research interests include the development of signal/data processing algorithmic solutions for smart chemical sensing based in sensor arrays or microspectrometers integrated typically using microsystem technologies.