Research Article

Orthogonal Signal Correction to Improve Stability Regression Model in Gas Sensor Systems

Rachid Laref, Diaa Ahmadou, Etienne Losson, and Maryam Siadat

University of Lorraine, LCOMS, EA7306, 57070 Metz, France

Correspondence should be addressed to Maryam Siadat; maryam.siadat@univ-lorraine.fr

Metal oxide sensors are the most often used in electronic nose devices because of their high sensitivity, long lifetime, and low cost. However, these sensors suffer from a lack of response stability making the electronic nose systems useless in industrial applications. The sensor instabilities are particularly caused by incomplete recovery process producing gradual drifts in the sensor responses. This paper focuses on a signal processing method combining baseline manipulation and orthogonal signal correction technique in order to reduce effectively the drift impact from the sensor outputs. The proposed signal processing is explored using experimental data obtained from a gas sensor array responding to various concentrations of pine essential oil vapors. Partial Least Square method is then applied on the corrected dataset to establish a regression model for the estimation of gas concentration. In this work, we show essentially how our drift correction approach can help to improve significantly the stability of the regression model, while ensuring good accuracy.

1. Introduction

Gardner defines the electronic nose (E-nose) as "an instrument, which comprises an array of electronic chemical sensors with partial specificity and an appropriate pattern-recognition system, capable of recognizing simple or complex odors" [1]. The metal oxide sensors (MOX) are the most used in this instrument because they are very sensitive to many gases, are commercially available, have a long lifetime, and have low cost [2]. However, these sensors show a lack of reproducibility (or instability) which limits the translation of laboratory results to industrial applications [3]. Instability of sensor responses can be due to several problems. (1) Drifts: sensor response signals always tend to show a small variation even if the E-nose is exposed to same gas and concentration under constant environmental conditions. Many reasons can explain this variation: (i) sensor aging due to thermomechanical fatigue after successive gas expositions [4] and (ii) sensor poisoning because of the exposure to high concentration or to an aggressive chemical or silicone vapors [5]. (2) Environmental disturbances: the humidity variation and temperature or pressure fluctuations change

also the sensor responses [6]. (3) Sampling condition: the mechanism of MOX technology, which is the exchange of oxygen molecules between gas mixture and metal film, makes the acquisition cycles very long, since in many applications the steady-state responses of sensors are never reached [7].

A lot of correction methods have been investigated to improve sensor response stabilities; they are based on different approaches: univariate or multivariate methods. In univariate technique the correction is applied on each sensor individually. Among these methods the baseline manipulation, largely used in industry [8], consists of transforming a sensor response with use of the initial response value. Three kinds of transformation can be made to correct the baseline of sensor signals: differential, relative, and fractional corrections. Generally baseline manipulation is used as preprocessing of the sensor output [9]. Filtering techniques like Fourier bandpass filter, moving median filter, or discrete wavelet transform have been also utilized by many researchers in this field to remove drift effects from measurements [10, 11]. Among the above-mentioned methods, discrete wavelet transform is more flexible because it can analyze the signal at different frequency bands with different resolutions.
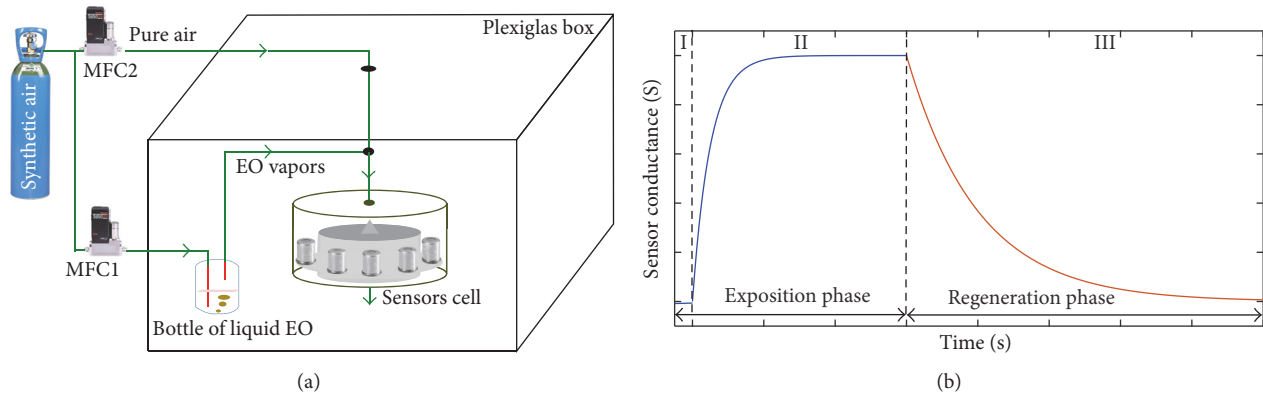
FIGURE 1: (a) Schematic representation of the experimental set-up and (b) typical temporal response of a MOX gas sensor during exposition and regeneration phases.

In case of E-nose measurements, since the drift effects are correlated, the multivariate methods allow capturing more information from all the sensors permitting modeling more complex or nonlinear drift effects [12]. In the literature around the E-nose research different multivariate methods can be found. For example, adaptive neural network methods like Self-Organizing Maps [13] show good performances; but they are limited to gas classification applications and in case of gas quantification it would be hard to obtain good results [14].

When gas sensors are exposed to the same gas under the same sampling and environmental conditions, any changes in the sensor response are related essentially to drift. In order to reduce the variance of this drift that follows one direction, the multivariate linear correction methods based on partial least square PLS or Principal Component Analysis, PCA, constitute the best approach [6, 15, 16]. Among these methods orthogonal signal correction (OSC) has been chosen because different studies have proven that this method is the most efficient [16–18].

The orthogonal signal correction (OSC) was proposed firstly by Wold for NIR spectra correction [19], and then several algorithms were published to improve its performance. The main idea of OSC technique is focused on removing the variance that is not correlated to the variable to estimate.

In this study, we combine baseline manipulation with OSC technique in order to remove the drift effects on a MOX sensor array. Then PLS is used to model the behavior of our gas sensors responding to different concentrations of pine essential oil vapors (EO) diluted in pure air. The combination of this correction approach gives a good quantification of essential oil vapors by using only a few components of PLS in the modeling.

2. Materials and Methods

The data used in this work are obtained from a home-made experimental equipment mainly composed of a gas sensor cell and an EO vapor diffuser (Figure 1). The sensor cell contains seven metal oxide gas sensors (TGS882, TGS2620,

SP31, SPAQ1, MQ3, and MQ138 produced by Figaro, FIS, and Hanowei companies) and the diffuser unit uses an air bubbling system in liquid EO. So, the concentration of the EO in pure air is controlled by the flow rates of two mass flow controllers, MFC1 and MFC2. The total flow rate (MFC1 + MFC2) through the sensor cell is maintained constant, which also allows generating different EO concentrations by varying the percentage of the MFC1 flow rate over the total flow rate. The sensors cell and liquid EO bottle are placed in Plexiglas chamber to keep the sensors in constant climatic conditions [20].

2.1. Measurement Protocol. According to our previous study [20], we performed series of measurements on different EO concentrations. Each measurement comprises an EO exposure phase followed by a sensor cleaning phase through dry synthetic air. The response type of MOX sensors during this measurement cycle is presented in Figure 1(b): the sensor conductance increases during the gas exposition and then decreases during the cleaning process through its previous baseline. Gas exposition time is fixed at 75 seconds to provide a quantifiable response of the sensors, and cleaning time is set to 350 s to permit an acceptable recovery of the sensor sensitive element. For the learning measurements, we have created nine constant concentrations in the range of 0.5% to 4.5%, producing pleasant odor for aromatherapy, with a step of 0.5%.

Sensor outputs are digitalized, filtered, and then recorded every second in terms of sensor conductance values. We have opted to express the sensor response in conductance rather than resistance because this parameter is more efficient when gas concentrations identification with n-type semiconductor metal oxide sensors is demanded [1].

Forty measurements have been realized for each of the nine EO concentrations randomly selected throughout the experiments. Each sensor output is characterized by 425 recorded points (75 points during gas exposition, 350 points at recovery process). Data are arranged on a dataset formed from 2975 columns {425 data ∗ 7 sensors} and 360 rows {40 measurements ∗ 9 concentrations}.
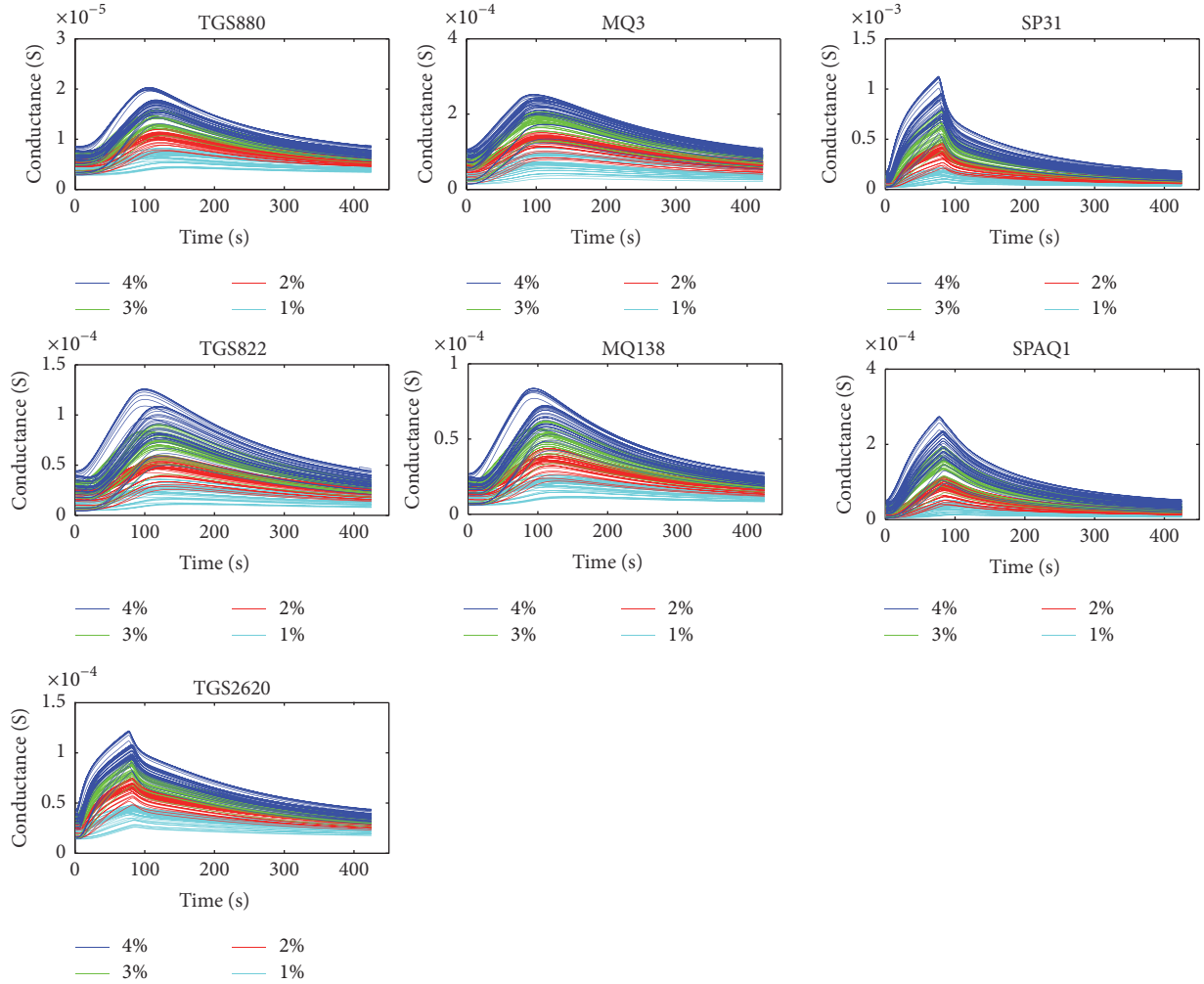
FIGURE 2: Temporal responses at 1 to 4% EO concentrations for each of the seven gas sensors (TGS880, TGS822, TGS2620, MQ3, MQ138, SP31, and SPAQ1).

2.2. Instability of the Measurements. To illustrate the instability of the temporal responses, we have grouped on the same axis the signals for 1, 2, 3, and 4% EO concentration of each sensor (Figure 2). We can easily reach the conclusion of an important disparity of the sensor responses obtained for the same EO concentration and the same conditions. This instability can be explained by the sensor drift which is mainly due to incomplete recovery process. In Figure 3, we have reported the temporal responses of TGS2620 sensor during several successive measurements at different concentrations. In this figure, the observation of the different baselines (sensor conductance at the beginning of a measurement) shows that the baseline value depends highly on the gas concentration used in the previous measurement. As shown in Figure 3, starting a measurement at 3% EO, the sensor baseline is 26 μS when following a measurement at 2% and 35 μS if the previous experiment was at 4% EO. Moreover, Figure 3 highlights that this drift alters not only the baseline but also the sensitivity of the sensors.
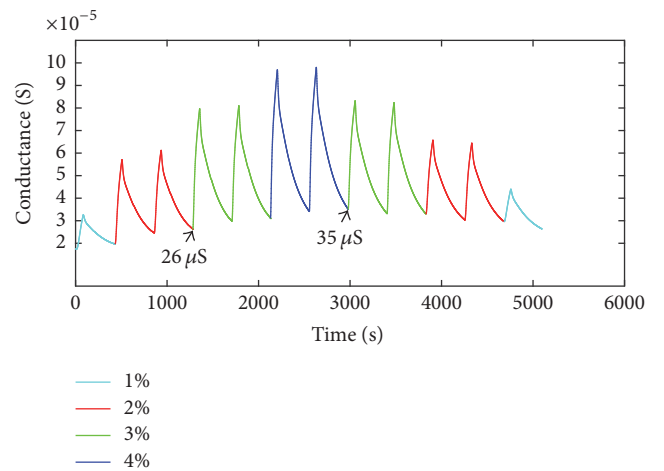

FIGURE 3: Illustration of the TGS2620 gas sensor instability in terms of its initial conductance during successive measurements by varying the EO concentrations.

This elementary comparison confirms the nonefficiency of the sensor recovery process causing noticeable drifts. In fact, these drifts will be greater in case of real-time and continuous measurement with an E-nose.

3. Drift Correction and Calibration

We have grouped all the sensor responses $G(t)$ in a dataset which is represented by matrix of 396 rows and 2975 columns, where rows correspond to the observations and columns to the corrected temporal response of sensors. Concentrations of the various experiments are grouped in vector Y (396).

3.1. Drift Correction. The reliability of E-nose results depends strongly on how the sensor outputs are treated particularly to minimize noises and drift affects (shown in previous section). So, the signal processing has a key role in E-nose performance and many studies were already done on this subject.

3.1.1. Baseline Manipulation. Baseline manipulations are very often cited in literature to remove the drift effects on sensor responses [9, 10]. Usually, baseline corrections are based on the initial response of a sensor. In this work, we have opted for sensor conductance (G_{final}) obtained at the end of the precedent cleaning process [21] and fractional correction on temporal response $G(t)$, as shown in

$$G_c(t) = \frac{G(t) - G_{final}}{G_{final}}. \tag{1}$$

This manipulation is made for temporal signals of each sensor. We observe only a very light amelioration concerning the disparity of the sensor responses obtained at same measurement condition and gas; so the concentration discrimination is still not conceivable (feasible).

For further development of the drift correction, the dataset named X containing corrected data after baseline manipulation is used.

3.1.2. Orthogonal Signal Correction (OSC). Prior to applying the regression modeling, we followed the baseline manipulation of OSC technique to reduce more efficiently the drift effects from the sensor signals. We show that the use of this correction technique improves the calibration processes making it reliable and stable.

The main objective of the OSC technique is to remove the variance which is not correlated to the variation of concentration Y. This procedure is done by the suppression of nonrelevant information of gas response in matrix X. So, only information orthogonal to Y is removed, and this condition will guarantee that the information useful for the calibration is largely saved.

The algorithm for OSC [19] is based on the following steps:

(i) Use Principal Component Analysis (PCA) to decompose X into scores t.

(ii) Orthogonalize the first score t_1 (first component containing maximum of information) to Y in order to obtain new score t_n as follows:

$$t_n = \left(1 - Y\left(Y^T Y\right)^{-1} Y\right) t_1, \tag{2}$$

where Y^T is the transpose of Y.

(iii) Calculate weight vectors W as follows: $W = X^{-1} t_n$ using PLS.

(iv) Calculate the new score t, $t = XW$.

(v) Compute the loading vector P, $P^T = t^T X/(tt_n)$.

(vi) The new corrected matrix X_{corr} is given by

$$X_{corr} = X - tP^T. \tag{3}$$

In order to test the benefits of the OSC technique, the new dataset (X_{corr}) composed of 396 observations is divided into "training set" which contains 75% of observations and the "test set" (25% of observations). Both of these datasets cover the concentration range.

For each gas sensor, applying OSC technique makes the responses at the same EO concentrations more similar. Figure 4 shows the temporal responses of the test set after OSC treatment for the 1, 2, 3, and 4% EO concentrations. Comparing Figures 2 and 4, we can observe that the dispersion of temporal responses at same concentration is significantly attenuated. Figure 4 also highlights how it could be easier to discriminate the different EO concentrations after temporal signal correction with OSC.

For better perception of the OSC impact on gas quantification, we have plotted in Figure 5 PCA scores of all our data before and after the correction with OSC technique. PCA plot of the dataset (before OSC) confirms the impossibility of concentration discrimination, but we can clearly observe on Figure 5(b) the improvement brought by OSC correction which permits a successful separation of all the EO concentrations, even with a very small step (0.5%).

3.2. Calibration. As the predictors in our dataset are highly correlated and their number is very large (number of columns) by comparison with the number of observations (number of lines), the use of multiple linear regression (MLR) model is not suitable because of the existing multicollinearity [19]. To deal with the multicollinearity problem, regression model should be performed on independent variables. So, we have utilized the partial least square (PLS) analysis to find independent components that can explain as much as possible the covariance between X and Y. These components can be used for regression modeling and guaranteed a good prediction [22]. However regression methods can suffer from the overfitting or the underfitting: if we take a large number of components, the model shows poor performance for the recognition of new data, and the use of a small number of components may not be sufficient to reach a good precision. Then, the number of components should be optimized.

We have performed the calibration of our E-nose by using PLS regression as recognition method. In this aim, dataset
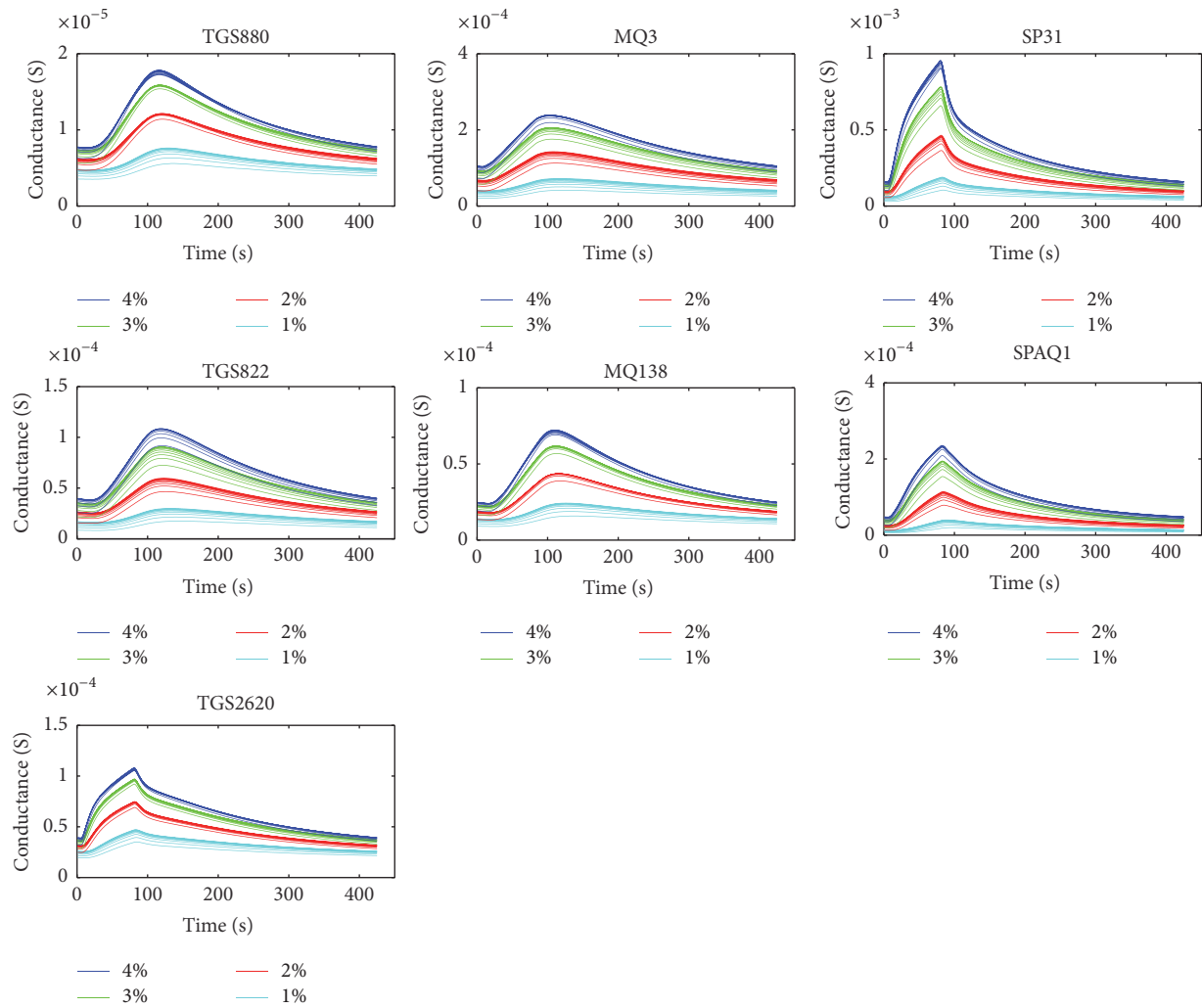
FIGURE 4: Temporal responses at 1 to 4% EO concentrations after using OSC technique for each of the seven gas sensors (TGS880, TGS822, TGS2620, MQ3, MQ138, SP31, and SPAQ1).
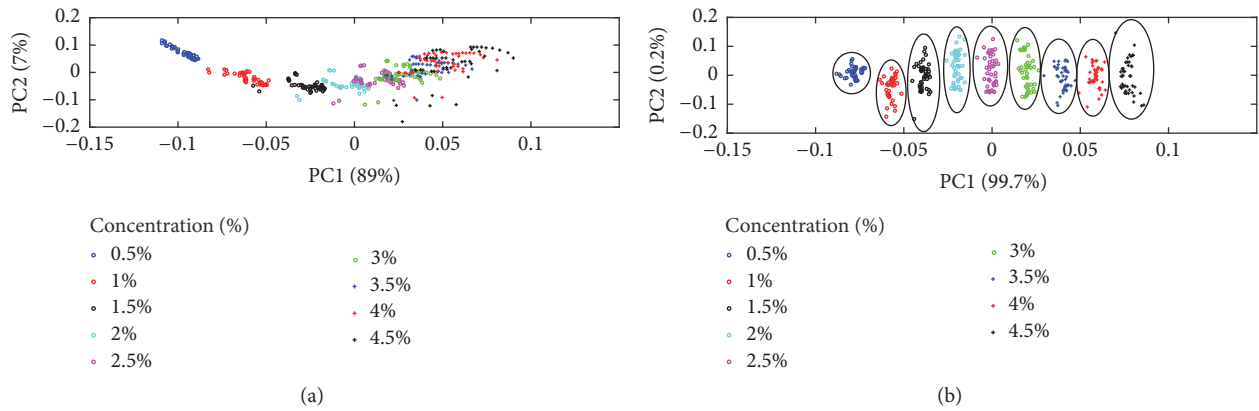

FIGURE 5: PCA score plots: (a) before OSC correction and (b) after OSC correction.

Table 1: CV values of regression coefficient and TOTAL_RMSE along with number of OSC + PLS components.

CV	Number of variables						
	1	2	3	4	5	6	7
A_0	**0,0430**	0,0476	0,0470	0,0474	0,0474	0,0470	2,5036
A_1	**0,0020**	0,0135	0,0208	0,0231	0,0299	0,0352	2,5372
A_2	**0,0015**	0,0051	0,0141	0,0124	0,0060	0,0072	−2,5057
A_3	**0,0019**	0,0113	0,0105	0,0110	0,0099	0,0093	2,6227
A_4	**0,0022**	0,0096	0,0316	0,2056	−0,1426	−0,1391	2,4622
A_5	**0,0015**	0,0198	0,0248	0,0242	0,0179	0,0139	−2,5424
A_6	**0,0026**	0,0096	0,0127	0,0144	0,0379	0,0394	−2,4941
A_7	**0,0020**	0,0051	0,0122	0,0228	0,0572	0,0348	−2,5388
TOTAL_RMSE	**0,2454**	0,2449	0,2447	0,2446	0,2446	0,2446	0,494

Table 2: CV values of regression coefficient and TOTAL_RMSE along with number of PLS components.

| CV	Number of variables						
	1	2	3	4	5	6	7
A_0	−0,0176	0,0674	0,1342	0,0479	0,1913	0,0839	**0,0573**
A_1	0,0030	0,2424	0,3032	0,0754	0,0282	0,01317	**0,0130**
A_2	0,0023	0,0614	0,0403	0,02821	0,1543	−0,4366	**−0,3317**
A_3	0,0028	0,0456	0,4243	−0,2263	0,0503	0,04381	**0,0316**
A_4	0,0026	−0,8605	2,1887	0,1037	9,1080	0,0993	**0,0740**
A_5	0,0019	−0,0188	−0,0463	−0,02986	−0,0257	−0,01194	**−0,0111**
A_6	0,0033	−0,1652	−0,3465	0,06188	0,3475	−0,1358	**−0,0810**
A_7	0,0024	−0,0341	−0,2107	−0,0621	−0,0203	−0,0167	**−0,0133**
TOTAL_RMSE	0,4442	0,3725	0,3569	0,30796	0,2522	0,24428	**0,2433**
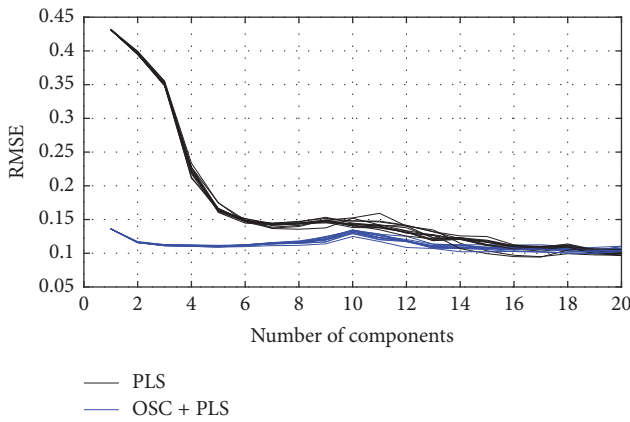
Figure 6: Plot of RMSE versus number of PLS or OSC + PLS components.

is divided into 12 segments with 33 observations each: one segment is used as data test to evaluate the performance of the calibration model when the other segments (11 segments) are used as learning data. To obtain a valuable indicator based on cross validation, this operation is performed 12 times, and successively the data test is changed. So, each of the 396 observations becomes a predicted one. The RMSE was calculated by averaging the twelve RMSEs.

In Figure 6 we illustrate the value of the RMSEs versus of number of components in the case of performing PLS on raw data or performing PLS on OSC corrected data. We can observe that both PLS and PLS + OSC give a good accuracy but PLS uses 18 components to reach RMSE = 0.1% while OSC + PLS needs only one component.

3.2.1. Feature Selection. To investigate the stability of the pattern, we compare between the variability of regression coefficients obtained by PLS or OSC + PLS. Each sensor output is characterized by 425 points; hence the number of regression coefficient is $(425 * 7 + 1)$ making the comparison extremely challenging. To reduce this dimensionality we decide to use the average of the 425 points of a sensor signal as characteristic feature [23]. So we only have $(7 + 1)$ coefficients to compare.

3.2.2. Stability Test. The coefficient of variation (CV) is calculated for all the regression coefficients in the two cases (PLS analysis or OSC + PLS analysis) for seven cycles. In the first cycle we started to build a model using one component and we added one more component in each cycle until we used all of them (7 components) in the final cycle. Dataset was divided on 12 subsets, so in one cycle we have calculated 12 times the regression coefficients and the RMSE. At the end of each cycle, CVs are calculated as the ratio of the standard deviation over the mean value of each coefficient and TOTAL_RMSE as the average of twelve RMSE. These results are presented in Tables 1 and 2.
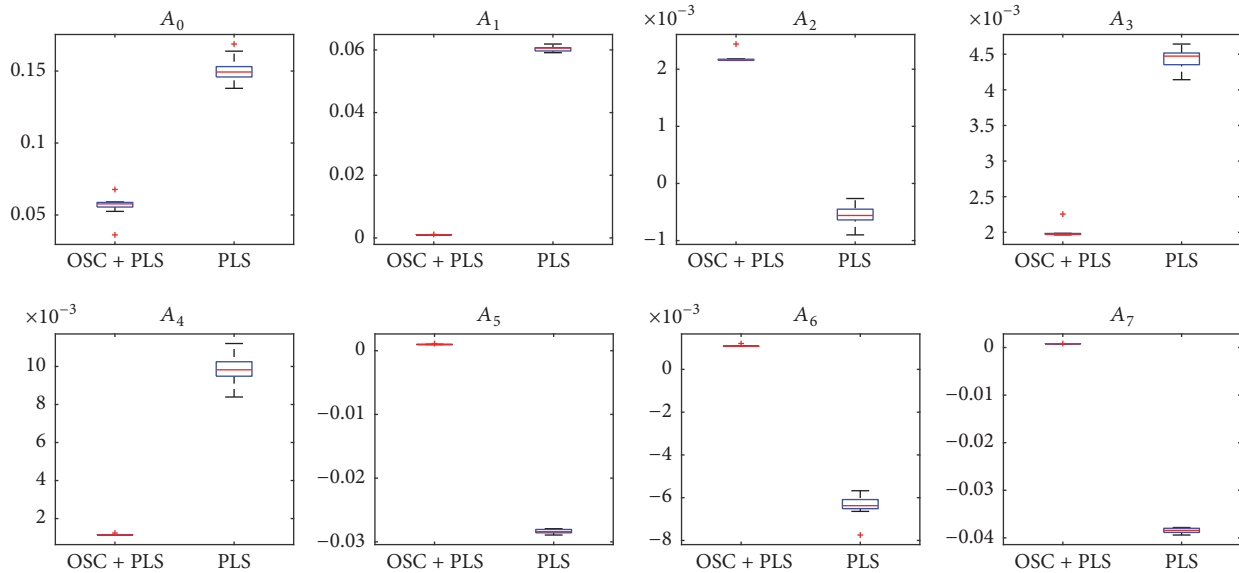
FIGURE 7: Boxplots of the distributions of the coefficients regression value for OSC + PLS and PLS models.

As we can see, in the case of OSC + PLS the CVs of the regression coefficient are approximately 10 times lower than those obtained in the case of applying only PLS. Moreover, to obtain the best result for TOTAL_RMSE, we need 1 component if applying OSC and 7 components without OSC.

Consequently, we have chosen the model with one component for OSC + PLS and the model with 7 components for PLS because they give the best results, and also because they have approximatively the same RMSE allowing us to compare the stability of regression coefficients independently. This statement is confirmed in Figure 7 which shows the distribution Boxplot of the regression coefficients using PLS and OSC + PLS. Distribution boxes are very narrow in the case of using OSC + PLS, highlighting a very poor disturbance.

For better comparison of regression coefficients stability we have plotted on the same figure the absolute value of CV; Figure 8 shows that the CV of regression coefficients of OSC + PLS model is ten times less than those of PLS model.

4. Conclusion

The main challenge in E-nose field is based on sensor signal processing, particularly to correct the gas sensor drift affects. As a first step, a fractional baseline correction is suggested by the use of a reference value corresponding to the sensor conductance taken at the end of the cleaning phase (G_{final}). Afterwards, PLS regression was combined with orthogonal signal correction to improve regression modeling performances for gas quantification. In this paper, the proposed signal processing is performed on a dataset obtained from experiments carried out on various dilutions of pine EO vapors in dry synthetic air. The Principal Component Analysis of the dataset shows clearly how signal treatment with OSC technique is essential for the EO concentration discrimination. Regarding the performances of the regressing
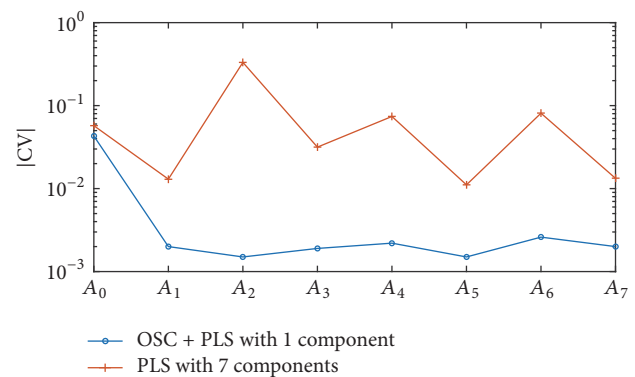

FIGURE 8: Magnitude of CV for the regression coefficients of OSC + PLS and PLS model.

model, the two methods PLS and OSC + PLS were compared in terms of stability and accuracy. Each of the two methods gives high accuracy, but PLS without OSC needs more variables to reach the same performance as OSC + PLS. We have investigated the stability of the regression model by comparing the variability of regression coefficients of these two methods. Our results show incontestably that for the same accuracy OSC + PLS model has high robustness comparing with PLS model.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. W. Gardner and P. N. Bartlett, "A brief history of electronic noses," *Sensors and Actuators B: Chemical*, vol. 18, no. 1–3, pp. 210-211, 1994.

[2] A. D. Wilson and M. Baietto, "Applications and advances in electronic-nose technologies," *Sensors*, vol. 9, no. 7, pp. 5099–5148, 2009.

[3] S. Marco and A. Gutiérrez-Gálvez, "Signal and data processing for machine olfaction and chemical sensing: a review," *IEEE Sensors Journal*, vol. 12, no. 11, pp. 3189–3214, 2012.

[4] R. K. Sharma, P. C. H. Chan, Z. Tang, G. Yan, I.-M. Hsing, and J. K. O. Sin, "Investigation of stability and reliability of tin oxide thin-film for integrated micro-machined gas sensor devices," *Sensors and Actuators, B: Chemical*, vol. 81, no. 1, pp. 9–16, 2001.

[5] G. Korotcenkov and B. K. Cho, "Instability of metal oxide-based conductometric gas sensors and approaches to stability improvement (short survey)," *Sensors and Actuators, B: Chemical*, vol. 156, no. 2, pp. 527–538, 2011.

[6] C. Di Natale, E. Martinelli, and A. D'Amico, "Counteraction of environmental disturbances of electronic nose data by independent component analysis," *Sensors and Actuators, B: Chemical*, vol. 82, no. 2-3, pp. 158–165, 2002.

[7] J. G. Monroy, J. González-Jiménez, and J. L. Blanco, "Overcoming the slow recovery of MOX gas sensors through a system modeling approach," *Sensors (Switzerland)*, vol. 12, no. 10, pp. 13664–13680, 2012.

[8] E. L. Hines, E. Llobet, and J. W. Gardner, "Electronic noses: a review of signal processing techniques," *IEE Proceedings - Circuits, Devices and Systems*, vol. 146, no. 6, pp. 297–310, 1999.

[9] J. W. Gardner and P. N. Bartlett, *Electronic Noses: Principles and Applications*, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, Oxford, UK, 1999.

[10] E. Llobet, J. Brezmes, R. Ionescu et al., "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, 2002.

[11] M. Zuppa, C. Distante, K. C. Persaud, and P. Siciliano, "Recovery of drifting sensor responses by means of DWT analysis," *Sensors and Actuators, B: Chemical*, vol. 120, no. 2, pp. 411–416, 2007.

[12] A. Ziyatdinov, S. Marco, A. Chaudry, K. Persaud, P. Caminal, and 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, 2010.

[13] S. Marco, A. Ortega, A. Pardo, and 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, no. 1, pp. 316–321, 1998.

[14] D. Hui, L. Jun-Hua, and S. Zhong-Ru, "Drift reduction of gas sensor by wavelet and principal component analysis," *Sensors and Actuators, B: Chemical*, vol. 96, no. 1-2, pp. 354–363, 2003.

[15] T. Artursson, T. Eklöv, I. Lundström, P. Mårtensson, M. Sjöström, and M. Holmberg, "Drift correction for gas sensors using multivariate methods," *Journal of Chemometrics*, vol. 14, no. 5-6, pp. 711–723, 2000.

[16] R. Gutierrez-Osuna, "Signal processing methods for drift compensation," in *Proceedings of the 2nd NOSE II workshop*, Linköping, Sweden.

[17] S. D. Carlo and M. Falasconi, "Drift correction methods for gas chemical sensors in artificial olfaction systems: techniques and challenges," 2012.

[18] M. Padilla, A. Perera, I. Montoliu, A. Chaudry, K. Persaud, and S. Marco, "Drift compensation of gas sensor array data by Orthogonal Signal Correction," *Chemometrics and Intelligent Laboratory Systems*, vol. 100, no. 1, pp. 28–35, 2010.

[19] S. Wold, H. Antti, F. Lindgren, and J. Öhman, "Orthogonal signal correction of near-infrared spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1-2, pp. 175–185, 1998.

[20] D. Ahmadou, E. Losson, M. Siadat, and M. Lumbreras, "Optimization of an electronic nose for rapid quantitative recognition," in *Proceedings of the 2014 International Conference on Control, Decision and Information Technologies, CoDIT 2014*, pp. 736–741, Metz, France, November 2014.

[21] D. Ahmadou, R. Laref, E. Losson, and M. Siadat, "Reduction of drift impact in gas sensor response to improve quantitative odor analysis," in *proceedings of the IEEE International Conference on Industrial Technology (ICIT)*, pp. 928–933, 2017.

[22] P. C. Jurs, G. A. Bakken, and H. E. McClelland, "Computational methods for the analysis of chemical sensor array data from volatile analytes," *Chemical Reviews*, vol. 100, no. 7, pp. 2649–2678, 2000.

[23] M. M. Macías, J. E. Agudo, A. G. Manso, C. J. G. Orellana, H. M. G. Velasco, and R. G. Caballero, "Improving short term instability for quantitative analyses with portable electronic noses," *Sensors (Switzerland)*, vol. 14, no. 6, pp. 10514–10526, 2014.