Analytica Chimica Acta 1013 (2018) 13–25

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Estimation of the limit of detection in semiconductor gas sensors through linearized calibration models

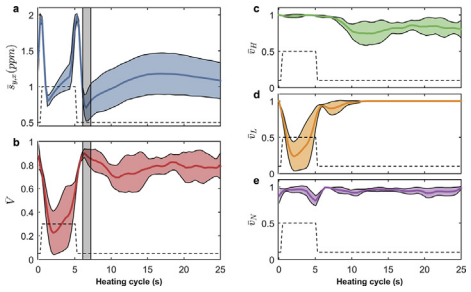
Javier Burgués [a, b, *], Juan Manuel Jiménez-Soto [b], Santiago Marco [a, b]

[a] Department of Electronics and Biomedical Engineering, Universitat de Barcelona, Marti i Franqués 1, 08028, Barcelona, Spain
[b] Signal and Information Processing for Sensing Systems, Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028, Barcelona, Spain

HIGHLIGHTS

- A methodology to estimate the LOD in non-linear chemical sensors is proposed.
- A set of statistical tests verify the main LOD assumptions at each detection point.
- The methodology was applied to a scenario of carbon monoxide and variable humidity.
- The optimum univariate detection point differs from manufacturer recommendations.
- The temporal stability of the LOD was explored.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 4 August 2017
Received in revised form
25 January 2018
Accepted 29 January 2018
Available online 7 February 2018

Keywords:
Semiconductor gas sensors
Metal-oxide sensors
Limit of detection
Non-linear
Humidity interference
Temperature modulation

ABSTRACT

The limit of detection (LOD) is a key figure of merit in chemical sensing. However, the estimation of this figure of merit is hindered by the non-linear calibration curve characteristic of semiconductor gas sensor technologies such as, metal oxide (MOX), gasFETs or thermoelectric sensors. Additionally, chemical sensors suffer from cross-sensitivities and temporal stability problems. The application of the International Union of Pure and Applied Chemistry (IUPAC) recommendations for univariate LOD estimation in non-linear semiconductor gas sensors is not straightforward due to the strong statistical requirements of the IUPAC methodology (linearity, homoscedasticity, normality). Here, we propose a methodological approach to LOD estimation through linearized calibration models. As an example, the methodology is applied to the detection of low concentrations of carbon monoxide using MOX gas sensors in a scenario where the main source of error is the presence of uncontrolled levels of humidity.

1. Introduction

Many potential applications of gas sensors involve the detection of very low concentrations of the target gases. For instance, in applications related to environmental monitoring the United States Environmental Protection Agency (EPA) specifies four principal pollutants and their corresponding maximum exposure

* Corresponding author. Department of Electronics and Biomedical Engineering, Universitat de Barcelona, Marti i Franqués 1, 08028, Barcelona, Spain.
E-mail address: jburgues@ibecbarcelona.eu (J. Burgués).

https://doi.org/10.1016/j.aca.2018.01.062
0003-2670/

concentrations (hourly) in outdoor air: 9 ppm of carbon monoxide (CO), 100 ppb of nitrogen dioxide (NO_2), 70 ppb of ozone (O_3) and 75 ppb of sulfur dioxide (SO_2) [1]. In the case of the commercial automotive market, a sensing range of 5 ppm is required to meet NO_2 regulations [2]. CO detectors integrated in smoking cessation monitors must comply with a detection threshold of 10 ppm of CO in exhaled breath in order to differentiate a smoker from a non-smoker [3].

The figure of merit that characterizes the ability of an instrument to differentiate between the absence of a substance (a blank value) and measurements with low concentration of the target gas is the limit of detection (LOD). The LOD is usually reported in concentration units. The LOD is extremely relevant for the applications described above because only instruments with LODs below the thresholds specified for the target gases could satisfy the applications requirements. The International Union of Pure and Applied Chemistry (IUPAC) defines the LOD for univariate signals in terms of statistical hypothesis tests, considering measurements as random variables [4]. IUPAC simplified formulas are derived from the upper prediction interval of the blank in a linear regression model which transforms the observed responses into concentration values. For binary detection (i.e. analyte present or not present) the estimated net signal or concentration is compared against the critical level (L_C), which only controls the false positives. The LOD (L_D) is, however, a more representative measure of the inherent detection capability of the analytical system because it also considers the false negatives. The distribution of the blanks and the LOD in the signal domain are modelled using a t-distribution with a maximum of 5% of false positives and 5% of false negatives. There are three strong underlying assumptions in the IUPAC LOD formulas based on the use of linear regression models for prediction purposes:

a) **Linearity**: A linear model can be fit to the data only if the relationship between the independent and dependent variables is linear. In other words, the residual errors of a linear fitting should have zero mean.
b) **Homoscedasticity** of the residual errors: If the variance of the residual errors is not constant along the independent variable, the standard error will be biased. A biased standard error will lead to incorrect prediction intervals. Extensions for the heteroscedastic case are described in the IUPAC norm but only for a particular type of heteroscedasticity: variance proportionate to response.
c) **Normality** of the residual errors. For Least Squares estimators to be maximum likelihood estimators of the linear fit the errors should be normal. Additionally, the use of *t*-test for hypothesis testing also requires that the residual errors are normally distributed.

Violation of any of these assumptions may bias the confidence and prediction intervals of the linear regression model. Because the IUPAC LOD is based on the upper prediction interval of the blank, the LOD will be biased too. Even when the assumptions are satisfied, it is important to remark that the LOD estimate has a domain of validity closely tied to the set of conditions covered by the calibration points, including obviously the conditions for the blanks. Departure from those conditions casts doubt on the validity of the LOD estimate (this is true for any analytical instrument).

Instruments based on linear devices such as electrochemical gas sensors usually satisfy these assumptions and therefore the LOD estimates are trustworthy. However, this is not the case for certain chemical sensor technologies. A clear example are semiconductor gas sensors such as, metal oxide semiconductor (MOX) [5], gasFETs [6] or thermoelectric sensors [7]. The non-linear response with respect to concentration will in many cases violate the linearity assumption. In addition, some of these technologies also present cross-sensitivity to environmental interferences (e.g. temperature, humidity or barometric pressure) and lack of stability [8–10]. This might yield non-normal or heteroscedastic residual errors in calibration.

In some cases, the LOD is calculated when the instrument is calibrated in laboratory conditions and then it is (incorrectly) assumed that in future and field operation the LOD will still be the same as in calibration. But, in field operation the range of variation of interferences such as temperature and humidity can go beyond the calibration range or the instrument might have drifted. Since in many chemical sensor applications we expect the system to run without human intervention for extended period of times (months or even years), a relevant question is the stability of the LOD in time. The validity of the LOD estimates in MOX sensors is a critical issue due to the intrinsic drift and cross-sensitivities to environmental conditions and other gases. However, this problem has not been properly addressed in the literature. Some works blindly trust the LOD obtained in calibration [11–14]. Other authors have approached the LOD validation based on the theory of hypothesis testing measuring the probabilities of false positive and false negatives, however, using relatively small samples sizes [15]. The results obtained with this approach are not trustworthy unless there are many measurements and replicates corresponding to concentration standards near the LOD, which is usually unknown.

Probably due to these issues, the LOD is rarely reported in archival papers concerning MOX sensors. A search in the current literature (as of July 2017) reveals that only a handful of articles deal with LOD estimation in MOX sensors [12–14]. Nevertheless, most of these works often ignore the IUPAC assumptions and, for simplicity, assume they hold in their experimental data. Despite the difficulties associated to the estimation of the LOD, the use of MOX sensors in applications that require a certain LOD value has obvious interest. MOX sensors are long-lasting devices which can be cheaply manufactured in a miniaturized form factor while achieving low power consumption (thanks to MEMS technology). The market of MOX sensors has been indeed very active in the last years due to the potential integration into portable devices such as smartphones, wearables or tablets. MOX sensors are being used in diverse applications that range from environmental monitoring [12,16,17], to safety and security [18], food [19] or medical applications [20]. Areas with high levels of pollutants could be easily localized thanks to portable measuring stations based on MOX sensors [21]. In the biomedical field, MOX sensors are currently proposed to monitor breath biomarkers for diagnostic of many diseases, including several types of cancer [22].

The combination of pulsed temperature modulation and univariate CO detection at low temperatures was pioneered in commercial products by Figaro Engineering in the successful product TGS-203 [23]. The marketing of this product with this recommended operational mode started as early as 1980. The sensor manufacturer shows that the best selectivity to CO is achieved when the sensor element temperature is kept under 100 °C. However, at this temperature the sensor becomes very susceptible to water vapor. To eliminate this influence, they propose to use a cycling high/low voltage pulses. They claim that the high heater voltage cleans the sensor surface and removes the influence of water vapor, while the low heater voltage conditions the sensor for measuring CO. This operation mode is still used by several manufacturers like Figaro (TGS3870 A-04), FIS (SB-500-12) or SGX Sensortech (MiCS sensors). Newer sensors using either beads or micromachined substrates are faster from a thermal point of view and they allow to use shorter thermal cycles, but the recommended operation mode remains basically the same. In all cases, the

detection of CO remains univariate for simplicity of use. The reduction of humidity influence in this operation mode has been confirmed later [24–26]. It is also true that the selectivity of MOX sensors can be improved by temperature modulation followed by a multivariate prediction model. However, from a user perspective, univariate detection has the advantage of simplicity and probably for this reason is still the recommended operation mode by the manufacturers. In this work, we will be using FIS SB-500-12 sensors. Our measurements are based on univariate features: single detection points at certain locations within the temperature modulation cycle. Although there is potential for improving the LOD by using the full response pattern, it is not the scope of this work to explore the estimation of the LOD from the multivariate response pattern. The problem of multivariate LOD estimation is that there is no well-accepted LOD formula, as it is the case for univariate estimation. There are several multivariate LOD proposals [27] and most of them produce different outputs [28]. Instead of dealing with the multivariate case, the idea behind this work is to optimize the location of the univariate detection point using the standard IUPAC univariate formula for LOD estimation and then applying a set of statistical tests to ensure that the underlying LOD assumptions are satisfied.

The present work, therefore, aims to several purposes: (i) to recall the main ideas underlying the IUPAC theory regarding the univariate limit of detection and their impact in non-linear chemical sensing; (ii) to give guidelines to detect situations in which heteroscedastic, non-normal and biased residuals in calibration hinder the application of the simplified LOD formulas; (iii) to propose a set of statistical tests to verify that the obtained LOD values in calibration are applicable to validation data; (iv) finally to illustrate the problems associated to the blind use of the LOD in future data when using MOX sensors. For that, we proposed a scenario of carbon monoxide (CO) detection under variable humidity conditions, which is representative of various real applications. In the selected scenario, humidity is the main interference and we consider the LOD as a figure of merit to optimize the measurement point for CO sensing reducing the effect of the humidity. We would like to stress that here humidity is a random perturbation and not a systematic influence, and it is not the interest of this work to estimate the humidity, but only to ascertain how humidity variations have an impact on the LOD. Obviously, humidity interference is a major concern for MOX performance. In the case of CO sensing, this has been extensively reviewed by Ref. [29]. Our contribution focuses on the methodological aspects related to the estimation of the LOD in chemical sensors through linearized models. We will present a methodology for evaluating the properties of the sensor data relevant to the LOD formulas.

2. Experimental and methods

To illustrate the challenges described in the introductory section we designed an experimental scenario of CO detection under random humidity conditions using temperature-modulated MOX sensors. The experimental setup was comprised of a test bench where dynamic mixtures of CO in humid synthetic air were generated, an electronic board composed of seven MOX sensors and a temperature/humidity sensor. The data was processed using MATLAB R2009 (The MathWorks) and libraries from the Statistics and Machine Learning Toolbox.

2.1. Sensor board

An electronic board including seven MOX sensors for CO detection (SB-500-12, FIS Inc.), a temperature/humidity sensor (SHT75, Sensirion AG) and the necessary conditioning circuits was assembled in-house. The commercial MOX sensors used in this work are manufactured with large tolerances in baseline (one order of magnitude) and sensitivity (a factor of two) [30]. To obtain statistics of the performance of the sensor family (e.g. mean and variance), several units of the same MOX sensor model were integrated in the sensor board. The Sensirion sensor provided reference humidity and temperature values with tolerance below 1.8% r.h. and 0.5 °C, respectively, every 5 s. According to the manufacturer datasheet, the long-term drift of the SHT75 sensor is less than 0.5% r.h./year and 0.04 °C/year. The humidity values will not be used for sensor data compensation, only to obtain a direct measurement of the humidity values within the sensor chamber.

The heater voltage of the MOX sensors was modulated in the range 0.2–0.9 V in cycles of 25 s, following the manufacturer recommendations (0.9 V for 5s, followed by 0.2 V for 20s) (see Fig. 1). For simplicity, in this work we are not implementing a closed loop control of the heater temperature which has been shown to produce more accurate results [31]. The seven MOX sensors were preheated for one week before starting the experiments to reach stable responses. The manufacturer recommends measuring the sensor response at a single detection point located at the end of low level of the heating cycle (see Fig. 1). The choice of this detection point, that we will call “nominal working point”, is specific to CO detection as the sensors are more sensitive to CO at low temperatures. Nonetheless, we recorded the sensor response during the whole heating cycle to evaluate the performance of the sensor in other measurement points. For that, the MOX read-out circuits consisted of voltage dividers with 1MΩ load resistors. The value of the load resistor allowed proper quantification of the sensor resistance, considering the large dynamic range (20 kΩ–10 MΩ) present at low concentrations of the analyte with the chosen temperature modulation waveform. The output voltage of the sensors was sampled at 3.5 Hz using an Agilent HP34970A/34901A DAQ configured at 15 bits of precision and input impedance greater than 10 GΩ. In this configuration, the errors introduced in the voltage measurements were considered negligible compared to the intrinsic variability of the sensor resistance due to the chemical transduction process.

2.2. Experimental design

Previous experience by the authors on CO sensing suggested that the LOD could be around 5 ppm or even smaller. At this point, it is important to remember that for LOD estimation the IUPAC

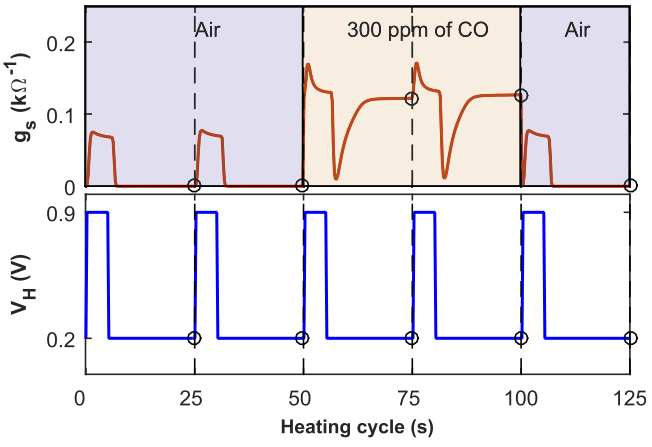


Fig. 1. FIS SB-500-12 operating mode and output signal. The top subplot represents the sensor conductance during an exposure to air (50 s), 300 ppm of carbon monoxide (50 s) and air (25 s). The bottom subplot represents the heater voltage. The black circles indicate the detection points recommended by the manufacturer.

recommends that the calibration points should be placed close to the LOD [4]. For this reason, we selected a calibration range of 0–9 ppm. At this point, it is important to remark that 9 ppm is not the upper limit for the detection of CO with MOX sensors, but only a reduced range to follow IUPAC recommendations for LOD estimation. In fact, MOX sensors are routinely used for CO detection in domestic premises with measurement ranges up to 400 ppm [29].

The experimental design was based on the literature recommendations for LOD estimation of using between five and ten calibration standards, eight to ten replicates [32] and equidistant concentration levels [33]. To generate the experimental dataset, five linearly equally spaced calibration standards in the range 0–9 ppm were selected ({0.0, 2.2, 4.4, 6.7 and 8.9} ppm). Ten repetitions were performed at each concentration, with each repetition having a relative humidity value randomly chosen from a uniform distribution between 15 and 70% r.h. (see Fig. 2). Hence, each experiment was composed of 50 conditions (5 concentrations x 10 repetitions) which were performed in random order to reduce the impact of factors that were not explicitly accounted for in the experimental design. The duration of each exposure was 15 minutes. A single calibration experiment lasted approximately 12 hours and was repeated 13 times for repeatability and reproducibility estimation. The dataset was therefore composed of 650 measurements (13 experimental days x 50 measurements/day) which were collected over 15 natural days.

2.3. Generator of dynamic gas mixtures

Gas mixing was performed using mass flow controllers (MFC), which controlled three different gas streams (CO, wet air and dry air). These streams were delivered from high quality pressurized gases in cylinders. The CO bottle contained 1600 ppm of CO diluted in synthetic air with 21 ± 1% O_2. The wet and dry air streams were both delivered from a synthetic air bottle with 99.995% purity and 21 ± 1% O_2. The humidification of the wet stream was based on the saturation method using a glass bubbler (Drechsler bottles) [34]. To reach higher humidity levels, the Drechsler bottles were submerged into warm water at 42 °C. The selected MFCs (EL-FLOW Select, Bronkhorst) had full scale flow rates of 1000 ml_n/min for the dry and wet air streams and 3 ml_n/min for the CO channel. Check valves protected the MFCs against reversed flow in the system. The gas mixtures at different values of concentration and humidity were transferred into a small-sized polytetrafluoroethylene (PTFE) test chamber (250 cm^3 internal volume) containing the sensors under test. Fig. 3 shows a schematic plan of the measuring system.

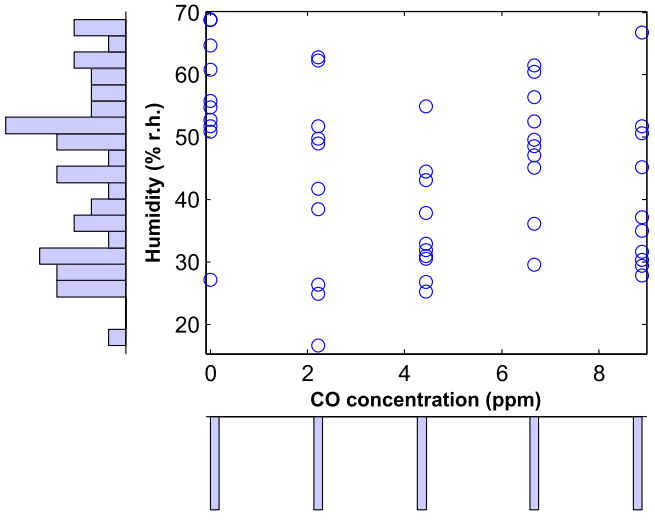


Fig. 2. Scatter plot of the nominal CO concentration (ppm) and the measured humidity (% r.h.) in the first experimental day. The marginal distributions of the nominal CO concentration and the measured humidity are displayed as univariate histograms on the horizontal and vertical axes of the scatter plot, respectively.

At the beginning of each experiment, the gas chamber was cleaned for 15 min using a stream of synthetic air at a flow rate of 240 ml_n/min. After that, the gas mixtures were released at a constant flow rate of 240 ml_n/min for 15 min each. The temperature variations inside the gas chamber, for each experiment, were below 3 °C. The highest relative uncertainty in CO concentration was 5.5% for the first non-zero concentration (2.2 ppm), which corresponds to an absolute uncertainty of 0.12 ppm. A desktop computer running Labview 2015 (National Instruments) commanded the MFCs and integrated the readings from the humidity sensor into a log file.

During each experimental condition, the seven sensor signals, indicative of the gas conditions presented to the sensors, were continuously sampled. In order to evaluate the sensor response to a certain experimental condition, portions of the sensor signal corresponding to a full heating cycle (henceforth called patterns) were extracted from the original signals. To ensure that the extracted patterns correspond to stable CO levels, only the last three cycles of each 15-minute condition were taken into account. These three fragments were averaged to reduce instrumental noise, producing a single multivariate conductance pattern $g_s(t)$, $t \in [0, 25]$ s associated to a stable gas concentration$c$ (ppm). The readings from the humidity sensor were also averaged during the same three heating cycles to provide a reference relative humidity value h (% r.h.).

2.4. Calibration models

As we said earlier in the introduction, MOX sensors present a non-linear response which might be best fit by non-linear models. Nevertheless, within a narrow range of small concentrations the sensor response can be considered quasi-linear and the use of linear models can be justified. This allowed us to use the sound LOD theory already developed for univariate linear calibration models.

In this work, we only considered linear univariate calibration models where the observed response y is predicted from the experimental analyte concentration x by a linear regression model of the form:

$$y = B + Ax + e_y \quad (1)$$

Where B and A are the real (unknown) regression coefficients and e_y is the residual error of y. In Eq. (1) it is assumed that the analyte concentration is error-free because the uncertainty in the preparation of concentrations was negligible (see Section 2.3).

For CO detection, the most widely used method is the Clifford-Tuma model (Eq. (2)) [35]. This empirical model uses a linear relationship between the logarithm of the sensor conductance g_s ($k\Omega^{-1}$) and the logarithm of the analyte concentration c (ppm):

$$\log(g_s) = \log(g_{air}S) + \beta \log(c) = \alpha + \beta \log(c) \quad (2)$$

Where g_{air} ($k\Omega^{-1}$) is the sensor conductance in clean air, S is a gas depending parameter and β is the so-called sensitivity to the gas. Log-log models such as the Clifford-Tuma model provide in general better fitting than normal models but their use for LOD estimation is questionable because the blanks ($c = 0$ ppm) need to be excluded due to the term $\log(c)$. As pointed out in the IUPAC recommendation [4], "*The blank is one of the most crucial quantities in trace analysis, especially in the region of the Detection Limit … Inadequate attention to the magnitude and variability of the overall*

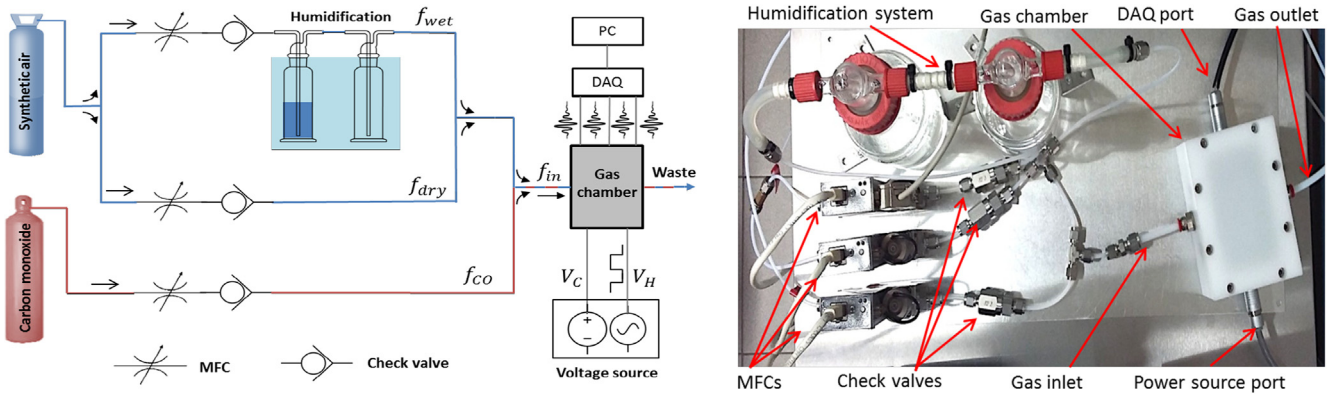


Fig. 3. Experimental test bench for the generation of dynamic gas mixtures and acquisition of the sensor signals. Left: Block diagram. Right: Picture.

blank, may lead to severe underestimation of the detection limit."

To introduce blanks into the model, we propose a log-normal model in which the logarithm is applied only to the sensor conductance (Eq. (3)):

$$\log\left(\frac{g_s}{g_{air}}\right) = \mathrm{B} + \mathrm{A}c + \varepsilon \tag{3}$$

where B is the intercept, A is the sensitivity and ε is the error term. Dividing g_s by g_{air} corrects the potential changes in sensor baseline. g_{air} was estimated from a blank sample with random humidity measured at the beginning of the day and all the measurements of that day were divided by this value. The use of Eq. (3) for LOD estimation purposes is valid only for a narrow range of small concentrations and for certain working points in which the sensor response (g_s) can be linearized without applying the logarithm to the concentration.

The precision of the calibration function was assessed by the residual standard deviation ($s_{y,x}$) and the estimated standard deviation of the estimated slope ($\sigma_{\widehat{A}}$) and the estimated intercept ($\sigma_{\widehat{B}}$), as recommended by the IUPAC [36]. Because $\sigma_{\widehat{A}}$ and $\sigma_{\widehat{B}}$ only depend on $s_{y,x}$ and the calibration design -which was fixed-, minimizing $s_{y,x}$ also minimizes $\sigma_{\widehat{A}}$ and $\sigma_{\widehat{B}}$. Therefore, we only report $s_{y,x}$ (Eq. (4)).

$$s_{y,x} = \sqrt{\sum_{i=1}^{n}\frac{(y_i - \widehat{y}_i)^2}{n-2}} \tag{4}$$

where n denotes the number of calibration samples, and y_i and $\widehat{y}_i$ are the nominal and fitted value for the calibration sample$i$, respectively.

2.5. Limit of detection (LOD)

In the IUPAC recommendations [37], the LOD is defined as the "smallest measure that can be detected with reasonable certainty for a given analytical procedure". IUPAC derives the LOD in terms of statistical hypothesis testing with default values for the probabilities of false positives (α) and false negatives (β) both equal to 0.05. The fundamental relations (Eqs. (5) and (6)) are formulated based on a generic symbol (L) that represents either net signal (blank-corrected) or analyte concentration:

$$P\left(\widehat{L} > L_C | L = 0\right) = \alpha \tag{5}$$

$$P\left(\widehat{L} < L_C | L = L_D\right) = \beta \tag{6}$$

Where $\widehat{L}$ is an estimate of L, L_C is the threshold used for the detection decision (i.e. analyte present or not) and L_D is the limit of detection or LOD. Assuming that $\widehat{L}$ is normally distributed with known variance (σ), the value of L_C is given by Eq. (7), after solving for L_C in Eq. (5)

$$L_C = z_{1-\alpha}\sigma_0 \tag{7}$$

Where $z_{1-\alpha}$ is the critical value of the normal distribution for significance level α and σ_0 is the true standard deviation of $\widehat{L}$ when $L = 0$. Similarly, L_D can be found by first solving for L_C in Eq. (6):

$$L_C = L_D - z_{1-\beta}\sigma_D \tag{8}$$

Where $z_{1-\beta}$ is the critical value of the normal distribution for significance level β and σ_D is the true standard deviation of $\widehat{L}$ when $L = L_D$. Then, solving for L_D in Eq. (8):

$$L_D = L_C + z_{1-\beta}\sigma_D \tag{9}$$

Inserting Eq. (7) into Eq. (9) yields:

$$L_D = z_{1-\alpha}\sigma_0 + z_{1-\beta}\sigma_D \tag{10}$$

In many practical applications, the true values σ_0 and σ_D in Eq. (10) are unknown, so a Student t-distribution is used to approximate the Gaussian curve. The critical value of the Gaussian distribution is replaced by the one-sided t-critical value ($t_{1-\alpha,\upsilon}$) for the chosen confidence level (α) and degrees of freedom (υ). Similarly, the true standard deviations σ_0 and σ_D are replaced by the corresponding estimates s_0 and s_D, respectively. Because s_D corresponds to the dispersion of $\widehat{L}$ at L_D, which is unknown (in fact, L_D is what we are trying to estimate), s_D is replaced by s_0 (assuming homoscedasticity). Finally, recalling that we chose $\alpha = \beta$, the formula of Eq. (10) is simplified to Eq. (11):

$$L_D = 2t_{1-\alpha,\upsilon}s_0 \tag{11}$$

Eq. (11) is a generic LOD formula that can be used to determine the LOD in the signal domain or in the concentration domain. The LOD is usually reported in concentration units but the measuring instruments report signals instead. Given a set of discrete measurements, there are two ways to compute the LOD: (a) Convert the instrument signals to concentration values and estimate the LOD in

the concentration domain using Eq. (11) or (b) compute the LOD in the signal domain using Eq. (11) and then transform the "signal LOD" to a "concentration LOD". Signals and concentrations can be related to each other by the use of a calibration curve. The IUPAC only considers the linear calibration function (c.f. Eq. (1)). In this case, the concentration ($\hat{x}$) corresponding to an observed gross signal (y) is estimated by:

$$\hat{x} = \frac{y - \hat{B}}{\hat{A}} = \frac{\hat{S}}{\hat{A}} \tag{12}$$

Where $\hat{B}$ is the estimated intercept, $\hat{A}$ is the estimated slope of the calibration function and $\hat{S}$ is the net (blank-corrected) response. Application of the fundamental LOD definitions (i.e. Eqs. (5) and (6)) to the concentration domain (i.e. option (a) above) requires normality in the estimator $\hat{x}$, which is a strong assumption considering that $\hat{A}$ will be estimated with uncertainty. Therefore, the IUPAC recommends the computation of the LOD in the signal domain (Eq. (11)) and the posterior transformation to concentration units through Eq. (12).

The only variable that we need to estimate in Eq. (11) is s_0, the standard deviation of the estimated net response $\hat{S}$ when the analyte is absent ($x = 0$). From linear regression [36], we know that the standard deviation of a predicted value y_p at position x_i is:

$$s_{y_p} = s_{y,x}\sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_j - \bar{x})^2}} \tag{13}$$

Where $s_{y,x}$ is the standard error of regression, n is the number of samples, x_j are the calibration standards and $\bar{x}$ is the mean calibration concentration. Solving for a blank ($x_i = 0$) yields:

$$s_0 = s_{y,x}\sqrt{1 + \frac{1}{n} + \frac{\bar{x}^2}{\sum (x_j - \bar{x})^2}} = s_{y,x}\eta \tag{14}$$

where $\eta = \sqrt{1 + \frac{1}{n} + \frac{\bar{x}^2}{\sum (x_j - \bar{x})^2}}$ is a design parameter reflecting the calibration design (i.e. relative position of the calibration standards and number of samples) [37].

Substituting Eq. (14) into Eq. (11), yields the LOD in the signal domain (S_D):

$$S_D = 2t_{1-\alpha}s_{y,x}\eta \tag{15}$$

The LOD in the concentration domain is obtained by inserting Eq. (15) into Eq. (12):

$$x_D = \frac{S_D}{\hat{A}} \cdot \left(\frac{K}{I}\right) = \frac{2t_{1-\alpha}s_{y,x}\eta}{\hat{A}} \cdot \left(\frac{K}{I}\right) \tag{16}$$

The correction factor $\left(\frac{K}{I}\right)$ accounts for the uncertainty in the regression slope ($\hat{A}$) [37] and it is defined in Eq. (17):

$$\frac{K}{I} = \frac{1 - r(B,A) \cdot \left(\sigma_{\hat{B}}/\sigma_0\right) \cdot \left[t_{1-\alpha,v} \cdot \left(\sigma_{\hat{A}}/A\right)\right]}{1 - \left[t_{1-\alpha,v} \cdot \left(\sigma_{\hat{A}}/A\right)\right]^2} \tag{17}$$

Where r(B, A) is the correlation coefficient between the slope and intercept of the calibration line (Eq. (18)), $\sigma_{\hat{B}}/\sigma_0$ is the relative standard deviation of $\hat{B}$ (estimated intercept), σ_0 is the standard deviation of the blanks and $\sigma_{\hat{A}}/A$ is the relative standard deviation of $\hat{A}$.

$$r(B,A) = \bar{x} \Big/ \sqrt{\sum_{i=1}^{n} x_i^2 / n} \tag{18}$$

The interpretation of the correction factor $\left(\frac{K}{I}\right)$ is as follows. If the slope (A) is estimated with low uncertainty (i.e. $(\sigma_{\hat{A}}/A) \to 0$), the term in square brackets in Eq. (17) goes to zero, and both K and I are equal to 1, yielding $\left(\frac{K}{I}\right) = 1$. In this case, the LOD is not affected by the correction factor. At the other extreme, when the uncertainty in estimating A is very high (as $\sigma_{\hat{A}}/A$ approaches $1/t_{1-\alpha,v}$), the denominator (I) goes to zero, the ratio $\left(\frac{K}{I}\right)$ goes to infinity and x_D is unbounded. As it can be seen in Eq. (16), the estimated slope ($\hat{A}$) is the only parameter of the regression line that affects x_D. However, the correction factor includes the effect of the uncertainty of $\hat{B}$ when estimating A. If A and B are estimated individually, the correlation coefficient r (B, A) is zero and the ratio $\left(\frac{K}{I}\right)$ only depends on the uncertainty of $\hat{A}$. When A and B are estimated jointly, they will be negatively correlated and the numerator (K) in Eq. (17) will increase with the uncertainty of $\hat{B}$.

2.5.1. Statistical tests to check the underlying LOD assumptions

In the preceding section, we presented a formula for estimating the LOD in the concentration domain, assuming certain properties of the instrumental signals. In particular, the IUPAC LOD definition [4] assumes that the residual errors are unbiased, normally distributed and homoscedastic. The violation of any of these assumptions may yield inconsistent LOD values. In this section, we present a set of statistical tests to systematically verify the underlying LOD assumptions:

- **Homoscedasticity:** Levene's test [38] was used to test the homogeneity of variance across the residuals of the I concentration levels (J samples each). This test assumes normally distributed data but is less sensitive to departures from normality than other tests for comparison of variances.
- **Normality:** The Saphiro-Wilk test [39] was applied to the residuals of each concentration level to assess normality (I tests with J samples each). This test assumes that the sample size is large enough (30 or more) because for small sample sizes normality tests have little power to reject the null hypothesis of normally distributed data. Nonetheless, it has been proved that Saphiro-Wilk test is the most powerful normality test for small sample size [40]. When the residuals at every concentration level are normally distributed the full set of residuals will be normally distributed too. Based on this, an additional Saphiro-Wilk test was applied to all the residuals (one test with IxJ samples) to increase the power of the test.
- **Linearity:** If the data behaves linearly, the residual errors in each concentration level should be unbiased or, in other words, symmetrically distributed around zero. A two-tailed t-test of zero mean [41] was applied to the residuals of each concentration level and to all residuals to verify the linearity assumption. This test assumes normally distributed data. The power analysis of this test indicated that the minimum detectable changes for the group tests (10 samples) and for the full set (50 samples) were 1.5 and 0.3 ppm respectively, which we considered acceptable values.

Table 1
Summary of statistical tests used to verify the LOD assumptions.

Assumption	Test	Null hypothesis	Alternative hypothesis
Homoscedasticity	Levene's test	The multiple data samples have equal variances	At least two of the data samples do not have equal variances
Normality	Saphiro-Wilk test	Sample data is normally distributed	Sample data comes from a non-normal distribution
Linearity	One-sample T-test	The sample data comes from a normal distribution with mean equal to zero	The sample distribution does not have a mean equal to zero

The Holm-Bonferroni correction for multiple comparisons was applied to the family of tests to reduce the false positive rate [42]. Let H_1, H_2, H_3 be the null hypothesis of the homoscedasticity, normality and linearity assumptions (see Table 1), sorted by ascending p-values P_1, P_2, P_3. The Holm-Bonferroni algorithm will first find the minimum index k such that the p-value P_k satisfies the following condition:

$$P_k > \frac{\alpha}{m+1-k} \tag{19}$$

Where m is the number of hypothesis ($m = 3$ in our case) and α is the significance level ($\alpha = 0.05$ in our case). The null hypothesis $H_1, \cdots, H_{k-1}$ are rejected and $H_k, \cdots, H_3$ are not rejected. We define the validity of a hypothesis H, as the Boolean variable v:

$$\begin{aligned} v &= 0, \ \text{if } H \textit{ is rejected} \\ v &= 1, \ \text{if } H \textit{ is not rejected} \end{aligned} \tag{20}$$

v_H, v_N and v_L represent the validity of the homoscedasticity, normality and linearity assumptions, respectively. The joint validity of the LOD assumptions ($v_{H,N,L}$) is a Boolean variable defined as the AND operation (represented by the logic operator $\wedge$) of the marginal validities:

$$v_{H,N,L} = v_H \wedge v_N \wedge v_L \tag{21}$$

2.5.2. Validity of x_D

The value of x_D only makes sense if it is positive, the calibration function is estimated with reasonable accuracy and the underlying LOD assumptions are satisfied. The constraint in Eq. (22) is used to avoid negative values resulting from Eq. (16) due to the random variable $\widehat{A}$.

$$x_D > 0 \tag{22}$$

The constraints in Eqs. (23) and (24) set an upper limit in the uncertainty of the regression parameters (slope and intercept, respectively).

$$\sigma_{\widehat{A}} \Big/ A \leq 1 \tag{23}$$

$$\sigma_{\widehat{B}} \Big/ \sigma_0 \leq 1 \tag{24}$$

The joint validity of the LOD assumptions (see Section 2.5.1) must be equal to one:

$$v_{H,N,L} = 1 \tag{25}$$

The values of x_D which did not satisfy the above constraints were considered invalid. Combining Eqs. (22)–(25), we obtain the formula for the validity of x_D (V):

$$V = (x_D > 0) \wedge \left(\sigma_{\widehat{A}} \Big/ A \leq 1\right) \wedge \left(\sigma_{\widehat{B}} \Big/ \sigma_0 \leq 1\right) \wedge (v_{H,N,L} = 1) \tag{26}$$

Here, V is a Boolean variable which takes the value '1' if x_D is considered valid and '0' otherwise.

2.6. Optimization of the working point

To optimize the working point of a given sensor, we used the measurements of the first experimental day to fit the model described by Eq. (3) for every working point t_i of the response pattern. Then, from the subset of the models that satisfied Eq. (26), we selected the working point that minimized the error of regression $s_{y,x}$ (Eq. (4)). We denote this optimum point as $t_i^*(s)$, to indicate that there is one optimum point for each sensor s.

To report robust indicators of x_D and V, for a given working point and a given sensor s, the individual estimates $x_D(s,d)$ and $V(s,d)$ obtained in different experimental days d ($d = 1, \ldots, N$) were averaged according to Eqs. (27) and (28):

$$\overline{x}_D(s) = \frac{1}{N}\sum_{d=1}^{N} x_D(s,d) \tag{27}$$

$$\overline{V}(s) = \frac{1}{N}\sum_{d=1}^{N} V(s,d) \tag{28}$$

where $\overline{x}_D(s)$ and $\overline{V}(s)$ represent the average LOD and average validity for sensor s, respectively.

2.7. Validation of the optimum working point

Previously we described the assumptions imposed by the IUPAC on the calibration model to reliably calculate a LOD value. The IUPAC recommendations only concern the calibration process and expect that the unseen future data used in the prediction step have the same properties as the calibration data. While this can be the case in linear stable instruments operating under controlled lab conditions, MOX sensors tend to drift over time and are very sensitive to interferences. Slight changes in experimental conditions during the measurements, influences of temperature, pressure or humidity, can also cause the baseline to drift away from its original level. As a result, data obtained days after calibration might be shifted with respect to the calibration data or show different variance (see Fig. 4). In any of these cases, the LOD value obtained in calibration will be biased and the number of false positives and false negatives can be different from the 5% established by the IUPAC norm.

The calibration models built on the kth day were externally validated using the samples of the consecutive days k+1, k+2, ..., N. The concentration of the external validation samples was predicted using the inverse of Eq. (3) and the root mean squared error in prediction (RMSEP) was calculated:

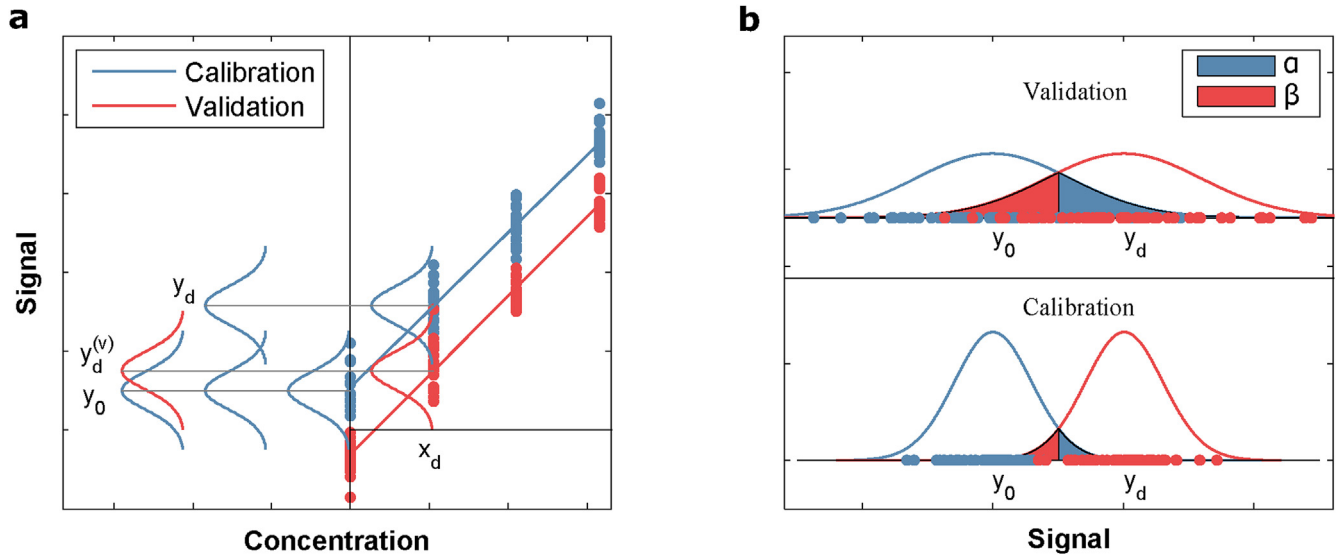


Fig. 4. Simulated examples of degradation of calibration model due to: (a) shifted validation data; (b) increased variance of the validation data. The blue gaussian curve centered at y_0 is the response in blank samples and the red gaussian curve centered at y_d is the response at the LOD level. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

$$\text{RMSEP} = \sqrt{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 / n} \quad (29)$$

Where n denotes the size of the validation set, and x_i and $\hat{x}_i$ are the true and predicted concentration for sample i, respectively.

3. Results and discussion

In order to illustrate the proposed methodology in a real data set and face some of the challenges associated to LOD estimation in non-linear sensors, we designed a scenario representative of real world applications: CO detection under random variable humidity conditions. The working principle of MOX sensors is based on the change of the sensor conductance when the sensor is exposed to

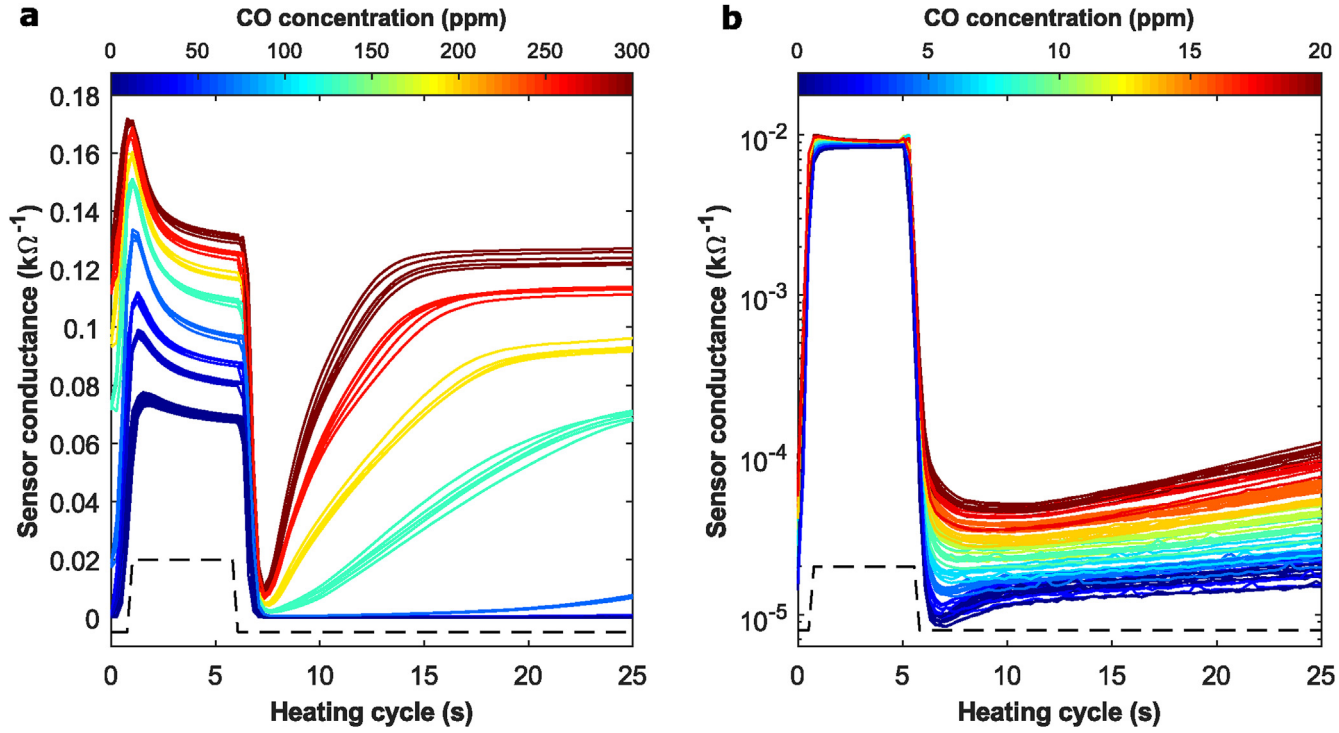


Fig. 5. Multivariate sensor conductance patterns of a FIS SB-500-12 sensor exposed to (a) 0–300 ppm of CO and (b) 0–20 ppm of CO. The colormap indicates the CO concentration (see colorbar at the top of each subplot). The black dashed line represents the heater voltage (a.u.). The logarithm is applied to the sensor conductance in (b) to facilitate the visualization of the low heating temperatures.

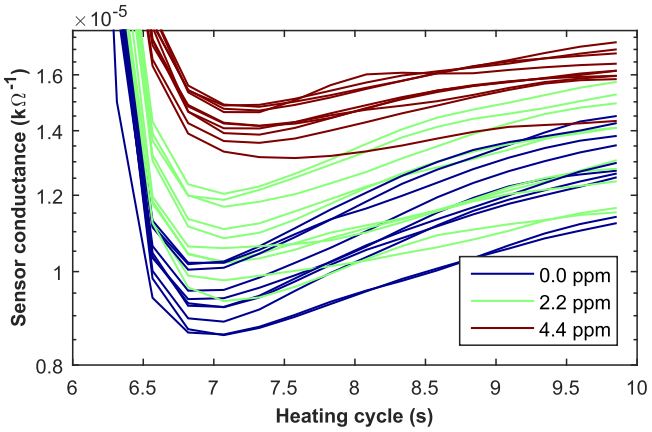


Fig. 6. Zoom-in of the region $t = 6 - 10$ s of the logarithmic sensor conductance patterns shown in Fig. 5b. Only the three lowest concentration levels (0.0, 2.2 and 4.4 ppm) are displayed for visual clarity.

the target gas. The sensitivity to the target gas depends on the temperature (working point) of the sensing surface. When the sensor is exposed to relatively high CO concentrations the sensitivity to CO is maximized at the end of the low heating cycle (Fig. 5a). This is why the nominal detection point recommended by the manufacturer is $t = 25$ s. At lower concentration ranges (Fig. 5b), similar values of the sensitivity are obtained in a range of measurement points in the low level of the heating cycle ($t \in [5, 25]$ s). At the lower end of the concentration range, the separability across concentrations is higher in a narrow range of working points located at the beginning of the low temperatures ($t = 6.5 - 7.5$ s). Because this is barely visible in Fig. 5b and it is a relevant finding, we have created a zoomed-in version of that area only for the three lowest concentrations (see Fig. 6).

At high concentration ranges (e.g. 0–300 ppm), the log conductance – log concentration response curve shows a non-linear behavior, for every working point of the response pattern (Fig. 7a). The response at high working temperatures and at the beginning of the low temperatures ($t < 7.5$ s) follows a slight quadratic behavior (it cannot be seen in the figure), whereas for most working points in the low heating temperatures ($t > 7.5$ s) the response might be approximated by a higher order polynomial. Because the IUPAC univariate LOD formulas are based on a linear regression model, no working point satisfied the LOD assumptions in this concentration range. We found that the logarithm of the sensor conductance in a reduced concentration range (0–9 ppm) linearized the response for most working temperatures. In this reduced concentration range, the working point can affect both the slope of the calibration graph (sensitivity) and the distribution of the data in each concentration level. In the example shown in Fig. 7b, the sensor output at the working point $t = 6.3$ s showed more sensitivity and less variability than at the nominal working point ($t = 25$ s). This resulted in a lower LOD estimate ($x_D = 1.82$ versus 4.77 ppm) and smaller residual error of the calibration model ($s_{y,x} =$ 0.55 versus 1.29 ppm). The reliability of the LOD estimate was also higher at $t = 6.3$ s because the LOD assumptions were satisfied, whereas at $t = 25$ s the homoscedasticity assumption was violated. The Levene's test determined that there was more variability at 2.2 ppm than at 4.4 ppm. Predictions with unbiased, normally distributed and homoscedastic residuals are the basis of LOD theory. The LOD values obtained when any of these assumption is violated are not trustworthy. This highlights the importance of assessing the validity of the underlying LOD assumptions before reporting a LOD value.

To find the optimum working point for each sensor, the procedure described in Section 2.6 was applied. The measurements corresponding to one experimental day (d = 11) were excluded from the dataset due to abnormal readings caused by an external factor. The MFC controlling the CO injection channel reported

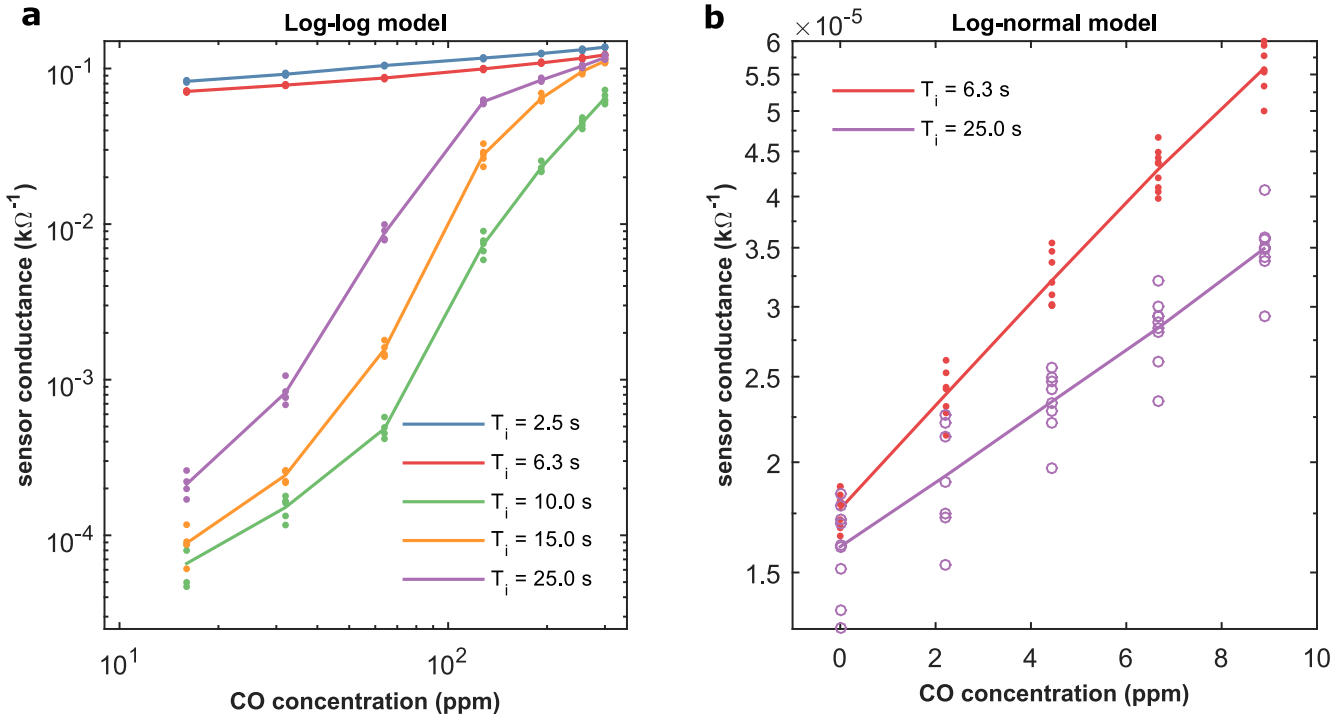


Fig. 7. Dependence between sensor conductance and CO concentration at several detection points of the heating cycle. (a) log-log model in the range 0–300 ppm, (b) log-normal model in the range 0–9 ppm. In both (a) and (b), each point is the average of ten replicate measurements each one having a random humidity value in the range 30–70% r.h.

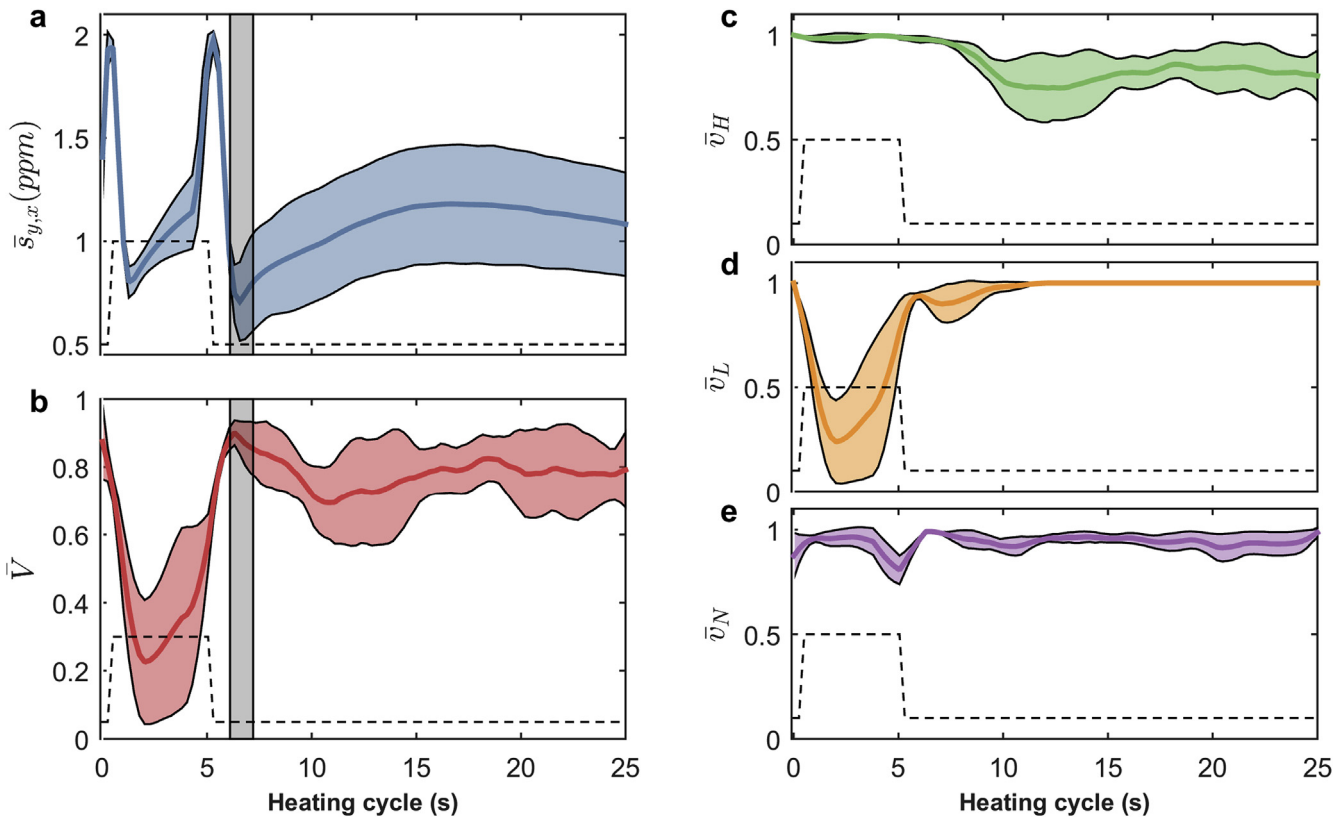


Fig. 8. Main parameters for the optimization of the working point in the log-normal calibration model (Eq. (3)) in the range 0–9 ppm, as a function of the heating cycle. (a) Standard error of regression; (b) validity of x_D; (c–e) validity of the homoscedasticity, linearity and normality assumptions, respectively. The values were first averaged across calibration days and the dispersion of the average values among sensor units is represented as an area plot. The solid line surrounded by a shaded area indicate the mean ± 1 mean absolute deviation (MAD). MAD $= \frac{1}{n}\sum_{i=1}^{n}|x_i - mean(x)|$. The vertical gray shaded area in subplots a-b indicate the optimum working range for the sensor family.

unusually low flow rates, probably due to an accidental manipulation of the manometer that controls the injection of CO into the gas chamber was accidentally manipulated. Similarly, one of the seven MOX sensors (sensor #2) was not considered for the analysis because its response was clearly different from the other six sensors. This sensor, which probably belongs to a different fabrication batch, was acting as an outlier for the statistical analysis. As a result, the following analysis was performed using data extracted from the remaining twelve days and six sensors. Fig. 8 shows a summary of the main parameters used during the optimization process, for the entire sensor family. The values were averaged across experimental days -which can be understood as repetition of the calibration procedure-to produce more robust estimates. The first thing we can see in Fig. 8 is that the temperature transitions ($t = 0$ s and $t = 5$ s) are not good detection areas because s_{yx} was quite high, which means that the calibration model was not very accurate in this area. High heating temperatures ($t \in [0,\ 5]$ s) produced better calibration models (values of s_{yx} were moderate) but the obtained models rarely passed the linearity test. It should be recalled that values of V, v_H, v_N or v_L close to 1 indicate that the given constraint was valid in 100% of the repetitions (i.e. experimental days). The logarithmic transformation was more effective in linearizing the sensor response at low heating temperatures ($t \in [5,\ 25]$ s). The normality assumption was satisfied in both high and low heating temperatures. The data became slightly heteroscedastic in the second half of the heating cycle ($t \in [10,\ 25]$ s). The optimum working points for the six sensors were located at $t = \{6.5,\ 7.6,\ 6.8,\ 6.6,\ 6.1$ and $7.1\}$ s, respectively. The optimum working point was slightly different for each sensor but an optimum range consistent for the sensor family was found at the beginning of the low heating temperatures ($t \in [6.1,\ 7.6]$ s). Within this narrow range, minimum s_{yx} values and maximum values of V were obtained. This variability in optimum working points may be related to device tolerances regarding temperature at the sensing surface. Due to the mini-bead technology of FIS for this sensor series, we can have small tolerances both in thermal dynamics and steady temperature values at low power cycle. These small variations on the surface temperature, then will lead to variations in the chemical dynamics of the reactions taking place and ultimately instability of the optimum working point across devices.

Using the optimum working point of each sensor s, calibration models were computed for every experimental day d. The LOD $x_D(s,d)$ and the validity of the LOD $V(s,d)$ of these models were averaged across experimental days to produce more robust estimates (Eqs. (27) and (28)). A comparison between $\bar{x}_D(s)$ and $\overline{V}(s)$ and the equivalent values obtained at the nominal working point is shown in Fig. 9. Compared to the nominal detection point, the optimized models improved both the mean LOD (2.3 versus 4.0 ppm) and the mean validity (95% versus 80%). Because the optimum LOD value is at the low end of the concentration range tested in this work, to fully confirm the results more tests in a reduced range centered around the LOD (e.g. 0–4 ppm) might be required.

If the LOD value is going to be taken as reference for future operation, it is key to check the stability of this parameter. In other words, it is important to check if the calibration model remains stable. The prediction error may increase due to a shift of the data, a change of sensitivity or an increase in the variance. In any of these

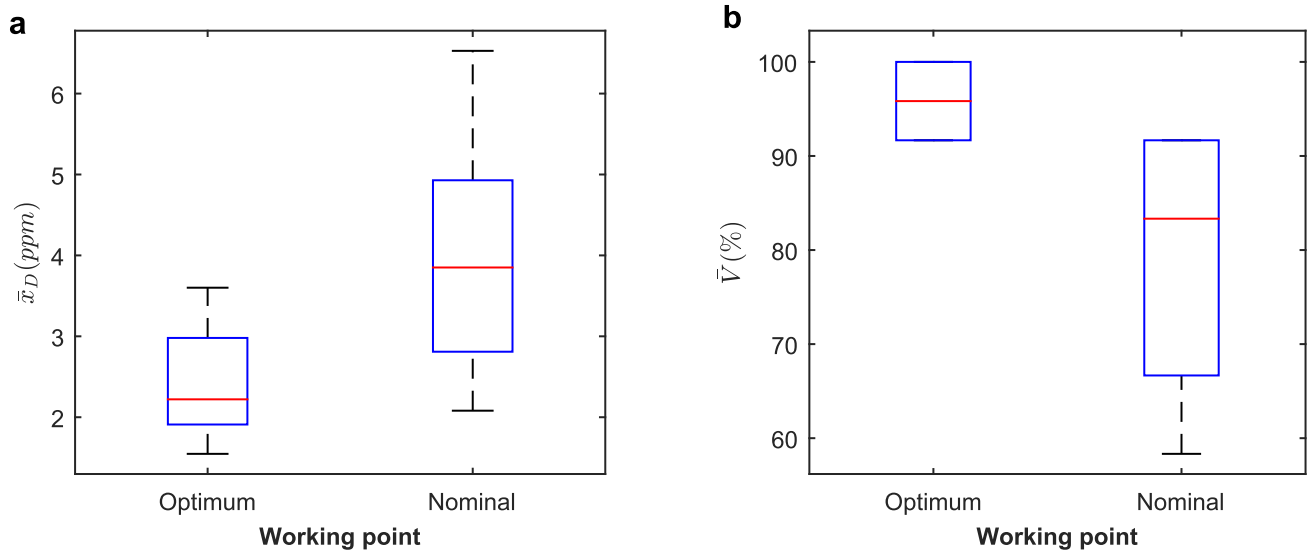


Fig. 9. Comparison between the optimized working point and the nominal working point in terms of (a) LOD and (b) its validity. The values were first averaged across calibration days and the dispersion of the average values among sensor units is represented as a box plot. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers.

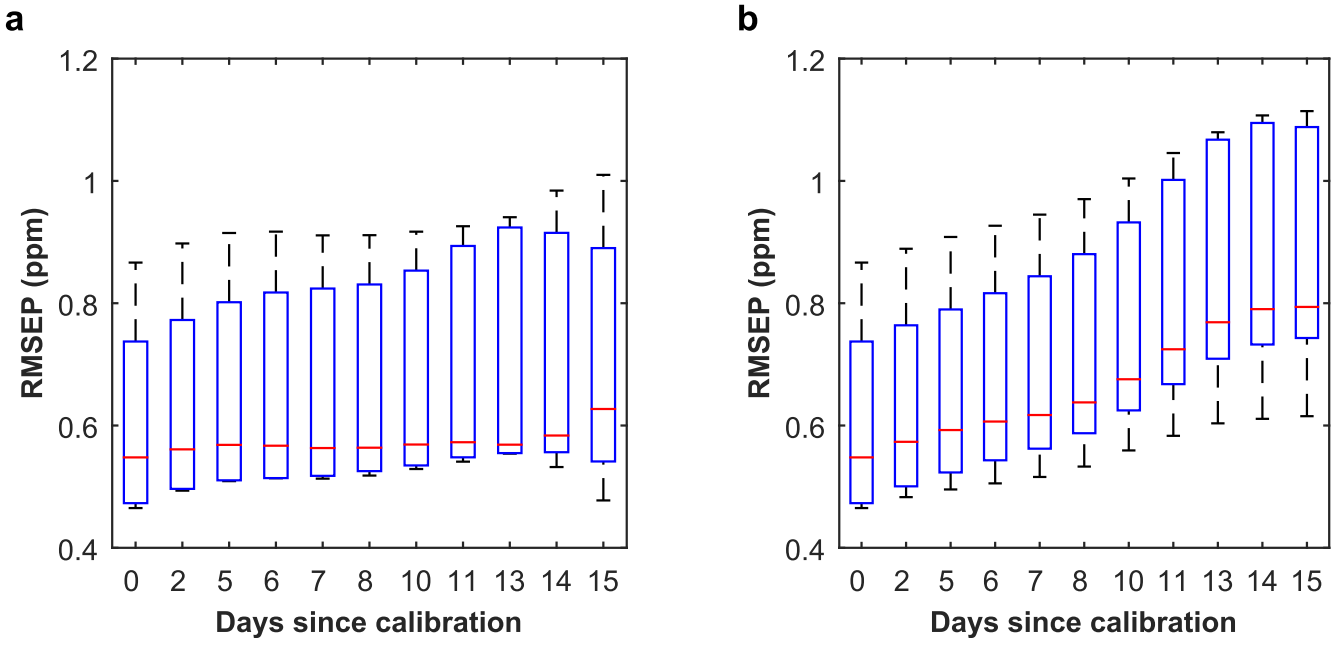


Fig. 10. Temporal evolution of the RMSEP of the optimized calibration models in the range 0–9 ppm. (a) Baseline correction; (b) No baseline correction. The dispersion of the values among sensor units is represented as a box plot. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. The x-axis indicates the elapsed time between the calibration day d_c and the validation day d_v. Each x value combines those groups of days in which $\{d_v - d_c = x\}$. $x = 0$ represents the RMSE in calibration.

situations, the reported LOD will not represent the detection capabilities of the sensor beyond the calibration day. The temporal evolution of the mean RMSEP of the optimized models is presented in Fig. 10. In the left graphic, we can see that the mean RMSEP slightly increased as the elapsed time between calibration and validation increased. In 15 days, the average RMSEP increased 0.09 ppm (0.61 versus 0.70 ppm), which represents a 3.9% of the mean LOD value (2.3 ppm). This means that the LOD value can be trusted in the studied temporal period provided that baseline correction was performed (i.e. dividing g_s by g_{air} in Eq. (3)). This baseline subtraction method is based on the measurements of a blank sample, which might not be easy to obtain in a real scenario. Blank samples could be generated either by exposing the sensor system to a reference clean air or by directly measuring the target atmosphere when the target gas is not present. The first approach increases the complexity of the system by adding a container to store the reference gas and the extra fluidic components. The second method assumes that at certain time the gas sensor will not be exposed to the target gas or at least not to concentrations near the limit of detection. For example, background concentrations of 150 ppb can be sometimes found in “clean” air. Since this residual concentration is one order of magnitude below the estimated limit

of detection, the baseline correction method will still work.

If the baseline is not corrected, the stability of the models is degraded (Fig. 10b). We can see that the mean RMSEP monotonically increased as the elapsed time between calibration and validation increased. In 15 days, the average RMSEP increased 0.32 ppm (0.64 versus 0.96 ppm), which represents a 13.9% of the mean LOD value (2.3 ppm). This may put into question the reliability of the LOD in a 2-week period. Further investigation revealed that the sources of error were small shifts in the measurements from one day to another, probably due to the instability of the sensor response at blank measurements, the open-loop control of the sensor temperature and the intentional changes in the levels of interference (humidity). The sensitivity and the variance of the measurements were not significantly affected. These findings greatly encourage baseline correction whenever possible.

4. Conclusions

We have proposed a methodology to overcome the challenges associated to LOD estimation in non-linear sensors, in this case MOX sensors aiming at CO detection in the presence of uncontrolled humidity levels. A main contribution of the paper is the selection of the sensor measurement point that should show small prediction error and simultaneously fulfilled the statistical requirements imposed by the IUPAC for linear univariate LOD estimation. The calibration models with logarithmic transformation at the input in a reduced concentration range (0–9 ppm) satisfied the linearity assumption at low temperatures of the heating cycle. The working points that yielded the most accurate calibration models and highest validity ratios of the LOD assumptions were all found at the beginning of the low level of the heating cycle ($t \in [6.1,\ 7.6]$ s). Average LOD values of 2.3 ppm and average validity ratios of 95% were found in the optimized models. The nominal working point suggested by the manufacturer yielded higher LOD values (4.0 ppm) and satisfied the LOD assumptions only in 80% of the repetitions.

The calibration models were stable in a scenario of varying levels of interferences, if baseline correction was performed. In this case, a maximum RMSEP increase of 0.09 ppm was observed (3.9% of the mean LOD value). This suggests that the LOD obtained in calibration is representative of the detection capabilities of the analytical system at least 2 weeks after calibration. Nonetheless, longer experiments might be required to confirm these results. If the changes in baseline were not corrected, the average RMSEP increased 0.32 ppm in 15 days, which represents a 13.9% of the mean LOD value. This might question the reliability of the LOD estimates in the studied time period.

This study demonstrated that linearized MOX sensors calibrated with optimized univariate features and periodic baseline adjustments can be used for gas sensing applications requiring a detection threshold. However, we should distinguish between two scenarios: (i) the LOD required by the application is much higher than the estimated LOD of a given sensor; (ii) the LOD required by the application is comparable to the sensor's LOD. Whereas in the first case the sensor will definitely meet the application requirements, in the second case the use of MOX sensors is questionable. It should be noted that MOX sensors are characterized by their intrinsic instability and, therefore, taking the LOD as a hard threshold in the second scenario might result in more false positives or false negatives than it was expected. This advice is especially important in critical applications in which a false positive or a false negative could result in legal consequences. Due to the high tolerance in sensitivity between MOX sensor units, the values reported in this work are specific to the sensor units used in this work. If the same experiments were repeated using different units the results might slightly change.

The main limitation of the current proposal is that the LOD must lie within the range of linear behavior. The current approach will fail if the LOD is larger than this limit. Additionally, it must be remarked that the optimum working temperature and the reported LOD values are specific to the scenario of carbon monoxide as target gas and humidity as the main interference. If the sensors were exposed to a complex gas mixture, the LOD values will probably increase. The methodology proposed in this paper could be applied to find a better working temperature for the given gas mixture.

Acknowledgements

This work was partially funded by the Spanish MINECO program, under grants TEC2014-59229-R (SIGVOL), PCIN-2013-195 (SENSIBLE) and BES-2015-071698 (SEVERO-OCHOA). The Signal and Information Processing for Sensor Systems group is a consolidated Grup de Recerca de la Generalitat de Catalunya and has support from the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya (expedient 2014-SGR-1445). This work has received support from the Comissionat per a Universitats i Recerca del DIUE de la Generalitat de Catalunya and the European Social Fund. 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.

References

[1] EPA, Criteria air pollutants, Am. Child. Environ. (2015) 1–22. http://www.epa.gov/criteria-air-pollutants.
[2] P.K. Sekhar, E.L. Brosha, R. Mukundan, W. Li, M.A. Nelson, P. Palanisamy, F.H. Garzon, Application of commercial automotive sensor manufacturing methods for NOx/NH3 mixed potential sensors for on-board emissions control, Sensor. Actuator. B Chem. 144 (2010) 112–119, https://doi.org/10.1016/j.snb.2009.10.045.
[3] E. Beard, R. West, Pilot study of the use of personal carbon monoxide monitoring to achieve radical smoking reduction, J. Smok. Cessat. 7 (2012) 12–17.
[4] L.A. Currie, Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995), Pure Appl. Chem. 67 (1995) 1699–1723.
[5] A. Burresi, A. Fort, S. Rocchi, B. Serrano, N. Ulivieri, V. Vignoli, Dynamic CO recognition in presence of interfering gases by using one MOX sensor and a selected temperature profile, Sensor. Actuator. B Chem. 106 (2005) 40–43.
[6] C. Bur, M. Bastuck, D. Puglisi, A. Schuetze, A.L. Spetz, M. Andersson, Discrimination and quantification of volatile organic compounds in the ppb-range with gas sensitive SiC-FETs using multivariate statistics, Sensor. Actuator. B Chem. 214 (2015) 225–233.
[7] S. Udina, M. Carmona, G. Carles, J. Santander, L. Fonseca, S. Marco, A micromachined thermoelectric sensor for natural gas analysis: thermal model and experimental results, Sensor. Actuator. B Chem. 134 (2008) 551–558.
[8] J.W. Gardner, P.K. Guha, F. Udrea, J.A. Covington, CMOS interfacing for integrated gas Sensors : a review, IEEE Sens. J. 10 (2010), https://doi.org/10.1109/JSEN.2010.2046409.
[9] A.C. Romain, J. Nicolas, Long term stability of metal oxide-based gas sensors for e-nose environmental applications: an overview, Sensor. Actuator. B Chem. 146 (2010) 502–506, https://doi.org/10.1016/j.snb.2009.12.027.
[10] M. Padilla, J. Fonollosa, S. Marco, Improving the Robustness of Odour Sensing Systems by Multivariate Signal Processing, 2013. Hum. Olfactory Displays Interfaces Odor Sens. Present. Inf. Sci. Ref. IGI-Global.
[11] L. Valentini, I. Armentano, J.M. Kenny, C. Cantalini, L. Lozzi, S. Santucci, Sensors for sub-ppm NO 2 gas detection based on carbon nanotube thin films, Appl. Phys. Lett. 82 (2003) 961–963.
[12] J. Nicolas, A.-C. Romain, Establishing the limit of detection and the resolution limits of odorous sources in the environment for an array of metal oxide gas sensors, Sensor. Actuator. B Chem. 99 (2004) 384–392, https://doi.org/10.1016/j.snb.2003.11.036.
[13] J. Fonollosa, A. Vergara, R. Huerta, S. Marco, Estimation of the limit of detection using information theory measures, Anal. Chim. Acta 810 (2013) 2013–2015, https://doi.org/10.1016/j.aca.2013.10.030.
[14] A.M.Z.B. Erna, S.T.T. Rowell, W.I.E.S.C. Ynkar, Comparison of metal oxide-based electronic nose and mass spectrometry-based electronic nose for the prediction of red wine spoilage, J. Agric. Food Chem. 56 (2008) 3238–3244.
[15] R. Boqué, M.S. Larrechi, F.X. Rius, Multivariate detection limits with fixed probabilities of error, Chemometr. Intell. Lab. Syst. 45 (1999) 397–408, https://doi.org/10.1016/S0169-7439(98)00195-6.

[16] C. Borrego, A.M. Costa, J. Ginja, M. Amorim, M. Coutinho, K. Karatzas, T. Sioumis, N. Katsifarakis, K. Konstantinidis, S. De Vito, E. Esposito, P. Smith, N. André, P. Gérard, L.A. Francis, N. Castell, P. Schneider, M. Viana, M.C. Minguillón, W. Reimringer, R.P. Otjes, O. v Sicard, R. Pohle, B. Elen, D. Suriano, V. Pfister, M. Prato, S. Dipinto, M. Penza, Assessment of air quality microsensors versus reference methods: the EuNetAir joint exercise, Atmos. Environ. 147 (2016), https://doi.org/10.1016/j.atmosenv.2016.09.050.
[17] M. Kuske, M. Padilla, A.C. Romain, J. Nicolas, R. Rubio, S. Marco, Detection of diverse mould species growing on building materials by gas sensor arrays and pattern recognition, Sensor. Actuator. B Chem. 119 (2006) 33–40, https://doi.org/10.1016/j.snb.2005.02.059.
[18] T. Hübert, L. Boon-Brett, V. Palmisano, M.A. Bader, Developments in gas sensor technology for hydrogen safety, Int. J. Hydrogen Energy 39 (2014) 20474–20483.
[19] A. Loutfi, S. Coradeschi, G.K. Mani, P. Shankar, J.B.B. Rayappan, Electronic noses for food quality: a review, J. Food Eng. 144 (2015) 103–111, https://doi.org/10.1016/j.jfoodeng.2014.07.019.
[20] M. Righettoni, A. Amann, S.E. Pratsinis, Breath analysis by nanostructured metal oxides as chemo-resistive gas sensors, Mater. Today 18 (2015) 163–171.
[21] G. Neri, First fifty years of chemoresistive gas sensors, Chemosensors 3 (2015) 1–20.
[22] C. Di Natale, R. Paolesse, E. Martinelli, R. Capuano, Solid-state gas sensors for breath analysis: a review, Anal. Chim. Acta 824 (2014) 1–17.
[23] F.U.S.A. Inc, TGS203-Carbon Monoxide Sensor Specification & Technical Information for Carbon Monoxide Sensors, 1999 pdf.dzsc.com/88888/2008617161056961.pdf.
[24] P. Van Geloven, M. Honore, J. Roggen, S. Leppavuori, T. Rantala, The influence of relative humidity on the response of tin oxide gas sensors to carbon monoxide, Sensor. Actuator. B Chem. 4 (1991) 185–188.
[25] G. Korotcenkov, B.K. Cho, Instability of metal oxide-based conductometric gas sensors and approaches to stability improvement (short survey), Sensor. Actuator. B Chem. 156 (2011) 527–538, https://doi.org/10.1016/j.snb.2011.02.024.
[26] G. Korotcenkov, B.K. Cho, Engineering approaches to improvement of conductometric gas sensor parameters. Part 2: decrease of dissipated (consumable) power and improvement stability and reliability, Sensor. Actuator. B Chem. 198 (2014) 316–341.
[27] A.C. Olivieri, N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas, H. Mark, Uncertainty estimation and figures of merit for multivariate calibration (IUPAC Technical Report), Pure Appl. Chem. 78 (2006) 633–661, https://doi.org/10.1351/pac200678030633.
[28] A.M. de la Peña, A. Espinosa-Mansilla, M.I.A. Valenzuela, H.C. Goicoechea, A.C. Olivieri, Comparative study of net analyte signal-based methods and partial least squares for the simultaneous determination of amoxycillin and clavulanic acid by stopped-flow kinetic analysis, Anal. Chim. Acta 463 (2002) 75–88.
[29] N. Barsan, U. Weimar, Understanding the fundamental principles of metal oxide based gas sensors; the example of CO sensing with SnO2 sensors in the presence of humidity, J. Phys. Condens. Matter 15 (2003) R813.
[30] F.I.S. Inc, FIS GAS SENSOR SB-500-12, 2017. http://www.fisinc.co.jp/en/common/pdf/SB50012E%7B_%7DP.pdf.
[31] T. Conrad, P. Hiry, A. Schutze, PuMaH-a temperature control and resistance read-out system for microstructured gas sensors based on PWM signals, in: Sensors Conference, 2005 IEEE, 2005, p. 8.
[32] I. Lavagnini, F. Magno, A statistical overview on univariate calibration, inverse regression, and detection limits: application to gas chromatography/mass spectrometry technique, Mass Spectrom. Rev. 26 (2007) 1–18.
[33] J. Mocák, I. Janiga, E. Rábarov Á, Evaluation of IUPAC limit of detection and iso minimum detectable value electrochemical determination of lead, Nov. Biotechnol 9 (2009) 91–100.
[34] G. Nelson, Gas mixtures: Preparation and Control, CRC Press, 1992.
[35] P.K. Clifford, D.T. Tuma, Characteristics of semiconductor gas sensors II. Transient response to temperature change, Sensor. Actuator. 3 (1982) 255–281.
[36] K. Danzer, L.A. Currie, Guidelines for calibration in analytical chemistry. Part I. Fundamentals and single component calibration (IUPAC Recommendations 1998), Pure Appl. Chem. 70 (1998) 993–1014.
[37] L.A. Currie, Detection: International update, and some emerging di-lemmas involving calibration, the blank, and multiple detection decisions, Chemometr. Intell. Lab. Syst. 37 (1997) 151–181.
[38] H. Levene, Contributions to Probability and Statistics, Essays Honor Harold Hotelling, Stanford University Press, 1960, pp. 278–292.
[39] S.S. Shapiro, M.B. Wilk, An analysis of variance test for normality (complete samples), Biometrika 52 (1965) 591–611.
[40] N.M. Razali, Y.B. Wah, others, Power comparisons of shapiro-wilk, Kolmogorov-smirnov, lilliefors and anderson-darling tests, J. Stat. Model. Anal 2 (2011) 21–33.
[41] G.E.P. Box, W.G. Hunter, J.S. Hunter, Statistics for Experimenters: an Introduction to Design, Data Analysis, and Model Building, JSTOR, 1978.
[42] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. (1979) 65–70.