



# Global calibration models for temperature-modulated metal oxide gas sensors: A strategy to reduce calibration costs

Albert Miquel-Ibarz<sup>a</sup>, Javier Burgués<sup>a</sup>, Santiago Marco<sup>a,b,\*</sup>

<sup>a</sup> Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona, Spain

<sup>b</sup> Department of Electronics and Biomedical Engineering, University of Barcelona, Martí i Franqués 1, 08028 Barcelona, Spain

## ABSTRACT

Tolerances in the fabrication of metal oxide (MOX) gas sensors lead to inter-device variability in baseline and sensitivity, even for sensors of the same fabrication batch. This has traditionally forced the use of individual calibration models (ICMs) built specifically for each sensor unit, which requires an expensive and time-consuming calibration process and hinders sensor replacement. We propose Global calibration models (GCMs) built using the responses of multiple sensor units, and then applied to a new sensor unit that is not part of the calibration set. GCM have been already successfully applied to transfer calibration models between sensor arrays (electronic noses) for classification tasks. In this work, we investigate the use of such models for regression purposes in temperature-modulated sensors, aiming at the quantification of low concentrations of carbon monoxide (CO) in the presence of variable humidity levels (20–80% r.h. at  $26 \pm 1$  °C). Using a laboratory dataset containing data from 6 replicas of the FIS SB-500–12 model, we evaluate the performance of global models built with data from 1 to 4 sensors when applied to unseen sensor units. Results show that the performance of global models improves with an increasing number of sensors in the calibration set, approaching the performance of individual calibration models ( $1.38 \pm 0.15$  ppm for GCM;  $1.05 \pm 0.24$  ppm for ICM), and surpassing their performance only if few calibration conditions per sensor are available ( $2.09 \pm 0.10$  ppm for GCM;  $2.76 \pm 0.22$  ppm for ICM, if only 5 samples per sensor are used).

## 1. Introduction

Metal oxide semiconductor (MOX or MOS) gas sensors are one of the most successful low-cost technologies in the market for measuring gases and volatile organic compounds (VOCs) at the parts-per-million (ppm) and sub-ppm levels. They are used in application fields such as environmental monitoring, indoor air quality, industrial safety, automotive exhaust, and in-cabin air quality monitoring, residential alarms, and biomedicine, among others. MOX sensors are also often used in electronic noses (e-noses) or sensor arrays to quantify odor intensity and classify odor types [1]. Despite being broadly used in so many fields, MOX generally display certain limitations such as lack of selectivity [2], non-linear response [3], cross-sensitivity to environmental conditions [4,5], high-power consumption [6], slow response time [7,8], and inter-device variability [9,10]. The sensor metrological performance can be improved by a combination of new developments in sensor technology [11], innovative operation principles [12–15] proper calibration methods [16] and, last but not least, signal processing and machine learning [17].

In particular, the inter-device variability in baseline and sensitivity due to tolerances of the fabrication process has forced the use of

individual calibration models, i.e., models tailored for each specific sensor unit, that require costly and time-consuming calibration campaigns. Despite some authors presented methods to reduce the calibration costs [18], the need for individual calibration hinders the use of chemical sensors in mass-applications, especially because faulty sensors cannot be directly replaced [19,20]. We state that the difficulties to transfer calibration models to different (otherwise identical) units are probably the highest barrier for the large-scale deployment of not only temperature modulated MOX sensors, but also sensor arrays and electrochemical sensors, which also exhibit high inter-device variability. When a calibration model trained in one device is applied to other identical devices, there is a non-acceptable degradation of performance.

In order to facilitate the use of predictive models trained with a master instrument to other slave devices, the field of calibration transfer methods have flourished. This field did develop originally for spectroscopic instrumentation [21], but it was applied originally to electronic noses by Balaban et al. [22]. They extended the calibration model of one sensor array to an uncalibrated replica by defining mapping structures (intercept-only functions, linear regression, and transfer matrices) between the response spaces of the two units (master and slave). Since then, many different calibration transfer strategies have been proposed to

\* Corresponding author at: Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona, Spain

E-mail address: [smarco@ibecbarcelona.eu](mailto:smarco@ibecbarcelona.eu) (S. Marco).

<https://doi.org/10.1016/j.snb.2021.130769>

Received 26 May 2021; Received in revised form 13 September 2021; Accepted 14 September 2021

Available online 28 September 2021

0925-4005/© 2021 The Author(s).

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

extend the individual calibration models to uncalibrated replicas of a sensing instrument. The basic idea behind any calibration transfer strategy is to (1) build an individual calibration model with a set of calibration samples from a reference instrument (the master), (2) collect smaller sets of samples with uncalibrated replicas of the same instrument (the slaves), (3) find possible transfer functions to map the responses of the slave and the master systems and (4) use the mapping of the slave responses to transfer the calibration model between replicas. In most cases the mapping structure has proven to be simple.

The transformation of data from the slave device to match the data of the master has been proposed by many different methods including direct standardization (DS), piecewise direct standardization (PDS) [23], but also regression algorithms such as artificial neural networks or partial least squares [22]. For instance, Zhang et al. [24] found that different e-nose systems are related by homogeneous linear functions and performed a linear calibration transfer using six twin e-noses, with four MOX sensors (of different models) each. Global affine transformation (GAT) by robust weighted least squares fittings (RWLS) was used to map slave units to the master, while a Kennard-Stone sequential algorithm (KSS) [25] was applied to select an optimum subset of representative samples. Artificial neural networks (ANNs) were used to predict the gas concentrations based on the sensor responses; however, this was only validated with calibration samples. One drawback of this methodology is that it requires all sensors to be placed together in the same controlled chamber, which might be impractical in realistic scenarios. Fonollosa et al. [26] studied the calibration transfer between MOX sensor arrays, each composed by 8 sensors, on a laboratory experiment including several gas compounds and sensor drift. In their work, they trained one sensing system (master) with a non-linear multiclass regression model to classify the different gases and their concentrations. Then, they used direct standardization (DS) to map the samples from uncalibrated replicas (slaves) to the reference space of the master device.

In temperature modulated MOX sensors, tolerances in the heater resistance induce temperature shifts in the hotplate when excited at constant voltage. Fernández et al., used DS, PDS, Orthogonal Signal Correction (OSC) and Generalized Least Squares Weighting for calibration transfer. In this case, PDS provided the best results with the minimum number of transfer samples [27]. In recent years, the problem of calibration transfer has received a strong boost. Calibration transfer can be reformulated as a problem of multitask learning: the calibration model is learned in different domains that correspond to different devices. The key point is to realize that these learning problems are related, and information can be shared to improve their performances. Calibration transfer makes emphasis on those problems where the number of samples across domains differ. In other words, the number of transfer samples is smaller than the samples used for master calibration. This has been named transfer-sample based multitask learning [28]. Similarly, in other works the main idea has been to learn a domain adaptation technique in such a way that the feature distribution similarity is maximized but preserving the information content of the source. The domain adaptation maybe based on linear projections [29], or autoencoders [30]. Calibration transfer is a very active topic of research and we have today sufficient evidence that the use of a limited number of calibration transfer samples greatly improves the transferability of the master calibration model. However, several difficulties remain: despite recent progress the calibration transfer approach still requires acquiring a number of transfer samples per slave instrument. Additionally, the method gives an excessive weight to the master instrument. An unlucky selection of the master instrument can lead to bad performance for the full batch. In other words, the obtained models for the slave devices may inherit negative characteristics due to an unperforming master device. Finally, calibration transfer does not properly address the specific problem of sensor replacement.

Global calibration methods try to overcome these issues by finding a unique calibration model that can be applied to multiple units of the

same sensor model without the need of calibration transfer samples. The underlying hypothesis is that there is a common mapping structure between the responses of multiple units of the same sensor model and the gas concentrations. The challenge in practice is that this common structure is hidden by the systematic differences of the sensors and time-drifts. A promising solution to find this structure using global calibration methods is multi-unit calibration [31]. Multi-unit calibration consists of building a calibration model with the matrix of responses of multiple sensor units exposed to the same calibration conditions, and then applying this global model to new uncalibrated replicas. These models are optimized so that the prediction accuracy in new replicas is maximized. Obviously, homogenization of the different responses via normalization and scaling is key to achieve good results. Solórzano et al. [31] compared the performance of individual and global calibration models applied to a MOX sensor array (e-nose) aiming at the classification of six gases under variable humidity levels in a controlled laboratory experiment. Individual calibration models had the best overall classification rate (100%), followed by global models (99%) and direct transfer models (91%), the latter ones consisting of applying the individual model of one unit to a different unit. Despite the overall good results, direct transfer models failed to correctly classify 62.5% of the carbon monoxide (CO) samples (with 25.6% false negatives to air), while global models predicted all the CO samples correctly and had 5% of false positives to air. Their results confirmed that individual models were local to the sensor array employed for calibration, indicating possible overfitting and master system dependencies when applied to different units.

In this work we evaluate the feasibility of global calibration models (GCMs) for temperature-modulated MOX sensors aiming at the prediction of CO concentration in variable humidity conditions. This is the first time that global models are studied for a regression problem or using temperature-modulated sensors. Our proposal involves the use of Orthogonalized Partial Least-Squares (O-PLS) calibration method [32] in combination with Repeated Stratified K-Fold cross-validation for model optimization. The model performance is evaluated in external validation samples acquired several weeks after calibration using the limit of detection (LOD) as a figure of merit. We compare the prediction error and temporal stability of global models versus individual models, and study how the number of sensors and samples included in the calibration set affects the performance.

## 2. Materials and methods

### 2.1. Data set

For this study we used a public dataset [32,33] containing recordings from 6 replicas of a temperature-modulated MOX sensor (SB-500-12, FIS Inc. [34]) exposed to gas mixtures of carbon monoxide (range 0–20 ppm) were and humid synthetic air (range 20–80% r.h. at  $26 \pm 1$  °C) inside a controlled in-house gas mixing station. Each gas exposure lasted for 15 min. Before each exposure the sensor chamber was cleaned for 15 min with synthetic air at an identical nominal flow. The sequence of gas concentrations and relative humidity levels was randomized. The SB-500–12 uses a mini-bead type sensing element of tin dioxide material placed in an external housing which contains an active charcoal filter. The multivariant sensor conductance is measured using a voltage divider and a load resistor of 1 M $\Omega$ . The output voltage of the sensors was sampled at 3.5 Hz using an Agilent HP34970A/34901 A data acquisition unit configured at 15 bits of precision and input impedance greater than 10 G $\Omega$ .

We refer the reader to Burgués et al. [32,33] for more detailed information about the experimental setup. According to the datasheet, the sensor can exhibit tolerances of a factor of 10 in baseline (4–40 k $\Omega$ ) and a factor of 2 in sensitivity (1.05–2.1). The recommended operation mode is temperature cycling using a squared heating waveform (0.2 V for 20 s followed by 0.9 V for 5 s) with a total period of 25 s. In the dataset, the

sensor conductance was continuously recorded at a sampling frequency of 1 Hz. Data is organized in 15 measurement campaigns, each one of 24 h duration and consisting of the same 100 measurement conditions (10 concentration levels x 10 humidity levels). For the purpose of this study, we used the subset of the data where the CO concentration is <10 ppm, as previous studies suggested this is the most relevant range for LOD estimation [32,33] while also being a relevant CO range for air quality applications. We also disregarded sensor unit #2, as this was identified as an outlier in previous studies [32,33].

## 2.2. Individual vs global calibration models

Individual calibration models were obtained by training a calibration model with data acquired by one sensor. Global models were built using signals from multiple sensor units, and the model was then applied to a new replica that has been left out of the calibration set (Figs. 1–2). The number of available measurements for building a global model increases with the number of sensors included in the model. For example, assuming a scenario in which  $m$  calibration samples are measured by a set of  $N$  sensors, we can build  $N$  individual models (one per sensor) with  $m$  calibration points each or build a single global model with  $m*N$  calibration points (Table 1).

## 2.3. Preprocessing and transformation

The raw sensor conductance signals were preprocessed to improve the signal-to-noise ratio (SNR), reduce non-linearity, and correct drift. Median filtering (window size of 3 s) followed by a moving average filter (window size of 3 s) was applied to remove spurious spikes and smooth the signals. The filtered signals were logarithmically transformed to linearize the relationship between sensor resistance and gas concentration. To correct for baseline drift (visible after several days of measurements) and compensate differences in the nominal sensor resistances, the sensor conductance was divided by the conductance in clean air. The latter one was measured at the beginning of each daily measurement campaign when the sensor chamber was flushed with clean air. We acknowledge that obtaining clean samples can be a limiting factor in certain open sampling conditions, where the sensors are constantly exposed. Although we have not been able to explore said

cases, different reference points or preprocessing strategies might be then inevitably set.

After being corrected, the preprocessed signals were transformed prior to building the calibration models. For individual models, the training signals were concatenated in one matrix and the heating points were standardized (mean-centered and scaled to unit variance), while for global models they were only mean-centered to improve generalization (a complete empirical demonstration can be found in the Sections A and B of the Supplementary Material). The same process is applied to the test data, always using the transformation parameters of the training data. Since the training pool for global models includes data from multiple sensors, two approaches can be considered: global and individual transformations. In global transformations, the preprocessed signals from the sensors in the training set are first arranged into the matrix of predictors  $\mathbf{X}$ , and then  $\mathbf{X}$  is transformed. In individual transformations, each sensor in the training data matrix is transformed with its own individual parameters prior to building  $\mathbf{X}$ .

## 2.4. O-PLS calibration models

Orthogonalized partial least squares (O-PLS) is a multivariate regression model based on PLS, the latter being one of the most popular multivariate regression models for gas sensor data due to its inherent ability to deal with high-dimensional and collinear sensor data [35]. By using PLS we can find which sets of measurements along the heating point provide relevant information to predict the gas concentrations. Such information is found maximizing the correlation of the X block (signals) variance with the Y block (concentrations) variance, the joint variability of X and Y, by the means of a least-square fitting. After performing PLS, we obtain a set of weights,  $\{w_{j,k}\}$ , (weighted linear combinations of heating points) that are used to calculate a regression vector,  $\beta$ . Although the latter is often used to make predictions, data is usually projected into the set of weights to visualize the scores and understand how the model behaves. However, to make a proper visual analysis the score space must have low dimensionality. To reduce the dimensions of the score space without losing relevant predictive information, an Orthogonal Signal Correction filter can be applied over the PLS set of weights (O-PLS). The resulting O-PLS model has the same regression vector,  $\beta$ , but explains all the score information with just two

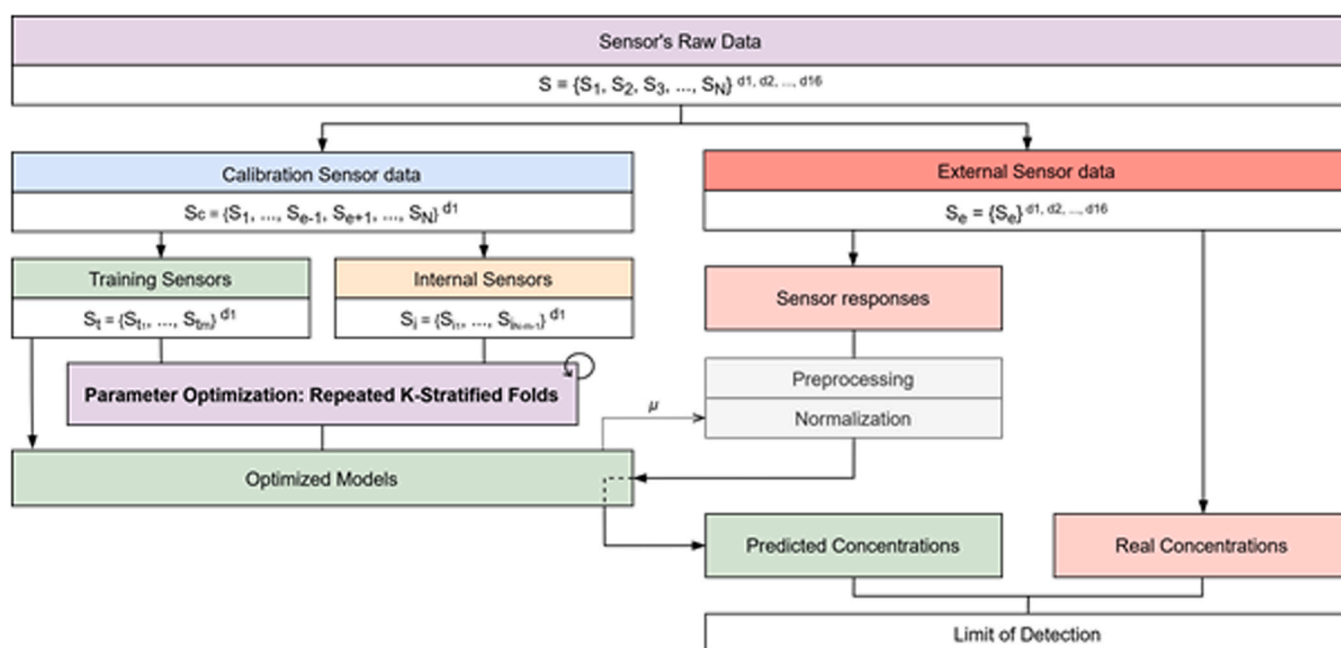


Fig. 1. Flow diagram of the proposed methodology to build and validate global calibration models.

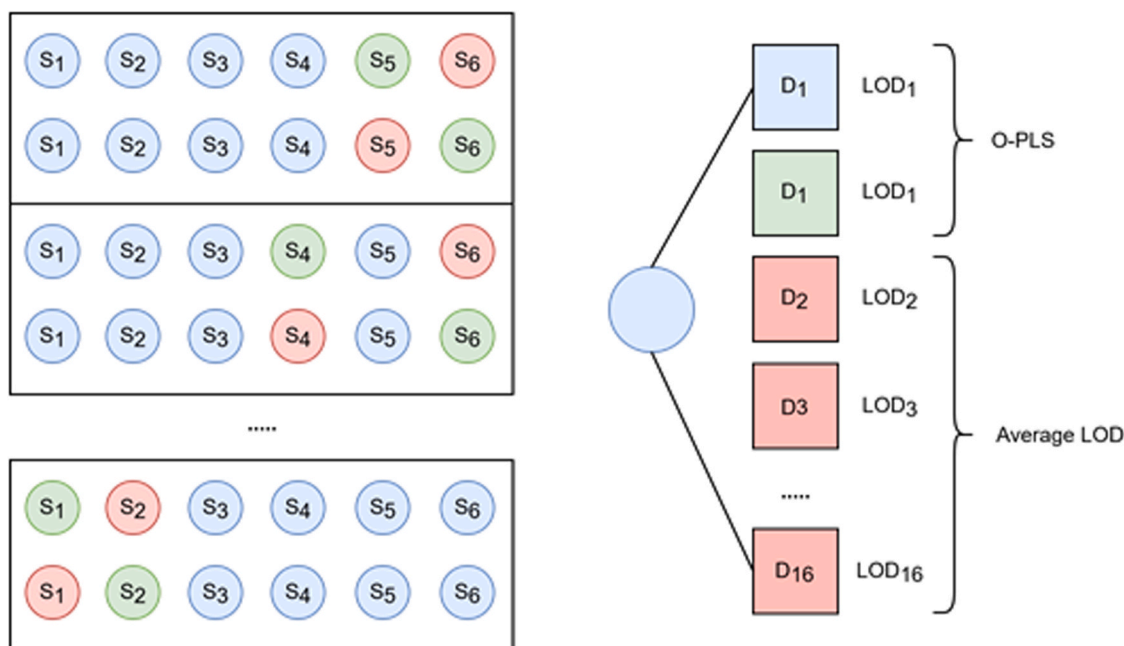


Fig. 2. Simplified flow diagram for a global calibration model that uses 4 training sensors. All possible training (blue), internal validation (green) and external validation (red) sensor sets are evaluated. For a given set, day 1 data is used to build and optimize an O-PLS model, which is externally validated using data from the following days. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Number of calibration samples available in the dataset for individual and global models.

Individual	Global		
	2 sensors	3 sensors	4 sensors
1 sensor			
50	100	150	200

weights,  $\{w_1, w_{2\perp}\}$ . This allows an intuitive visualization of the full structure of the model in the form a two-dimensional score plot, regardless of the number of latent variables of the original PLS model.

The underlying hypothesis behind OSC is that the first weight,  $w_1$ , of the PLS model is the direction that condenses the relevant predictive information; thus, the information contained in any weight orthogonal to  $w_1$  is considered an irrelevant source of noise. Since PLS weights  $\{w_{2,k}\}$  are not orthogonal to  $w_1$ , the first step in OSC is to orthogonalize them using the Gram-Schmidt method [36]. The resulting orthogonal weights,  $\{w_{2,k}\}_\perp$ , are then rotated using the Singular Value Decomposition (SVD) [37] of the covariance matrix of the PLS scores of the calibration data. The resulting first new weight,  $w_{2\perp}$ , condenses almost all the variance of the set of non-orthogonal weights,  $\{w_{2,k}\}$ . Thus, an O-PLS model is composed by only two weights orthogonal to each other: the first one,  $w_1$ , contains the relevant predictive information, and the second one,  $w_{2\perp}$ , contains all the noise-related variance. The orthogonalization just improves model interpretability, it does not change the regression vector of the model, so the performance is preserved.

## 2.5. Limit of detection (LOD)

As a figure of merit to evaluate the prediction performance of a calibration model in the low concentration range, we use the limit of detection (LOD). The LOD indicates the minimum concentration of the target analyte that can be reliably distinguished from the absence of the same analyte, and is defined by the International Union of Pure and Applied Chemistry (IUPAC) [38] for univariate models as:

$$L_D = 2 \ t_{1-\alpha, \nu} \ s_{y,x} \ \hat{A}^{-1} \quad (1)$$

Where  $t_{1-\alpha, \nu}$  is the one-sided t-critical value for the chosen confidence level ( $\alpha$ ) and degrees of freedom ( $\nu$ ),  $s_{y,x}$  is the standard error of regression computed on cross-validation or external validation samples, and  $\hat{A}$  is the slope of the calibration curve. The beauty of the LOD is that combines repeatability (scattering) and sensitivity (slope) into a single number, allowing for a more precise characterization of the measurement system than simply using the root mean squared error (RMSE). However, the univariate LOD formula cannot be directly applied to a multivariate model, such as O-PLS. To circumvent this limitation, we use a methodology proposed by us in a previous work [32] that computes the LOD in O-PLS models by using the scores along  $w_1$  (the first orthogonal O-PLS direction) as a surrogate scalar variable to which the univariate LOD formula (Eq. 1) can be applied.

## 2.6. Model optimization and validation

The main parameter to optimize in an O-PLS model is the number of latent variables (LVs) of the underlying PLS model. To avoid overfitting, the dataset is split into two disjoint subsets: calibration and external validation. The first experimental day is used as the calibration set and the data from the remaining days for external validation. The optimization set was therefore composed of the sensor responses to the 50 calibration conditions of Day 1 whereas the external validation set contained the responses to the remaining 450 conditions (50 conditions x 9 days). A cross-validation (CV) procedure based on Repeated Stratified K-Fold [39] was applied to the optimization set to find the optimum number of LVs without risk of overfitting.

To build an individual model for a given sensor using Stratified K-Folds (without repetition), the optimization set is randomly divided into K-folds containing balanced [CO] distributed samples (Fig. 3). In each iteration, a different fold is left-out the training pool and used as internal validation. Therefore, K-1 folds are transformed and used to train PLS models with different number of LVs. Afterwards, the orthogonalized models are applied to the left-out data to estimate the LOD. Once the K iterations are completed, the K LODs obtained for every O-PLS model with a given number of LVs,  $O\text{-PLS}_{LV}$ , were averaged to produce  $LOD_{LV}$ . Visual inspection of the resulting LOD vs LV boxplot is used to



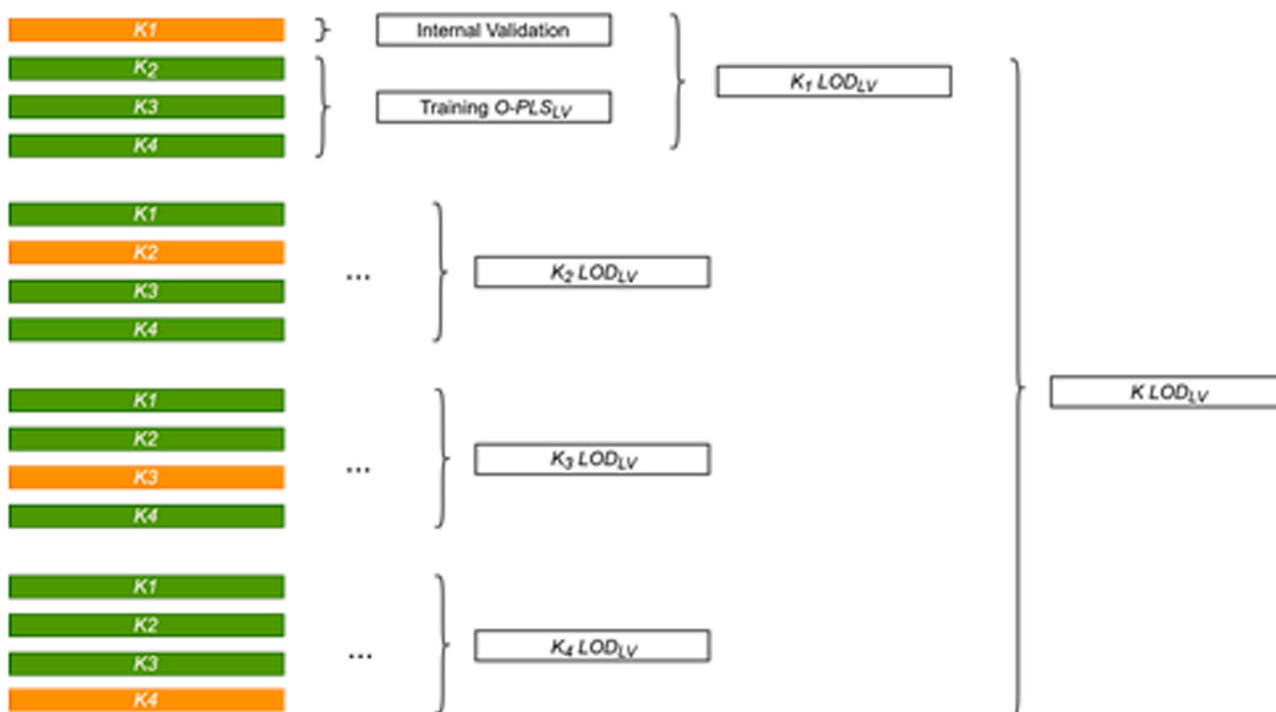


Fig. 3. Example of a K-Fold cross-validation process used to optimize Individual Calibration Models. Only 4 folds have been represented for visualization purposes.

choose the optimum number of LVs. To prevent overfitting, the stratified k-fold process is repeated  $R$  times with a different shuffling of the data. The  $(K\text{-LOD})_R$  of every  $O\text{-PLS}_{LV}$  are averaged through the  $R$  repetitions. After determining the optimum number of LV, the model is refit using all calibration data. Then, the optimized model is externally validated with samples collected by the same sensor in days 2, ...,  $N$ . This described process is independently done for every sensor in the dataset.

To build  $m$ -sensor global models given a dataset of  $N$  sensors, Repeated Stratified K-Fold CV is also used but the data partitions and optimization objectives are completely different (Fig. 4). First, an external validation sensor is selected. The unused sensors are used as the calibration set:  $m$  sensors are selected to train the model and  $N-m-1$  are used for internal validation. The K-Folds only contain data from the  $m$  training sensors, stratified by [CO] and sensor. One-fold is left out of the

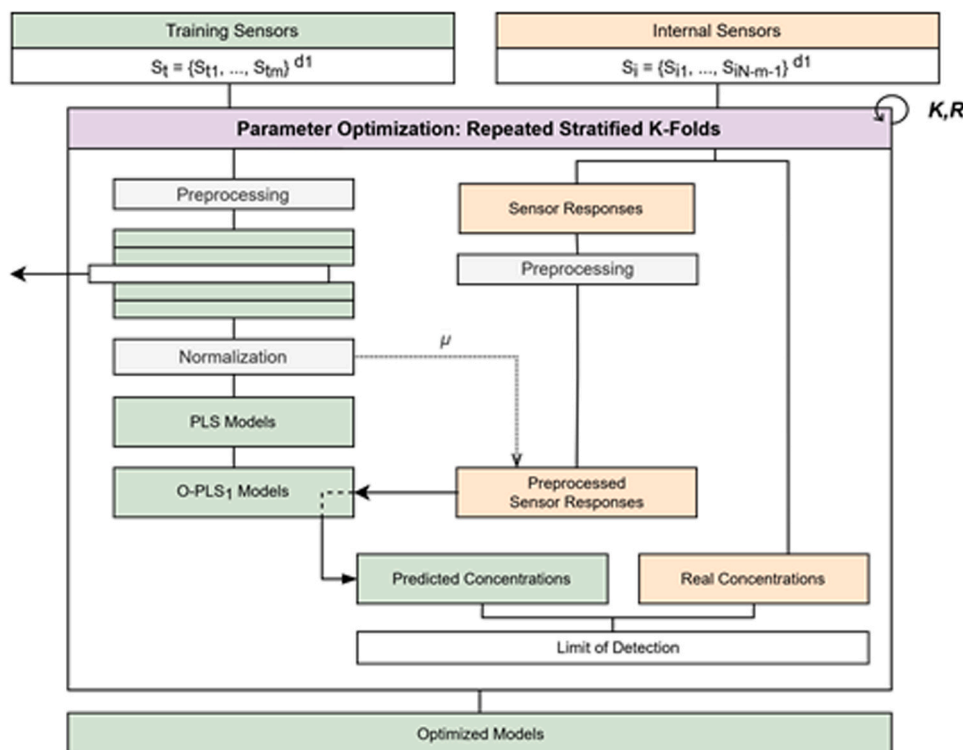


Fig. 4. Flow diagram of the repeated stratified K-Fold process used to optimize the number of LVs of the global models.

training pool. Once the training folds are transformed by sensor, PLS<sub>LV</sub> models are built, orthogonalized and applied to the internal validation sensors. The resulting LOD<sub>LV</sub> are averaged through the sensors, and this process is repeated for every fold. To avoid overfitting, this process is repeated multiple times reshuffling the data. Similarly to individual models, the boxplot of the resulting LOD vs LV curve is used to determine the optimum number of LVs and the model is refitted with all training data afterwards. Then, the optimized model is externally validated with all the samples collected by the external validation sensor in days 2,,*N*. The described process is repeated until all sensors are externally validated.

### 3. Results and discussion

#### 3.1. Raw signals

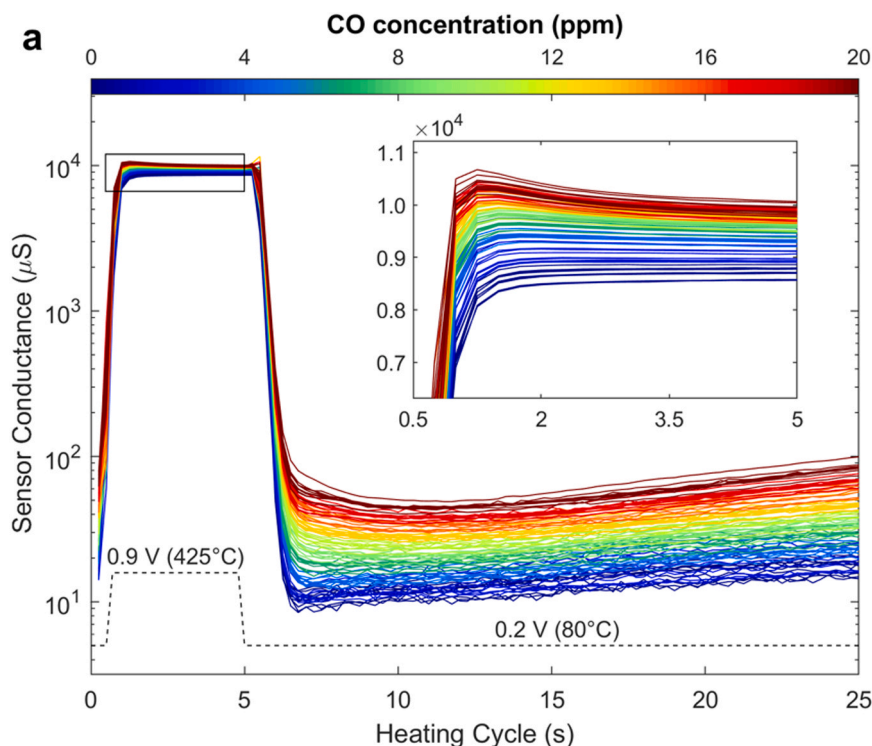
An example of the responses of one sensor unit to the generated mixtures of CO (0–20 ppm) and humid synthetic air (20 – 80% r.h. at  $26 \pm 1$  °C) in the first measurement day is shown in Fig. 5a. The conductance patterns clearly reflect the two temperatures of the heating pattern, i.e. high temperature for 5 s followed by low temperature for 20 s. Most of the sensitivity to CO happens at the low temperature regime, but the high level is also highly sensitive to CO (see inset). The inter-sensor variability is shown in Fig. 5b, by plotting the responses of the six sensors in the dataset to a fixed concentration of 20 ppm. As can be seen, there are not only baseline differences between units but also changes in sensitivity along the heating pattern. For example, unit #3 (green trace) shows the lowest conductance at high temperature but the highest conductance at low temperature. The baseline differences at high temperature amount for 20% in the worst case (unit #3 versus units #1 and #6). This variability is within the specs provided in the sensor datasheet [34]. An elegant way to visualize the differences in the responses of different units is by means of Principal Component Analysis (PCA), as shown in Fig. 6. Here we can identify two distinct sensor

behaviors according to how the data points spread in the space spanned by the first two principal components (PCs). Sensors 1, 4 and 5 (panels a-c) group into what we call Type-A response, whereas sensors 2, 3 and 6 (panels d-f) group into a Type-B response. The first PC, which accumulates nearly 94% of the total variance, captures the largest variations in the CO concentration. The second PC (2.06% variance) correlates slightly with the CO concentration mostly at low concentrations. Focusing on PC1 (the most important), we can see that the projection of the scores into this PC show a better separability for Type-A sensors at high concentrations and for Type-B sensors at low concentrations. This suggests that Type-B sensors could have a better performance in terms of LOD since PC1 will dominate the model parameters. One can also imagine that a calibration model built for a Type-A sensor will probably produce high prediction errors if directly applied to a Type-B unit.

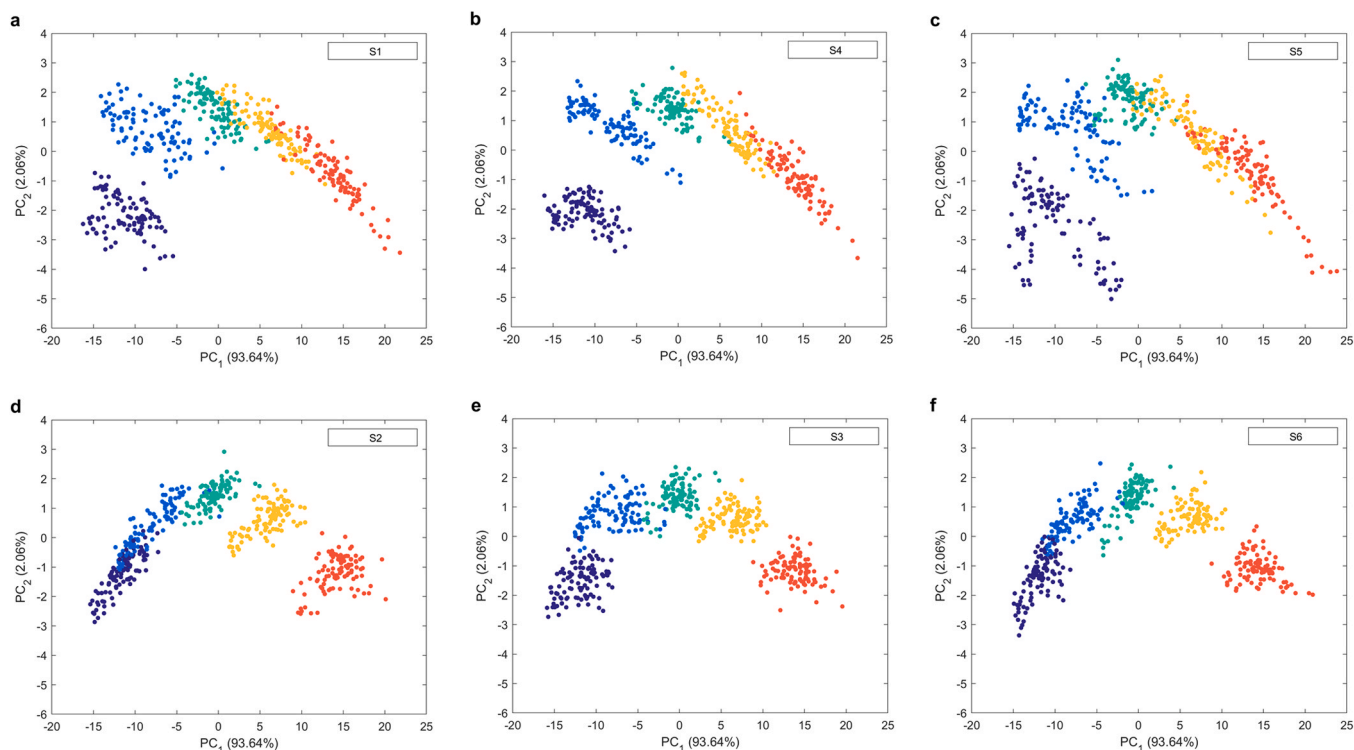
#### 3.2. Preprocessing and transformation

As an illustrative example, the preprocessed signals of one sensor unit are shown Fig. 7. It can be seen that the first seconds after the temperature transitions show higher sensitivity to the CO concentration, which was already observed in previous analysis of this same dataset after an elaborated data analysis [33]. The signal processing that we use in this study has highlighted this optimum detection point without any complicated data processing or analysis.

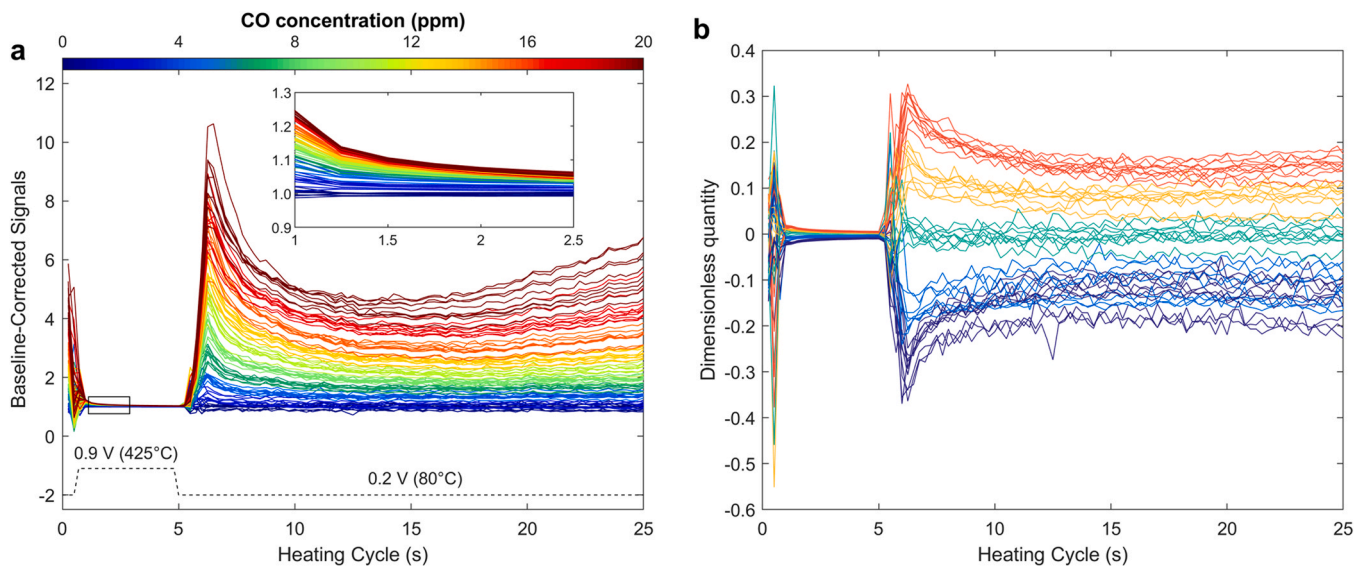
Fig. 8 compares two transformation approaches for global models (global standardization and sensor-specific mean-centering) by means of a PCA score plot. In global standardization (Fig. 8a–b), the first PC condenses the signal variance correlated to the CO concentration (Fig. 8a), whereas the second PC mostly captures the inter-device variability (Fig. 8b), hence modeling sensor dissimilarity. This variance derived by the sensors systematic differences condenses up to a 11% of the total variance, shadowing other relevant sources of variance. The scores along PC2 indicate which sets of sensors can be considered quasi-identical measuring systems (Fig. 8b). The scores of Type-A (S1, S4 and



**Fig. 5.** (a) Logarithmic conductance patterns of sensor unit #1 during one heating cycle, colored by CO concentration (see color bar on top). The black dashed line represents the heater voltage and the corresponding operation temperature. (b) Conductance patterns of the six sensor units when exposed to 20 ppm of CO (colored by sensor). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** Principal component analysis of the preprocessed and standardized sensor signals (each subplot represents the scores of a different sensor) showing two distinct sensor behaviors (a–c versus d–f). All sensor-standardized data from Day 1 was used to build the model, and the measurements taken in Days 2–15 were projected into this model. The colormap indicates the CO concentration, linearly spaced from 0 to 9 ppm (dark blue to red, respectively). The values next to each axis label indicate the variance captured by each principal component. To perform PCA, multivariate sensor data was processed using Singular Value Decomposition and a scatter plot function. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

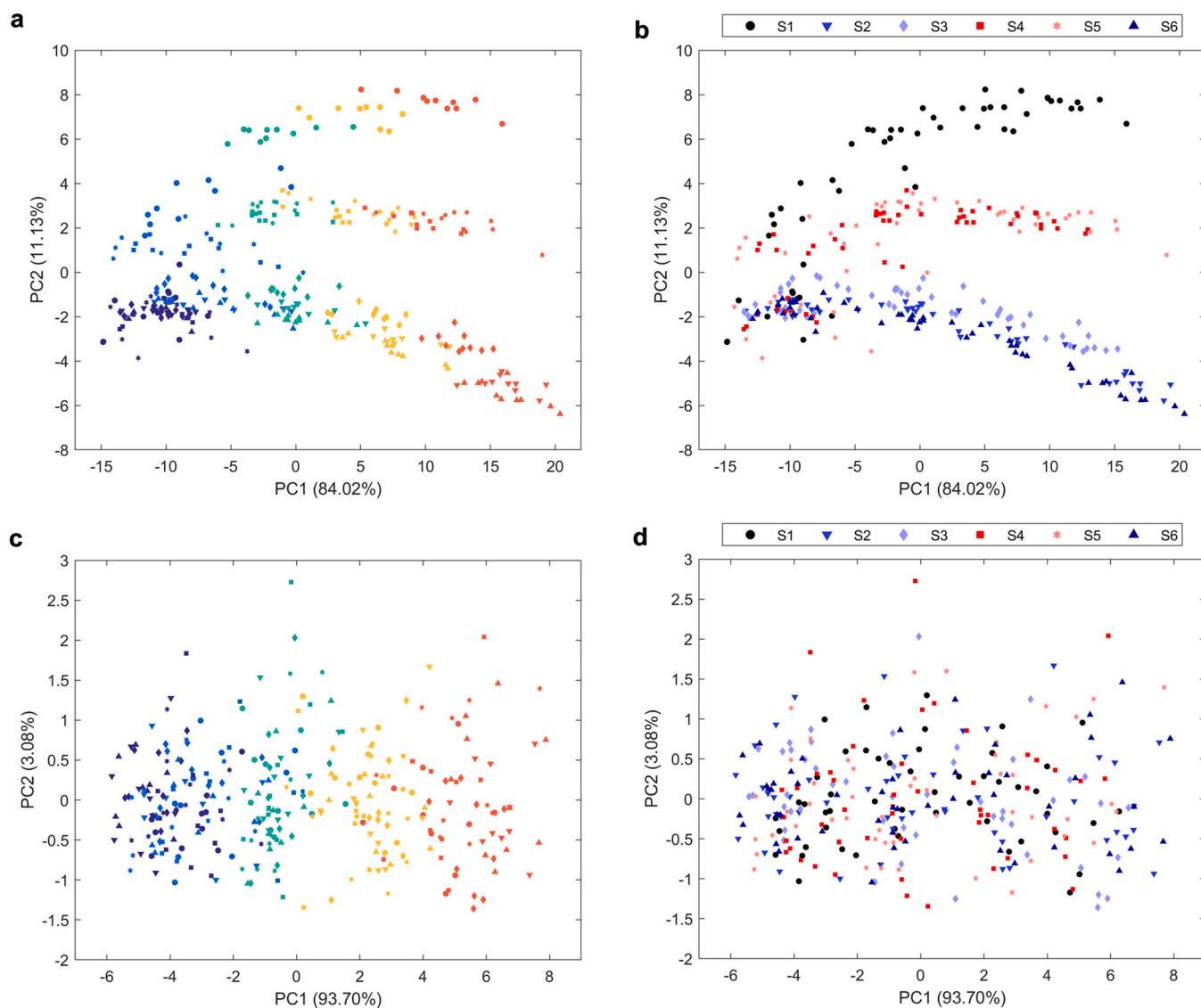


**Fig. 7.** Example of sensor signals after (a) baseline correction and (b) mean-centering. Data corresponds to sensor unit #4. The responses during the last heating cycle of all measurements taken in Day 1 are stacked into the same figure for visualization purposes. The colormap indicates the CO concentration (0–20 ppm). The black dashed line represents the heater voltage and the corresponding operation temperature. The signals in (b) are the ones used as input for the PLS model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

S5) and Type-B (S2, S3 and S6) sensors drift in the positive and negative PC2 direction, respectively, with increasing CO concentration. Despite the low number of sensor units in this study do not allow us to draw strong conclusions, Type-A sensors seem to have higher inter-device variability than Type-B sensors. A different scenario is portrayed when sensor-specific mean-centering is performed (Fig. 8c–d). Now, the

scores are clustered by CO concentration and not by sensor units, and the variance captured by PC1 has increased from  $\sim 84\%$  to  $\sim 94\%$ , probably leading to a better starting point for building a global calibration model.

Relevant information about the sources of variance in each transformation method can be extracted from the PCA loadings (Fig. 9). If global standardization is performed (solid line), PC1 seems to give



**Fig. 8.** Global standardization (a–b) and sensor-specific mean-centering (c–d) effects on the scores of the day 1 sensor signals seen in the reduced space of the first two principal components of a PCA. The scores are colored by concentration (panels a–c) and by sensor (panels b–d). The five CO concentrations are equally spaced, ranging from 0 to 8.89 ppm. The shapes of the scores also code the sensor number. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

slightly more weight to the low temperature level and penalizes the temperature transitions, while PC2 captures the variability in the high temperature of the pattern. If sensor-specific mean-centering is performed (dashed line), PC1 only weights the low temperature level (with higher weight to the spot with high sensitivity after the temperature transition), while PC2 captures the strong variance in the temperature transitions. Therefore, the inter-device differences are mostly caused by the variance of the responses in the high temperature regime, which was also seen in Fig. 5a–b. The reason why global or individual standardization do not perform well is that these differences are accentuated when all sensors are scaled using the mean standard deviation of the data (see more details in the [Section A of the Supplementary Material](#)).

### 3.3. Model optimization

Fig. 10 illustrates the optimization of a 4-sensor GCM and its external validation against an unseen sensor. In Fig. 10a, each box represents the internal validation LOD predictions when a given subset of four sensors is used for model training (averaged over 100 repetitions and 2 folds). In this example, four LVs could be a good choice to ensure the model is

neither underfitted nor overfitted. Increasing the model complexity beyond four LVs brings slight performance improvements in internal validation but probably not in external validation. Fig. 10b shows the scores of the training, internal and external validation samples of an optimized model. The first O-PLS weight condenses 98% of the variance due to CO concentration, whereas the second weight condenses the orthogonal variation which only relates to the cross-sensitivities. The separability between the clusters along the first direction indicates the ability of the model to distinguish between concentration levels. The score dispersion along the second direction shows the noise, which according to the second orthogonal weight pattern (red line in Fig. 10c) is mostly due to the misalignments of the sensor responses around the temperature transitions. Due to a slight lack of synchronization in sampling the conductance patterns of different sensor units, the fast conductance transitions under a high temperature gradient are an important source of interference. Interestingly, this noise does not seem to be related to humidity changes, which the model can inherently reject (further detail is provided in the [Section C of the Supplementary Material](#)). The first O-PLS weight direction in Fig. 10c concentrates the weight in the steady-state part of the low temperature regime of the



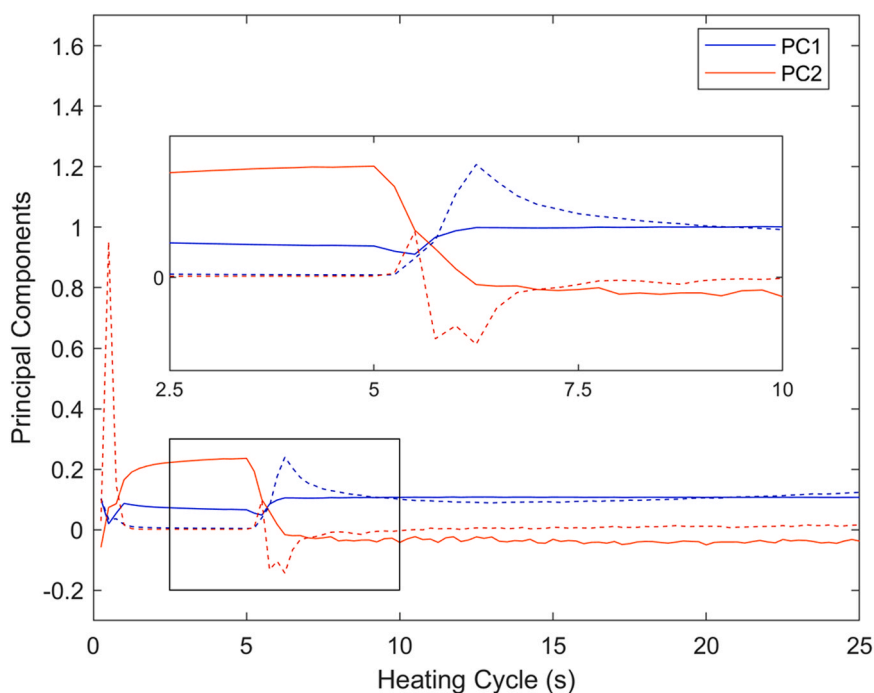


Fig. 9. Global standardization (solid lines) and sensor-specific mean-centering (dotted lines) effects on the first two principal components of the shown PCAs.

heating cycle, this is, rejecting the noise associated to the temperature transitions and resembling the first loading of the PCA models shown before.

It is important to note that the first orthogonalized weights of all the possible 4-GCMs trained models are remarkably similar. In fact, these weight directions are practically parallel (Section B of the Supplementary Material). This indicates that the inner structure that relates the sensor signals and the CO concentration remains constant among models. The second orthogonalized directions of the different models are collinear too, which indicates that there is also a common noise-filtering structure for all sensors. This partially ensures model transferability between the sensor units used for training, internal validation, and external validation.

Small differences between the OPLS scores obtained in training and external validation will have a negligible effect in the LOD estimation as long as these differences mostly occur in the second O-PLS direction. This is derived from the fact that the LOD estimation only compares the linear relationship between the scores of each sensor along the first O-PLS weight and the real CO concentration. This can be seen in Fig. 10d, which shows the scores along  $w_1$  versus the concentration. The LOD in this example is 1.35 ppm.

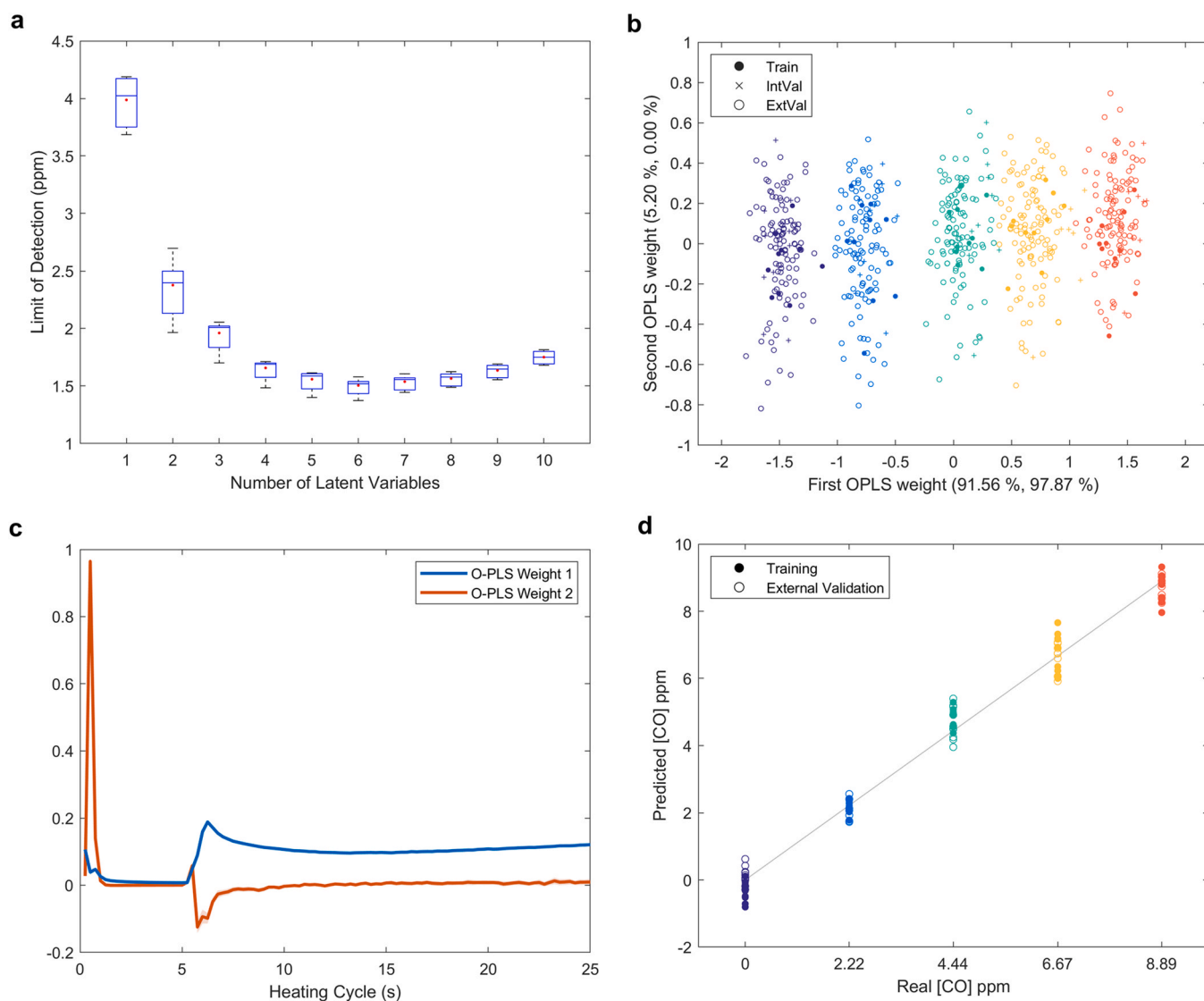
The regression vector of the OPLS model indicates how the sensor responses are weighted by the model to predict the gas concentration, thus showing which are the most relevant features along the heating cycle. Fig. 11 compares the normalized regression vectors of individual and global models. The regression vector coefficients of the individual models show more dispersion than in the global models, probably due to the sensor-specific nature of the individual optimization objectives. In contrast, the regression vectors of global models are very stable and, to some extent, converge to the average value of the individual regression vectors. The low dispersion of the GCMs regression vectors encouraged us to use the averaged regression vector to predict the LOD of every sensor in all the calibration campaigns. The predicted daily LOD settled at  $1.34 \pm 0.13$  ppm, which is slightly better than the one obtained using any of the original regression vectors of the GCMs ( $1.38 \pm 0.15$  ppm).

#### 3.4. Influence of training size and model complexity

Now we study how important is the training size of individual and global calibration models for estimating the daily LOD. For that, N-sensor GCMs were built using all possible combinations of sensors with varying training sizes. The characteristics of the different models and their optimization parameters are shown in the Supplementary Material (Tables D1 and D2). The results of this study are summarized in Fig. 12.

In Fig. 12 we compare the LOD of N-GCMs and ICMs built with different training sizes. The results show that the performance of N-GCMs increases with the number of sensors,  $N$ , and samples,  $m$ , used for training (Fig. 12a). The largest improvement generally occurs when the training size increases from 10% (5 samples/sensor) to 20% (10 samples/sensor). The improvement of GCMs saturates after using 30–40% of the training size, while ICMs keep improving until ~80% of the calibration samples have been used. ICMs and 4-GCMs produce a daily LOD of  $1.05 \pm 0.24$  ppm and  $1.38 \pm 0.15$  ppm, respectively. This is, fully trained ICMs outperform global models. However, when the number of calibration samples is very low (e.g., 5 samples per sensor or 10% of the training size), GCMs built with 2–4 sensors seem to outperform ICMs. In the case of 4 sensors and 5 samples we have an LOD of  $2.09 \pm 0.10$  ppm for GCM and  $2.76 \pm 0.22$  ppm for ICM. The reduced calibration costs of GCMs and their generic nature (i.e., applicable to new sensor units without acquiring transfer samples) could be, in some scenarios, more convenient than achieving the lowest possible LOD. In this dataset, 4-GCMs built with 10 samples/sensor (20% of the training size) represent a good tradeoff between calibration effort and performance (LOD =  $1.61$  ppm  $\pm$  0.14 ppm).

Fig. 12b explains whether the improvement of the GCMs is due to the increase of the number of training samples per sensor or to the fact that these samples come from more sensors. Three main conclusions can be drawn from this panel. First, 1-GCMs (direct transfer models) have the highest LOD, probably because models containing just one sensor in the training set are not global models per se. Second, the LOD curves of the remaining GCMs look like shifted versions of each other, suggesting that adding new sensors to the training set might be the main driver for the observed performance enhancement. Third, GCMs exhibit lower and more stable standard deviations than ICMs for any training size



**Fig. 10.** (a) Day 1 LOD of a 4-sensor GCM versus the number of LVs. Each box represents the interquartile range (IQR) of the distribution of LODs for the five used sensors, being the bottom and top edges the 25th and 75th percentiles, respectively. The mean and median of the distribution are represented by a horizontal blue segment and a red dot, respectively. The whiskers extend to the most extreme data points not considered outliers. The latter are considered standard outliers (empty circles) or extreme outliers (filled circles) depending on whether their value exceeds the interquartile range by more than 1.5 or 3 times, respectively; (b) Score plot of an optimized 4-sensor O-PLS model (external validation samples from all days are also shown). The color indicates the CO concentration (ranging from 0 ppm to 8.89 ppm). The values next to each axis label describe the of the signal-block and response-block variance contained in each direction. (c) O-PLS weights of all trained model; (d) Regression between t1 scores and true concentration, showing a linear dependency and homoscedasticity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

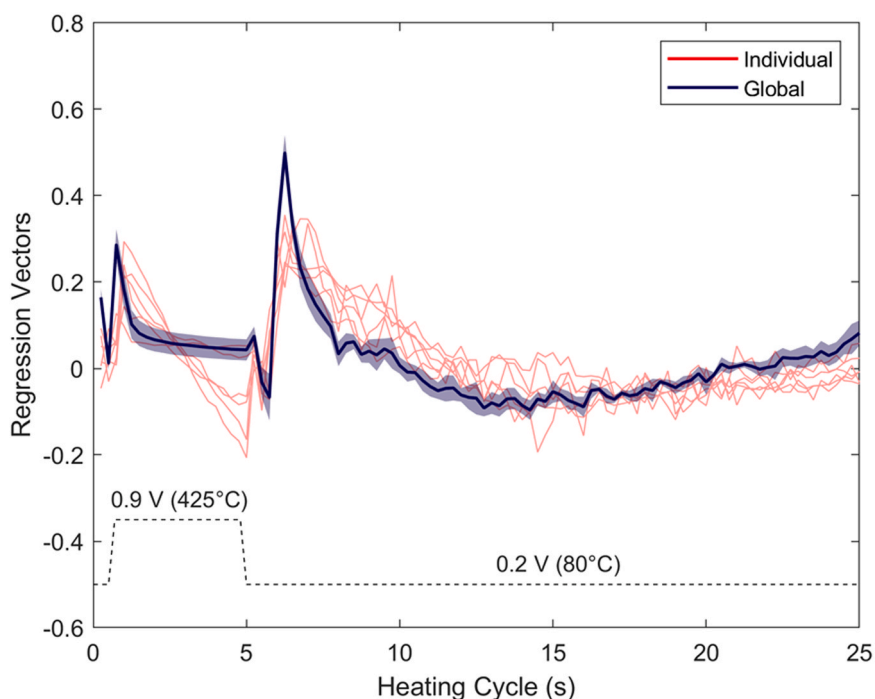
(0.20 ppm vs 0.33 ppm at 10% training size; 0.15 ppm vs 0.24 ppm at 100% training size). These points support the utility of GCMs: it is possible to learn the calibration models using only few calibration conditions applied to large sensor sets, instead of using many calibration conditions for only few sensors.

We may therefore argue that the developed GCMs succeed in capturing the underlying structure that relates the sensor responses to the steady-state gas concentrations. The presented methodology seems to be able to work as a generic calibrant for a FIS SB-500-12 sensor in the given controlled scenario and could potentially be extended to other sensing devices.

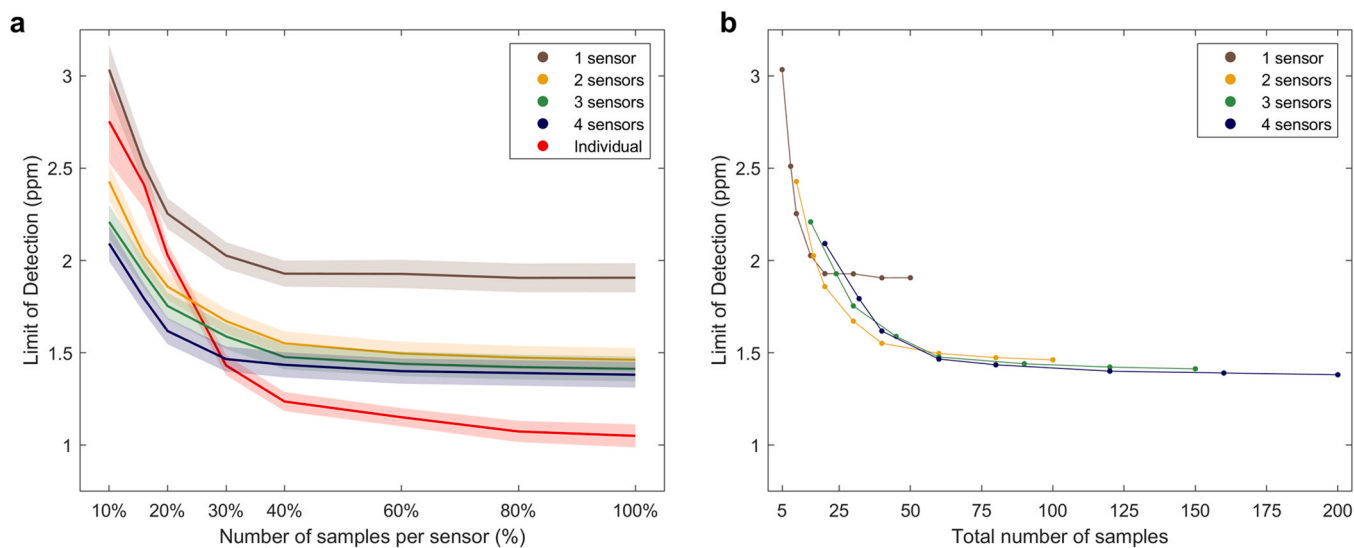
### 3.5. Temporal stability of global and individual models

Having studied the optimum training conditions for GCMs, we now benchmark the predictive performance of 4-GCMs against fully trained

ICMs (most challenging adversary) in a two-week period (Fig. 13 a–b). As Day 1 data is used to optimize the models, its cross-validation LODs won't be included in external validation LOD analysis, derived using the following unseen days. The mean daily LOD predicted using global models ( $1.38 \pm 0.15$  ppm) is 31% higher than the one given by individual models ( $1.05 \pm 0.24$  ppm). However, global models produce a LOD with higher temporal stability in terms of mean value and dispersion, which means that the LOD obtained in GCMs in Day 1 is less overfitted to the training data and therefore more representative of future performance. As the number of available samples for calibration decreases (use case of GCMs), the performance of ICMs will rapidly decrease (i.e., LOD increases) while GCMs will keep a relatively good performance (Fig. 13c–d). A possible interpretation is that ICMs with low sampling conditions (Fig. 13c) seem to be prone to suffer from noisy samples, whereas GCMs (Fig. 13d) naturally use more samples so noise resistance is enhanced.



**Fig. 11.** Orthogonal partial least squares regression vector comparison between Individual and Global models. The solid line indicates the mean regression vector over all day 1 optimized global models used to estimate the external validation daily LOD of the different calibration campaigns. The shaded area spans up to twice the observed standard deviation. The black dashed line represents the heater voltage and the corresponding operation temperature.



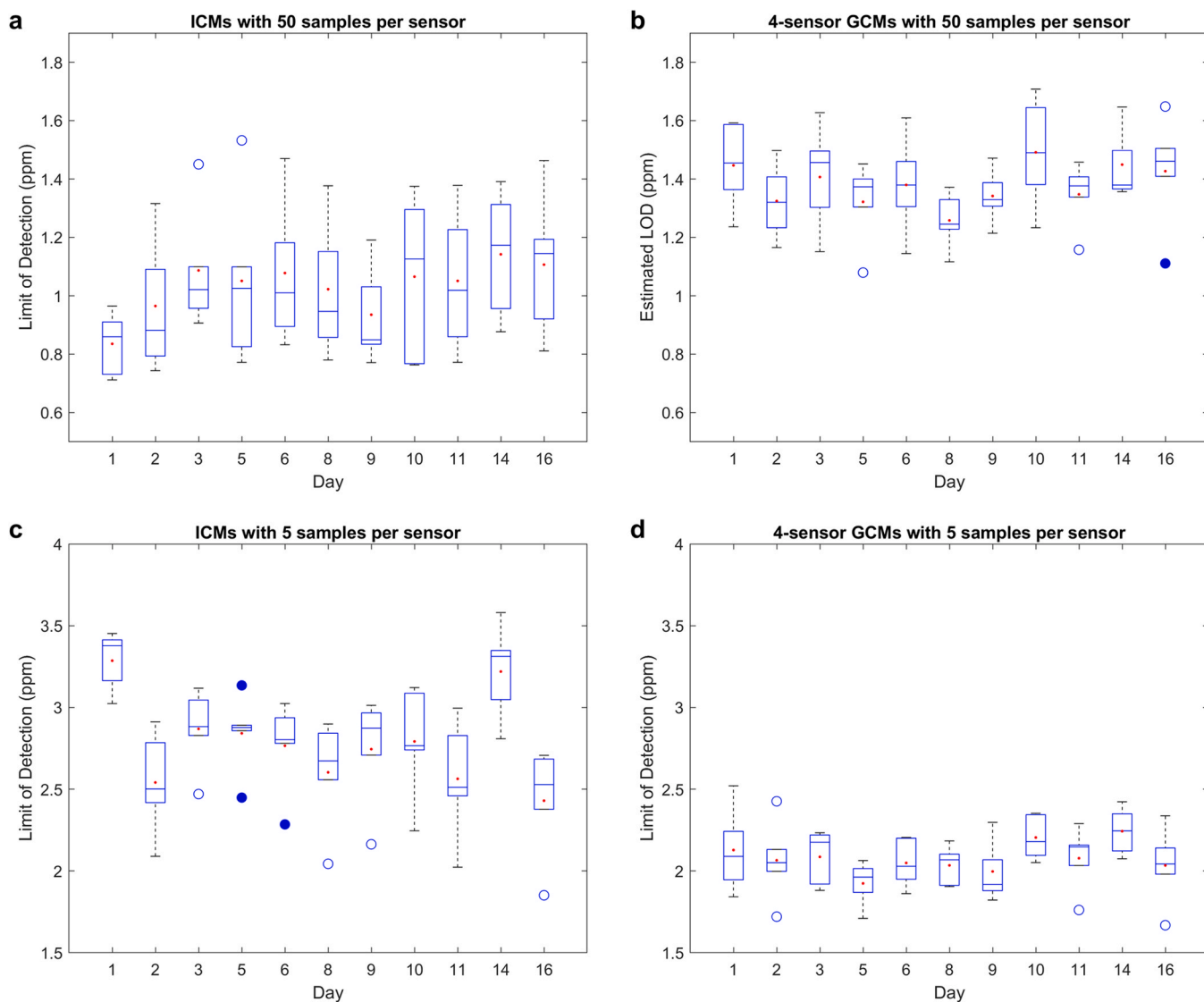
**Fig. 12.** Limit of detection of ICM and GCM models as a function of (a) number of samples per sensor; (b) total number of calibration samples. The solid line in (a) represents the mean daily LOD predicted by each model in external validation, while the shaded area indicates the standard deviation of the estimations.

#### 4. Conclusions

The main goal of this study was to investigate whether global calibration models could compare to the performance of individual calibration models for temperature-modulated MOX sensors. The main advantage of global models is that they can be directly transferred among different units of the same sensor model, reducing the calibration costs and effort, and allowing an easy replacement of faulty units. Using an experimental dataset with temperature-modulated MOX sensors aiming to the prediction of low levels of carbon monoxide in the presence of humidity interference, we studied the performance of O-PLS calibration models built with different number of sensors and varying

training sizes.

The main result is that global models can be effective with a minimal number of sensors in the training set. In the explored dataset, as few as 4 sensors were enough to learn the common underlying mapping between the sensor responses and the gas concentration after proper signal preprocessing. In the explored dataset, as few as 4 sensors were enough to learn the common underlying mapping between the sensor responses and the gas concentration after proper signal preprocessing. It is certainly possible that adding additional sensors might further improve the performance. We expect though that the law of marginal returns will manifest soon. In other words, the improvements will saturate for a sensor set of a small cardinality. While we have seen that global models



**Fig. 13.** Temporal evolution of the predicted daily LOD using (a, c) optimized ICMs and (b, d) optimized 4-GCMs. Lower LOD values are better. The models are always optimized using data from day 1, with  $n = 50$  samples/sensor in (a, b) and  $n = 5$  samples/sensor in (c, d), to make predictions in the following days. Each box represents the interquartile range (IQR) of the distribution of daily LODs for the six used sensors, being the bottom and top edges the 25th and 75th percentiles, respectively. The mean and median of the distribution are represented by a horizontal blue segment and a red dot, respectively. The whiskers extend to the most extreme data points not considered outliers. The latter are considered standard outliers (empty circles) or extreme outliers (filled circles) depending on whether their value exceeds the interquartile range by more than 1.5 or 3 times, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

improved their performance with an increasing number of training sensors, the largest improvements occurred when doubling the number of samples per sensor from 5 to 10. In contrast, the performance of individual models did not easily saturate, and achieved better daily LODs than the global models if enough samples were available (e.g., more than 15–20 samples). However, global models provided lower LODs if the number of samples per sensor was lower than 10. Therefore, for global models it is more convenient to increase the number of sensors in the training pool rather than increasing the number of samples per sensor.

Another finding of this study is that the LOD of global models seemed more stable in time than that of individual models. This means that the LOD computed during calibration was more representative of future model performance than the one obtained using individual models. Our interpretation of this result is that by rejecting the variability among sensor units, global models could also be partially rejecting drift directions.

These results define a clear use case for global models in scenarios where reduced calibration costs and generic models (i.e., applicable to new sensors without acquiring more samples) are more important than achieving the lowest possible LOD. At the same time, being able to share the calibration cost and effort of one single global model among numerous replicas, should be also especially attractive to sensor manufacturers, who solve inter-device variance through batch screening and individual calibrations. We envision that the same calibration strategy used in this paper for temperature-modulated sensors could be feasible for isothermal sensor arrays. Follow-up studies involving longer measuring campaigns, larger number of sensors and various interfering gases are necessary to confirm the results obtained in this first study. The methodology presented in this paper is intuitive and can be applied to other sensor technologies with vectorized responses (such as gasFETs or to isothermal sensor arrays) broadening the possible applications of chemical sensors to large-scale problems.



## CRedit authorship contribution statement

**JB, SM:** Conceptualization. **JB, SM:** Methodology. **AM, JB:** Software. **AM, JB:** Validation. **AM, JB:** Formal analysis. **AM:** Investigation. **SM:** Resources. **JB:** Data curation. **AM:** Writing – original draft. **JB, SM:** Writing – review & editing. **AM, JB:** Visualization. **JB, SM:** Supervision. **SM:** Project administration. **SM:** Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We would like to acknowledge, the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya (expedient 2017 SGR 1721); the Comissionat per a Universitats i Recerca del DIUE de la Generalitat de Catalunya; and the European Social Fund (ESF). Additional financial support has been provided by the Institut de Bioenginyeria de Catalunya (IBEC). IBEC is a member of the CERCA Programme/Generalitat de Catalunya.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.snb.2021.130769](https://doi.org/10.1016/j.snb.2021.130769).

## References

- J.W. Gardner, P.N. Bartlett, *Electronic noses. Principles and applications*, *Meas. Sci. Technol.* **11** (1999) 1087.
- A. Ponzoni, C. Baratto, N. Cattabiani, M. Falasconi, V. Galstyan, E. Nunez-Carmona, F. Rigoni, V. Sberveglieri, G. Zambotti, D. Zappa, Metal oxide gas sensors, a survey of selectivity issues addressed at the Sensor Lab, Brescia (Italy), *Sensors* **17** (2017) 714, <https://doi.org/10.3390/s17040714>.
- P.K. Clifford, D.T. Tuma, Characteristics of semiconductor gas sensors I. Steady state gas response, *Sens. Actuators* **3** (1982) 233–254, [https://doi.org/10.1016/0250-6874\(82\)80026-7](https://doi.org/10.1016/0250-6874(82)80026-7).
- K. Kamarudin, V.H. Bennetts, S.M. Mamduh, R. Visvanathan, A.S.A. Yeon, A.Y.M. Shakaff, A. Zakaria, A.H. Abdullah, L.M. Kamarudin, Cross-sensitivity of metal oxide gas sensor to ambient temperature and humidity: effects on gas distribution mapping, in: *Proceedings of the AIP Conf.*, (2017), 020025. (<https://doi.org/10.1063/1.4975258>).
- M. Holmberg, T. Artursson, Drift compensation, standards, and calibration methods, in: *Handb. Mach. Olfaction*, (2004), 325–346. (<https://doi.org/10.1002/3527601597.ch13>).
- J. Burgués, S. Marco, Santiago Marco, J. Burgués, S. Marco, J. Burgués, S. Marco, Low power operation of temperature-modulated metal oxide semiconductor gas sensors, *Sensors* **18** (2018) 339, <https://doi.org/10.3390/s18020339>.
- D. Martínez, J. Burgués, S. Marco, Fast measurements with MOX sensors: a least-squares approach to blind deconvolution, *Sensors* **19** (2019), <https://doi.org/10.3390/s19184029>.
- J. Fonollosa, S. Sheik, R. Huerta, S. Marco, Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring, *Sens. Actuators B Chem.* (2015) 618–629, <https://doi.org/10.1016/j.snb.2015.03.028>.
- W. Göpel, K.D. Schierbaum, SnO<sub>2</sub> sensors: current status and future prospects, *Sens. Actuators B Chem.* **26** (1995) 1–12, [https://doi.org/10.1016/0925-4005\(94\)01546-T](https://doi.org/10.1016/0925-4005(94)01546-T).
- M. Bruins, J.W. Gerritsen, W.W.J. Van De Sande, A. Van Belkum, A. Bos, Enabling a transferable calibration model for metal-oxide type electronic noses, *Sens. Actuators B Chem.* **188** (2013) 1187–1195, <https://doi.org/10.1016/j.snb.2013.08.006>.
- P.T. Moseley, Progress in the development of semiconducting metal oxide gas sensors: a review, *Meas. Sci. Technol.* **28** (2017), 082001, <https://doi.org/10.1088/1361-6501/aa7443>.
- I. Sayhan, A. Helwig, T. Becker, G. Müller, I. Elmi, S. Zampolli, M. Padilla, S. Marco, G. Mueller, I. Elmi, S. Zampolli, M. Padilla, S. Marco, G. Müller, I. Elmi, S. Zampolli, M. Padilla, S. Marco, Discontinuously operated metal oxide gas sensors for flexible tag microlab applications, *IEEE Sens. J.* **8** (2008) 176–181.
- F. Palacio, J. Fonollosa, J. Burgués, J.M. Gomez, S. Marco, Pulsed-temperature metal oxide gas sensors for microwatt power consumption, *IEEE Access* **8** (2020) 70938–70946, <https://doi.org/10.1109/ACCESS.2020.2987066>.
- E. Martinelli, D. Polese, A. Catini, A. D'Amico, C. Di Natale, Self-adapted temperature modulation in metal-oxide semiconductor gas sensors, *Sens. Actuators B Chem.* **161** (2012) 534–541, <https://doi.org/10.1016/j.snb.2011.10.072>.
- K.J. Johnson, S.L. Rose-Pehrsson, Sensor array design for complex sensing tasks, *Annu. Rev. Anal. Chem.* **8** (2015) 287–310, <https://doi.org/10.1146/annurev-anchem-062011-143205>.
- T. Baur, M. Bastuck, C. Schultealbert, T. Sauerwald, A. Schütze, Random gas mixtures for efficient gas sensor calibration, *J. Sens. Sens. Syst.* **9** (2020) 411–424, <https://doi.org/10.5194/jsss-9-411-2020>.
- S. Marco, A. Gutierrez-Galvez, Signal and data processing for machine olfaction and chemical sensing: a review, *IEEE Sens. J.* **12** (2012) 3189–3214, <https://doi.org/10.1109/JSEN.2012.2192920>.
- I. Rodriguez-Lujan, J. Fonollosa, A. Vergara, M. Homer, R. Huerta, On the calibration of sensor arrays for pattern recognition using the minimal number of experiments, *Chemom. Intell. Lab. Syst.* **130** (2014) 123–134, <https://doi.org/10.1016/j.chemolab.2013.10.012>.
- M. Padilla, A. Perera, 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: *Proceedings of the Int. Jt. Conf. Neural Networks (IJCNN 2010)*, 2010.
- O. Tomic, T. Eklöv, K. Kvaal, J.-E. Haugen, Recalibration of a gas-sensor array system related to sensor replacement, *Anal. Chim. Acta* **512** (2004) 199–206, <https://doi.org/10.1016/j.jca.2004.03.001>.
- R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, S.D. Brown, J. Ferré, Transfer of multivariate calibration models: a review, *Chemom. Intell. Lab. Syst.* **64** (2002) 181–192.
- 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, [https://doi.org/10.1016/S0925-4005\(00\)00617-1](https://doi.org/10.1016/S0925-4005(00)00617-1).
- Y. Wang, D.J. Veltkamp, B.R. Kowalski, Multivariate instrument standardization, *Anal. Chem.* **63** (1991) 2750–2756, <https://doi.org/10.1021/ac00023a016>.
- L. Zhang, F. Tian, C. Kadri, B. Xiao, H. Li, L. Pan, H. Zhou, On-line sensor calibration transfer among electronic nose instruments for monitoring volatile organic chemicals in indoor air quality, *Sens. Actuators B Chem.* **160** (2011) 899–909, <https://doi.org/10.1016/j.snb.2011.08.079>.
- R.W. Kennard, L.A. Stone, Computer aided design of experiments, *Technometrics* **11** (1969) 137–148, <https://doi.org/10.1080/00401706.1969.10490666>.
- J. Fonollosa, L. Fernández, A. Gutiérrez-Gálvez, R. Huerta, S. Marco, Calibration transfer and drift counteraction in chemical sensor arrays using direct standardization, *Sens. Actuators B Chem.* **236** (2016) 1044–1053, <https://doi.org/10.1016/j.snb.2016.05.089>.
- L. Fernandez, S. Guney, A. Gutierrez-Galvez, S. Marco, Calibration transfer in temperature modulated gas sensor arrays, *Sens. Actuators B Chem.* **231** (2016) 276–284.
- D. Zhang, D. Guo, K. Yan, D. Zhang, D. Guo, K. Yan, Learning classification and regression models based on transfer samples, *Breath. Anal. Med. Appl.* (2017) 113–135, [https://doi.org/10.1007/978-981-10-4322-2\\_7](https://doi.org/10.1007/978-981-10-4322-2_7).
- L. Zhang, Y. Liu, P. Deng, Odor recognition in multiple E-nose systems with cross-domain discriminative subspace learning, *IEEE Trans. Instrum. Meas.* **66** (2017) 1679–1692, <https://doi.org/10.1109/TIM.2017.2669818>.
- D. Zhang, D. Guo, K. Yan, A transfer learning approach for correcting instrumental variation and time-varying drift, *Breath. Anal. Med. Appl.* (2017) 137–156, [https://doi.org/10.1007/978-981-10-4322-2\\_8](https://doi.org/10.1007/978-981-10-4322-2_8).
- A. Solórzano, R. Rodríguez-Pérez, M. Padilla, T. Graunke, L. Fernandez, S. Marco, J. Fonollosa, Multi-unit calibration rejects inherent device variability of chemical sensor arrays, *Sens. Actuators B Chem.* **265** (2018) 142–154, <https://doi.org/10.1016/j.snb.2018.02.188>.
- J. Burgués, S. Marco, Multivariate estimation of the limit of detection by orthogonal partial least squares in temperature-modulated MOX sensors, *Anal. Chim. Acta* **1019** (2018) 49–64, <https://doi.org/10.1016/j.aca.2018.03.005>.
- J. Burgués, J.M. Jiménez-Soto, S. Marco, Estimation of the limit of detection in semiconductor gas sensors through linearized calibration models, *Anal. Chim. Acta* **1013** (2018) 13–25, <https://doi.org/10.1016/J.ACA.2018.01.062>.
- F.I.S. Inc, FIS Gas Sensor SB-500–12, 2017.
- P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, *Anal. Chim. Acta* **185** (1986) 1–17, [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9).
- S.J. Leon, Å. Björck, W. Gander, Gram-Schmidt orthogonalization: 100 years and more, *Numer. Linear Algebr. Appl.* **20** (2013) 492–532, <https://doi.org/10.1002/nla.1839>.
- G.H. Golub, C. Reinsch, *Singular value decomposition and least squares solutions*, in: *Linear Algebr.*, Springer, 1971, pp. 134–151.
- L.A. Currie, Detection: international update, and some emerging di-lemmas involving calibration, the blank, and multiple detection decisions, Contribution of the National Institute of Standards and Technology; not subject to copyright. Based on an invited lecture at t, *Chemom. Intell. Lab. Syst.* **37** (1997) 151–181, [https://doi.org/10.1016/S0169-7439\(97\)00009-9](https://doi.org/10.1016/S0169-7439(97)00009-9).
- R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Proceedings of the Fourteenth Int. Jt. Conf. Artif. Intell.*, **2** (1995), 1137–1143.

**Albert Miquel Ibarz** a Fundamental Physics student at the University of Barcelona. Since 2019 he has been an internship student at Institute for Bioengineering of Catalonia (IBEC) under the guidance of Dr. Santiago Marco and Dr. Javier Burgués. He is passionate about artificial intelligence, data science, philosophy, and their application to bioengineering, neuroscience and social sciences.

**Javier Burgués** received the BSc. degree in Telecommunication Engineering from the University Autónoma of Madrid, in 2010, the MSc. degree in Computer Science from the University of Southern California, in 2013, and the Ph.D. degree in Engineering and Applied Sciences from the University of Barcelona, in 2019. He currently works as R&D Technical Lead at ScioSense Germany GmbH, where he is responsible for the development of next-gen environmental sensors for automotive applications. His main research interests include signal processing and pattern recognition for chemical sensor data, electronic design, artificial intelligence, integration of chemical sensors into robotic platforms, and algorithm development for localization and mapping of chemical sources. More at <https://javierburgues.com>

**Santiago Marco** completed his university degree in Physics (1988) and Ph.D. (1993) from the University of Barcelona (UB). He held a European Human Capital Mobility grant for a

postdoctoral position at the Department of Electronic Engineering at the University of Rome "Tor Vergata" working on Electronic Noses. In 1995, he became Associate Professor at the Department of Applied Physics and Electronics at UB. In 2004 he had a sabbatical leave at AIRBUS-Innovation Works, Munich, working on Ion Mobility Spectrometry. In 2008 he was appointed leader of the Signal and Information Processing for Sensing Systems Lab at the Institute for Bioengineering of Catalonia. From 2020 he is Full Professor at the Department of Electronics and Biomedical Engineering at UB. His research concerns the development of signal/data processing algorithmic solutions for smart chemical sensing based in sensor arrays or microspectrometers integrated typically using Microsystem Technologies. He has published around 130 archival journals and around 250 conference papers. (more at <http://ibecbarcelona.eu/sensingsys>).