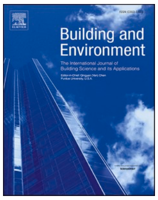
Contents lists available at ScienceDirect

Building and Environment

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



Evaluation of low-cost formaldehyde sensors calibration

Maria Justo Alonso [a,*], Henrik Madsen [b], Peng Liu [c], Rikke Bramming Jørgensen [d], Thomas Berg Jørgensen [a], Even Johan Christiansen [e], Olav Aleksander Myrvang [e], Diane Bastien [f], Hans Martin Mathisen [a]

[a] *Department of Energy and Process Engineering, NTNU, Kolbjørn Hejes v 1B, Trondheim, Norway*
[b] *Department of Applied Mathematics and Computer Science, DTU, Asmussens Allé, Building 303B, Kgs. Lyngby, Denmark*
[c] *Department: Architecture, Materials and Structures SINTEF Community, Høgskoleringen 13, Trondheim, Norway*
[d] *Department of Industrial Economics and Technology Management, NTNU, Sem Sælands vei 5, Trondheim, Norway*
[e] *Department of Electronic Systems, NTNU, O.S. Bragstads plass 2B, Trondheim, Norway*
[f] *Department of Technology and Innovation, University of Southern Denmark, Campusvej 55, Odense, Denmark*

ARTICLE INFO

Keywords:
IAQ
Low-cost sensors
CO_2
Air temperature
RH
Formaldehyde
TVOC

ABSTRACT

Low-cost sensors (LCS) are becoming ubiquitous in the market; however, calibration is needed before reliable use. An evaluation of the calibration of eight identical pre-calibrated formaldehyde LCS is presented here. The LCS and a reference instrument were exposed to a pollutant source(s) for the calibration measurements. After one year, some tests were repeated to check the drift and stability of calibration.

This paper presents methodologies for calibration using data with significant autocorrelations. Autocorrelation in sensor measurements might be present when performing a frequent sampling. To obtain reliable results, sensor calibration methodologies must consider autocorrelation or serial correlation between subsequent measurements. Experimental design can be used to reduce the risk of highly autocorrelated measurement.

Ordinary Least Squares Estimations should not be used when measurements are autocorrelated, as their central assumption is that the residuals are independent and identically distributed. Two alternative methods considering autocorrelation using a first-order Markov scaling are proposed: Maximum Likelihood and Restricted Maximum Likelihood Estimation (REML). REML has better compensations for the estimated parameters and the scaling parameters. Akaike information criterion was used to select the most significant parameters resulting in formaldehyde and temperature.

The results were presented for only one of the eight sensors. According to EPA's recommendations, the tested formaldehyde LCSs were Tier III, supplementary monitoring. The LCS over-and under-estimated the values obtained by the reference sensor, but they presented very similar dynamic responses, indicating that LCS could be used to detect concentration changes after calibration.

1. Introduction

Tightening building envelopes and using demand-controlled ventilation are commercial buildings' most applied energy-saving strategies [1]. When reducing supply airflow or infiltration rates, pollutants that otherwise would be ventilated away may be present at higher and even harmful concentrations [2]. Without correct implementation, retrofits targeting energy efficiency can adversely affect health due to the lower air change rates [3,4].

Formaldehyde is one compound widely found in household materials [5]. It is also produced in cooking, wood burning, other domestic activities [5], and waterproofing coatings [6]. However, it is associated with health risks such as mucous irritation [7] and is carcinogenic (group 1) to humans, according to the International Agency of Research on Cancer (IARC) [8]. Wolkoff [9] concluded that formaldehyde and benzene are generally reported as sensory irritation even before being smelled.

Norwegian indoor air quality should meet the air quality criteria based on health impacts defined in the building codes [10], the occupational health codes [11], and the public health legislation [12]. However, several defined pollutants are rarely measured due to the high cost of reliable sensors. Traditionally, air pollutants were measured with complex, expensive, and massive equipment at fixed locations. Thus,

[*] Corresponding author.
E-mail address: maria.j.alonso@ntnu.no (M. Justo Alonso).

https://doi.org/10.1016/j.buildenv.2022.109380
Received 29 April 2022; Received in revised form 6 July 2022; Accepted 6 July 2022
Available online 8 July 2022

Nomenclature

CO_2	Carbon Dioxide
CH_2O	Formaldehyde
CV	Coefficient of variation
HVAC	Heating, ventilation, and air-conditioning
IAQ	Indoor Air Quality
LCS	Low-cost sensor
MAE	Mean absolute error
ML	Maximum likelihood
MV	Measured value
NMB	Mean Normalized Bias
OLS	Ordinary Least Squares
PCC	Pearson Correlation Coefficient
PM	Particulate matter
RH	Relative humidity
REML	Residual maximum likelihood estimates
RMSE	Root mean squared error
TVOC	Total volatile organic compound

manufacturers have developed low-cost air quality sensors to measure air parameters and airborne pollutants. Technological advances in metal oxide semiconductors, for the detection of gaseous compounds [13] allowed the development of sensors with a much lower cost than the certified reference instruments. Typically, they are less accurate and suffer from cross-sensitivities with other pollutants. They are usually smaller sensors, measure constantly, provide real-time monitoring, and are easily deployed. They can send the information to IoT servers or record data in loggers. Thus, the availability of such sensors will likely continue to grow [14].

Current ambient air monitoring could be improved if LCS could produce reliable data under typical ambient conditions. However, preliminary tests [15–20] suggest poor to uncertain reliability. Some do not perform well under typical ambient conditions or do not correlate with data from regulatory measurement equipment [20]. A recent analysis of 112 studies on LCS [21] concluded that only a few studies followed US EPA guidelines [22] to examine performance. LCSs often suffer from errors due to internal causes (such as cross-sensitivities with other pollutants, drift, bias …) and external causes (such as temperature and relative humidity) [23]. Therefore, it is urgent to characterize the actual performance of LCS and educate the users about the potential and limitations of these sensors [14]. Understanding the sensors' limitations is needed to interpret the data output and its weakness [24]. A poor understanding of flaws (problems with experimental design) and limitations (factors that constrain the applicability of study findings) can lead to undesirable outcomes such as alarmistic behaviors. For some gaseous measurements, cross sensitivities to confounding compounds overpredict the measured pollutant [25] precisely as drift may [26].

With the universalization of the use of LCSs, better recommendations regarding placement should be delivered together with the sensor datasheets as the users know less about IAQ measurements. This knowledge is even more critical when using these sensors to control ventilation.

To compensate for their lower accuracy, dealing with the error is crucial no matter the accuracy required by the application of the sensor. Giordano et al. [27] revised the needs and challenges of achieving reliable data from particulate matter LCS. They summarized their knowledge on best practices in calibration considering data collection and model analysis, but they did not address the autocorrelation in the measurements.

1.1. Objectives

This study has two main objectives:

1-Testing the performance of eight formaldehyde sensors in comparison to laboratory-grade equipment. The evaluation uses measurements at the beginning of the sensors' lives and after one year.

2- Support that non-mathematician IAQ researchers can do a good calibration of LCS and that they are able to evaluate the results within the calibration range. The results of only one of these eight sensors are presented as the goal is to focus on the calibration procedure. For that, the article will address the following:

- The challenge of having frequent sampling yielding autocorrelated measurements and tests taken with very heterogeneous data collection periods
- Establish the best calibration estimation method that considers the autocorrelation of the calibration measurements and that the number of samples in the tests is not equal. The method must do a correct estimate of parameters and the uncertainties.

Contrarily to most of the existing articles evaluating LCS, this article focuses on the evaluation of the calibration process, which would also apply to other sensors when the data sampling results are autocorrelated. In this article, the experimental design is thoroughly described, discussed, and evaluated. The common application of R^2 evaluations to study correlations is confronted. It is mathematically wrong to use OLS when the residuals are not independent and identically distributed (iid). Thus, this article demonstrates an alternative methodology for this situation.

2. Methodology

In this study, eight identical Dart formaldehyde WZ-S LCS were calibrated using measurements in a laboratory's small chamber. The sensors were exposed to the same formaldehyde sources as laboratory-grade equipment. Some of the experiments were repeated after one year. The data obtained by low-cost and professional-grade sensors were compared to establish a model representing the sensor behavior and then estimate the residuals, i.e., the error in the model-based predictions. This article has created a procedure for estimating a weighting according to the autocorrelation based on the first-order Markov. Then a simple method using this weighting was created. To the authors' knowledge, the method presented here considering and weighting the autocorrelation using a first-order Markov scaling has not been used in the sensor calibration field.

2.1. Measurement equipment

2.1.1. Indoor air sensing stations and LSC sensors

Eight equal in-house mounted IAQ stations were assembled comprising LCSs to measure formaldehyde, TVOC, temperature, and RH. The LCSs were selected based on user-friendliness (these sensors had available information on the internet regarding mounting) and pre-calibrated from the factory (according to the producers, they should not need any pre-use calibration). Table 1 summarizes the LCS's model, type, and technical specifications. More information about the kit, the LCS not discussed in this article (commercial Sension LCS to measure particle matter SPS30, CO_2 SCD30) and their calibration can be found in Ref. [30].

The Dart Sensor WZ-S is a micro fuel cell formaldehyde sensor. In addition to formaldehyde, the other parameters were studied to see if they were confounding parameters. Air temperature and RH were measured with SHTC1 from Sensirion and TVOC with SGP30 from Sensirion. Sensors SGP30 and SHTC1 were integrated into the Arduino Shield SGP30_SHTC1 from Sensirion.

All the sensors were connected to an Arduino via a customized shield

Table 1
Technical specification LCSs data retrieved from, SVM30 [33], Dart WZ-S [35].

Sensor name	Parameter	Sensor type	Measurement range/size range	Accuracy collected from datasheets	Single unit price when bought in NOK
Dart Sensors WZ-S	Formaldehyde	MOS	0,03 - 2 ppm	≤0.001 ppm	148
Sensirion Arduino Shield SGP30_SHTC1[a]	TVOC Air Temperature Relative humidity	MOx CMOS	0–60′000 ppb −30 °C–100 °C 0%–100% RH	1 ppb or 6 ppb[b] ±0.3 °C ±3% RH	190

[a] This sensor uses SHTC1 for measuring T/RH and SGP30 for measuring TVOC. Sensirion does not recommend the use of this sensor for new designs anymore: https://www.sensirion.com/en/environmental-sensors/gas-sensors/multi-gas-humidity-temperature-module-svm30/.

[b] 1 ppb from 0 ppb to 2008 ppb, and 6 ppb from 2008 ppb to 11110 ppb.

card. The logged values were sent to a Raspberry Pi via USB cable. The Dart Sensors WZ-S were connected using a custom-written code. The complete code for Arduino and Raspberry Pi was available on GitHub. LCS collected data every 5 min, and measurements were converted to 30 min averages. The Arduino Shield SGP30_SHTC1 was connected using code from Adafruit [31] and Sensirion AG [32], available on GitHub. This sensor needs a pre-calibration file based on 12 h of calibration in the air. This pre-calibration is done because when the sensor is exposed (measuring or not) to conditions (RH and temperature) outside the recommendations, the RH signal may offset. After being in normal temperature and humidity, the sensor will slowly return to standard specifications [33] (removing the offset). The resulting calibration files to standard specifications were stored in the Raspberry Pi. This calibration to standard specifications was done previously to the one explained in the remaining of the article and should be done each time the sensors are used so that the following calibration makes sense. The data processing and further analysis were done with R software [34].

2.1.2. Reference monitoring equipment

Graywolf FM-801 was deployed as the reference instrument for formaldehyde measurements, and it was calibrated before the experiments. The Graywolf sensor uses photoelectric absorptiometry to read the sensor's absorbance change that formaldehyde induces. A small colorimetric sensor cartridge is used for passive diffusion sampling. Graywolf FM-801 measures the absorption change between each 30-min interval and then calculates the difference. The value reported by the unit represents the average of over 30 min. The sensor has a detection range from 25 μg/m^3 to 1230 μg/m^3 and an accuracy of ±10% for readings larger than 48 μg/m^3 [28]. The sensor suffers from cross-sensitivity to methanol, ethanol, isopropanol, carbon monoxide, phenol, acetaldehyde H_2, chloroform, limonene, styrene, acetaldehyde, ozone, H_2S, and $SO_{2,}$ among others [28]. Pegasor AQTM Indoor was deployed for measuring RH and temperature [29]. RH range: 0–100%; producer-reported accuracy ±1.5% within 0–90%. Temperature range: 40 - 80 °C; producer-reported accuracy ±0.2 °C in the range 0–40 °C.

2.2. The chamber setup

The calibration evaluation of the formaldehyde LCS was conducted in the 1.5 m^3 plexiglas mini environmental chamber in Trondheim, Norway showed in Fig. 1. The chamber is equipped with dedicated ventilation, heating, and humidification systems run as the tests required. The HVAC system consists of extract and supply fans to control the ventilation rate and a small computer fan for mixing the chamber's air, a radiator, and a humidifier.

The sensors were launched at least 2 h before introducing the air pollution source. All the ventilation supply was turned off at the beginning of each experiment, right before the start of pollutant generation and monitoring. The background concentrations of formaldehyde were negligible at the start of the experiments. The air exchange was reduced to infiltration, which was minimized by blocking the openings with duct tape. Each experiment was monitored with both the LCS and the reference sensor continuously. In some cases, the sensors measured at least 1 h after the pollutant source was stopped/removed.

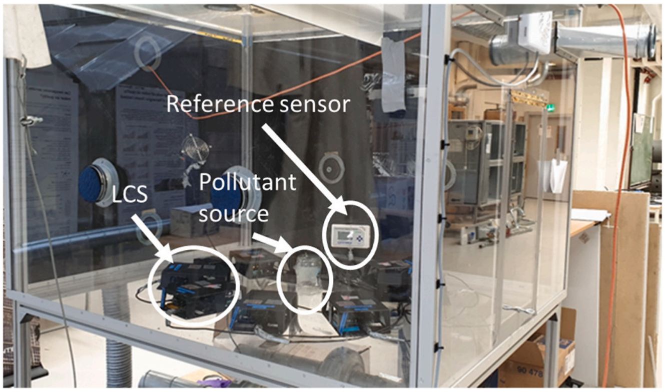


Fig. 1. Picture of the experimental setup with the formalin source in the center. The eight equal IAQ stations and the reference sensor in a circle equidistantly to the source.

The eight IAQ stations and the reference sensor stood in a circle around the pollutant source during the experiments, as shown in Fig. 1. For the experiments with chipboards, these were placed along the walls of the mini chamber.

2.3. Experiment description

The performance evaluation was done by comparing formaldehyde measurements with the eight equal LCSs and the reference sensor.

The calibration-tests details are summarized in Table 2. Test A was conducted under uncontrolled ambient temperature and RH in the chamber, thus, representing typical temperatures and RH in the laboratory in a Norwegian winter. In tests B, C, D, E, and F, a heater and a humidifier were run to control the temperatures and relative humidity. Tests with wood chipboards were repeated after one year, as Table 2 specifies, to study the repeatability of the results and drift of the equipment.

When wet, wood chipboards produce formaldehyde and TVOC at a higher range than under normal conditions. Formalin is a source of formaldehyde and methanol.

It was expected that in one year, the aging of the LCS would be negligible [27]; however, at the same time, problems with drift or problems with singular defective units would be identified. In-between calibrations, the sensors were used for routine measurements in schools. Effects of a differential exposure history during this period were considered by studying the recorded measurements. The recorded measurements were compared. In this case, the maximum measured concentrations did not differ by more than 9 %, and the effect of differential exposure history was assumed negligible.

2.4. Use of correlated data in calibrations

A calibration model is a regression model developed from the response of a sensor to known sources (customarily measured with a

Table 2
Description of calibration activities and resulting formaldehyde concentration reported as the highest 30-min average concentration with the reference sensor. TVOC is reported as the highest measured 5-min concentration by the 8 LCS. Conditions of temperature and relative humidity are shown as the average ± the standard deviation.

ID	Source	Test duration in minutes	Temperature and RH	Date	Activity description	Formaldehyde µg/m3	TVOC ppb
A	Formalin	150 min	20.2 ± 0.2 °C 28 ± 0.41%	Feb 2020	Beaker with (liquid 37%) formalin. Radiator and Humidifier off	336	1695
B	Wet wood chipboard	780 min	19.9 ± 0.1 °C 48.8 ± 13.7%	Feb 2020	1 wet board size 1m^2. Humidifier on	224	1802
C	Wet wood chipboard	180 min	21.1 ± 0.5 °C 26.8 ± 0.25%	March 2021	4 wet boards size 1m^2. Humidifier off	606	55
D	Wet wood chipboard	120 min	27.7 ± 0.8 °C 64.7 ± 9.7%	March 2021	3 wet boards size 1m^2. Radiator on, Humidifier on	451	105
E	Wet wood chipboard	90 min	23.5 ± 0.9 °C 72.6 ± 8.4%	March 2021	1 wet board size 1m^2 Radiator off, Humidifier on	41	214
F	Wet wood chipboard + wet glass wool insulation	180 min	21.2 ± 0.5 °C 35 ± 3.7%	March 2021	2 wet boards size 1m^2, 0.7m^2 wet insulation. Radiator and Humidifier off	22	278

reference sensor). To have a good calibration model is not so important to have a good fit of the measurements fed but to precisely predict/estimate new/unseen measurements within the calibration range.

Measurements have to be taken carefully to do a good calibration. For instance, if an additional measurement were taken only 1 s after the previous measurement, then the additional information of this new measurement would be limited. These two measurements are said to be serially correlated using statistical terminology, and such correlation in time for the same phenomena is often called autocorrelation.

Most often, standard OLS is used for analyzing data from calibration studies. OLS techniques assume that the measurements are independent and identically distributed (iid). Often this can be assured by an experimental design where the time distance between samples is large and constant.

However, the measurements might have been obtained in a more heterogeneous setting in some cases. An example could be that for some tests, several measurements might have been obtained during a few hours at that test, whereas sometimes, just a single measurement is conducted on other tests. Measurements within a single test can often be highly correlated in time. A high autocorrelation must be considered in a proper data analysis to obtain reliable conclusions.

The validity of using simple linear models or the effects of the experimental design are seldom discussed. Calibration quality is usually addressed as the coefficient of determination (R^2) between the evaluated and the reference instruments [21]. However, R^2 should only be used if the measurements are independent (not correlated) and evenly distributed. In the actual study, the lag-1 autocorrelation is 0.972, which is a high autocorrelation; thus, it has to be taken into account. This paper describes methodologies for conducting proper calibration analysis given such highly correlated data.

To understand the problem, let an extreme case be considered. It is assumed that the correlation between measurements taken on the same day was 1, whereas the correlation between measurements taken on two different days was 0. For simplicity, it will be assumed that 1000 measurements were obtained for one day, whereas for nine days, a single measurement was obtained for each day. Since the correlation was 1 within the same day, a single measurement contains all relevant information, and the 999 remaining measurements do not provide further information.

Using OLS, all measurements will have the same weight. The calibration line would be very close to the perfectly correlated observations on the day with 1000 measurements, and the influence from the nine other measurements will be minor. The resulting calibration line will thus be highly biased towards the line through the 1000 measurements. Using OLS, it is (wrongly) assumed that there are N = 1009 observations, and consequently, the uncertainty of the parameter estimates will be very low (proportionally to 1/N = 1/1009). Finally, R^2 will be very high (close to 1), and the Residual Standard Error will be very low, which does not reflect reality.

A proper weighting of the measurements is needed to do the correct analysis. Since the correlation was one within the same day, the 1000 measurements should effectively count only as one measurement. Since the correlation was assumed to be zero between days, the best calibration can be found using a weighting where the 1000 measurements were treated as a single measurement, and consequently, there are effectively only N = 10 observations. The uncertainty of the parameter estimates of the best calibration will be much higher (proportional to 1/10), and the Residual Standard Error will be much higher, but it would reflect the true calibration error needed if the calibrated sensor is expected to be used on a day in future.

This paper describes methods for calibration which take the actual autocorrelation into account and provide proper calibration curves and uncertainty estimates no matter how the measurements are taken. To the authors' knowledge, the suggested approach for handling autocorrelated measurements using a first-order Markov scaling has not been used to calibrate LCS.

2.4.1. Regression models used for calibration

The classical regression model is a statical relationship between a dependent variable Y$_t$ and *p* independent variables X$_{1t}$, X$_{2t}$, …, X$_{pt}$. For these sensor calibration experiments, the observations occur successively in time; therefore, an index *t* is introduced to denote the measurements at time *t*. In the calibration, the p independent (or explanatory) variables imply that adjusting for experimental conditions, like temperature and moisture, is possible.

A nonlinear function for the calibration curve can be used if a linear calibration curve does not fit the experimental data well for some calibration experiments. Thus, the general regression model will be introduced

$$Y_t = f(X_t, t; \theta) + \varepsilon_t \tag{1}$$

where $\theta = (\theta_1\theta_2, \ldots\theta_m)^T$ is a vector of the m unknown parameters, *f* is a known function of the $p + 1$ independent variables $X_t = (X_{1t}, X_{2t}, \ldots., X_{pt})^T$ and *t*.

The error term ε_t is assumed to be a random variable with a mean zero (E[ε_t] = 0), and the variance $Var[\varepsilon_t] = \sigma_t^2$, is assumed to depend on the time *t*. Furthermore, it is assumed that the residuals are correlated in time:

$$Cov\left[\varepsilon_{ti}, \varepsilon_{tj}\right] = \sigma^2 \Sigma_{ij} \tag{2}$$

where Σ_{ij} is a weight. In the following, it is assumed that the independent variable is known, i.e., $X_t = x_t$.

The central assumption in linear regression is that the sequence of error terms is a sequence of independent and identically distributed

(IID) random variables [37]. The above formulation contains a generalization that allows for varying uncertainty (variances) and a heterogeneous time sequence of autocorrelated error terms. Readers are referred to Madsen [37] for more details on how this is considered in the general nonlinear setting.

2.4.1.1. The general linear model. The calibration curve is often assumed to be linear, which allows using the general linear model or multiple linear regression model:

$$Y_t = X_t^T\theta + \varepsilon_t \tag{3}$$

where $X_t = (X_{1t}, X_{2t}, ..., X_{pt})^T$ is a known (non-random) vector and $\theta = (\theta_1, \theta_2, \ldots, \theta_m)^T$. The error term εt has zero mean and covariance $Cov[\varepsilon_{ti}, \varepsilon_{tj}] = \sigma^2\Sigma_{ij}$

N observations of the dependent and independent variables are assumed:

$$(Y_{t1}, x_{t1}), (Y_{t2}, x_{t2}), \ldots, (Y_{tN}, x_{tN}) \tag{4}$$

These observations occur successively in time, but observations at any given point in time, e.g. at non-equidistant time points, are allowed. This implies that very flexible experimental design and sampling times are allowed.

The total model for the N observations can be written

$$Y = x\theta + \varepsilon \tag{5}$$

where the design matrix x has dimension N × p. Following the definitions given in Eq. (3), $\mathrm{E}[\varepsilon_t] = 0$, and the covariance matrix for the residuals ε is $Var[\varepsilon] = \sigma^2\Sigma$ where $\Sigma = [\Sigma_{ij}]$.

Linear regressions assume that the residuals are independent and identically distributed (IID). This is described using the above formulation by putting $\Sigma = I$, being I the identity matrix leading to Ordinary Least Squares Estimation (OLS). However, this assumption is often violated in calibration problems and cannot be used.

2.4.1.2. Covariance and correlation structure. The covariance matrix Σ has to be specified appropriately to get reasonable estimates. Hence, both the variance and the correlation structure of the residuals ε_t have to be described.

The general formulation of the covariance matrix for the residuals is given by Eq. (2). By inspection of the data from these calibration experiments, it is seen that consecutive time residuals within a single test appear to be dependent. Similarly, it seems reasonable to assume that the variance is the same for all the residuals; thus, only the correlation structure must be specified.

In this case, it seems reasonable to assume that the correlation structure is an exponentially decaying function of the time distance between two observations, i.e.,

$$Cor\left[\varepsilon_{t_i}, \varepsilon_{t_j}\right] = \rho^{|t_i - t_j|} \tag{6}$$

where ρ is the correlation between two observations one-time unit apart. For hourly data, this is the hour-to-hour correlation. This corresponds to assuming a first-order Markov structure, or an Autoregressive first-order model, for the residuals [37]. Higher values of ρ coefficients denote a stronger correlation.

The assumption in linear regression is that Σ is known, but this is seldom the case in practice. In Ref. [37] a relaxation procedure is described, but for the above-mentioned problem, the likelihood function can be written assuming that the residuals are Gaussian.

2.4.1.3. Maximum likelihood estimates. The maximum likelihood method aims to describe the variation in the data by assuming a probability density and accounting for the autocorrelation. The consideration of autocorrelation will be advantageous in both forecasting and control.

As before equation (5) is considered for all N observationsand the residuals are assumed Gaussian, i.e., the measurements follow the Gaussian distribution

$$Y \in N\left(x\theta, \sigma^2\Sigma\right) \tag{7}$$

and contrarily to other authors, a first-order Markov correlation for the residuals Σ is assumed so that the correlation structure can be modeled and specified by

$$\Sigma_{ij} = Cor\left[\varepsilon_{t_i}, \varepsilon_{t_j}\right] = \rho^{|t_i - t_j|} \tag{8}$$

The above assumptions imply that the joint density for all observations, Y, is

$$f_y(y) = \frac{1}{\sqrt{\left(2\pi\sigma^2\right)^N det\Sigma}} \exp\left[-\frac{1}{2\sigma^2}(Y - x\widehat{\theta})^T\Sigma^{-1}(Y - x\widehat{\theta})\right] \tag{9}$$

Which implies that apart from a constant, the log-likelihood function for the unknown parameters is defined with equation (10):

$$\log L(\theta, \rho; Y) = -\frac{1}{2}\log \det(\Sigma) - \frac{N}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(Y - x\theta)^T\Sigma^{-1}(Y - x\theta) \tag{10}$$

The maximum likelihood estimates are found using numerical methods by maximizing equation (10). Estimates of the uncertainty of the parameter estimates are found using the observed Fisher Information Matrix; see Ref. [38] for details. In this article, the problem was implemented in R, and the GLS function was used in the NLME package. The Maximum Likelihood (ML) method and the residual maximum likelihood method (REML) were used. The ML method has the weakness that the variance estimates are biased, but this problem is handled by the REML method. The REML estimator corrects the estimated variance components for the degrees of freedom lost in estimating the fixed effect parameters; hence, the REML estimates the random effects more accurately. In practice, ML and REML give similar results and converge for large samples. Readers are referred to Ref. [38] for details.

2.5. Error determination

The idea of the error determination is to evaluate:

1) the accuracy that refers to how close the sensor reports to the true value or reference measurement,
2) the precision that responds to how consistently is the sensor reacting,
3) the bias that looks for systematic errors in reporting a value.

As already proven, it is impossible to demonstrate in this case that the residuals are independent and identically distributed. Therefore, using R^2 that quantifies the strength of the association (information about the goodness of a fit of a model) by equation (11) is not relevant for ML and REML. The ML or REML methods will "weight" the data, considering its information to consider it more reasonably. Thus, the fit will never be as good as when using the regression line considering all the data.

$$R^2 = 1 - \frac{\sum\left(C_{Lcs,i} - \widehat{C_{LCS,i}}\right)^2}{\sum\left(C_{Lcs,i} - \overline{C_{LCS,i}}\right)^2} \tag{11}$$

$$MAE = \frac{\sum_{i=1}^{N}\left|\widehat{C_{LCS,i}} - C_{ref,i}\right|}{N} \tag{12}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(\widehat{C_{LCS,i}} - C_{ref,i}\right)^2}{N}} \tag{13}$$

For precision metrics, mean absolute error (MAE; eq. (12)) and root mean squared error (RMSE; eq. (13)) will often be used. For equations (11)–(13) C_{ref} and C_{LCS} are formaldehyde concentrations measured by the reference monitor and LCS, respectively, $\widehat{C_{LCS,i}}$ is the predicted value and $\overline{C_{LCS,i}}$ the mean value.

MAE and RMSE calculate the average model prediction error, and their value can range from zero to values as high as the measured concentrations themselves. These parameters are useful to evaluate the fitted models' accuracy.

As these calibration models are built on different scales of formaldehyde concentrations, normalizing the accuracy metric is important so that models can be compared. Reporting normalized and absolute metrics is necessary when reporting errors. Normalizing performance metrics allows models to be appropriately compared between environments where concentration ranges are different [27]. The Coefficient of Variation (CV) and Mean Normalized Bias (MNB) are recommended guidelines by the United States Environmental Protection Agency (US EPA) for the evaluation of sensors [22] and thus, introduced in this article.

$$MNB = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(C_{LCS,i} - C_{ref,i}\right)}{C_{ref,i}} \tag{14}$$

$$CV = \frac{\sigma}{\mu} = \frac{\sqrt{\frac{\sum\left(C_{Lcs,i} - \overline{C_{LCS}}\right)^2}{N}}}{\overline{C_{LCS}}} \tag{15}$$

However, R^2, MAE, and RMSE do not account for autocorrelations. These performance indicators are created with the basic assumption that measurements are independent, which they are not in our case. They are nevertheless reported in this article to compare to other existing literature.

The Akaike information criterion (AIC) is a measure of the model's fit. AIC measures the quality of one model relative to the other as long as the models are constructed with the same estimate principle. For example, it will compare two different OLS models or two different REML models with different parameters, but not one OLS and one REML model. AIC provides a means for model parameter selection [39]. It is calculated using formula (16), where k is the number of estimated parameters in the model and$\widehat{L}$ the maximum value of the likelihood function for the model. As the Log-likelihood is a measure of model fit. The lower the number, the better the fit.

$$AIC = 2k - 2\ln\left(\widehat{L}\right) \tag{16}$$

For evaluation of the ML or the REML models, no error formulation is recommended as the evaluation would be very much dependent on the use of the sensor and other statistical parameters out of the scope of this article. In the case of ML and REML, a numerical method is often needed for finding the parameters which maximize the likelihood function.

2.6. Calibration procedure

The procedure for calibration followed the steps described below:

1 Check that all sensors react similarly to the exposure to the reference source. Before corrections can be studied, it should be controlled that all the units respond similarly to the same event [27]. Malings [40] defined intra-unit consistency when the variability is less than 20 % between equal units.
2 Log transformation of the data. To make it more normally distributed.
3 Study the calibration model most suitable to the available data, in this case, considering autocorrelated measurements and heterogeneous sampling lengths. The model sought was fitted using all the measured variables.
4 Repeat the fitting of the model only with the most significant variables chosen using the Akaike information criterion.
5 Evaluate the results based on the EPA suggested performance goals by application for MNB and CV according to Ref. [15]. This evaluation is done according to the values described in Table 3

3. Results and discussion

3.1. Raw measurements

Fig. 2 shows the raw measurements (out of the box) of all the eight sensors, and in black, the reference equipment (not for TVOC as a reference instrument was not available). For most of the measurements, all sensors react similarly. For temperature, the LCS overestimates the values compared to the reference sensor. For RH, the LCS over and underpredict the RH. For formaldehyde, LCS mostly overpredicts, being especially wrong in test A where the overprediction is especially high due to cross sensitivities with other gases. For TVOC, all LCS sensors predict similarly.

The sensors react similarly to the events to which they are exposed. The average difference among the LCS is 14%, 1%, 3%, and 18% for formaldehyde, temperature, RH, and TVOC, respectively. In the following, only the measurements and the models for calibrating sensor station S1 will be reported.

3.2. Calibration using formaldehyde, RH, temperature, and TVOC

The log transformation was done to have data that are more normally distributed.

The results for only one sensor are presented to exemplify the calibration methodology, but the results for all the other sensors were very similar.

Firstly, all the measured parameters were used to fit the calibration. Table 4 shows the parameters for OLS, ML, and REML after taking the logarithm of the data for sensor station 1.

OLS is just shown here for comparison to other literature, but given the performed experimental design (autocorrelation of the data and heterogenous sampling), its use is not recommended for the collected data. Given that not all the tests were equally long, with OLS, the longer ones will significantly affect the estimation. OLS is based on minimizing the sum of squares of the difference between the observed dependent variables in the dataset. Thus, when having autocorrelations, the fitting using OLS will be very good for the dataset fed-in. However, when using the calibration estimates to predict other "unseen" datasets, OLS estimates will perform poorly because they are "overfitted for the calibration dataset."

A practical example is given for clarity. If measurements of the size of a river are mostly taken after the snow-melting period, a huge river will

Table 3
EPA suggested performance goals by application for MNB and CV according to [15].

	MNB range	CV range
Tier I: Education and Information	$-0.5 < MNB < 0.5$	$CV < 0.5$ for all pollutants
Tier II: Hotspot Identification and Characterisation	$-0.3 < MNB < 0.3$	$CV < 0.3$ for all pollutants
Tier III: Supplemental Monitoring	$-0.2 < MNB < 0.2$	$CV < 0.2$ for all pollutants
Tier IV: Personal Exposure	$-0.3 < MNB < 0.3$	$CV < 0.3$ for all pollutants
Tier V: Regulatory Monitoring	$-0.07 < MNB < 0.07$	$CV < 0.07$ forO_3
	$-0.1 < MNB < 0.1$	$CV < 0.1$ for CO and $PM_{2.5}$
	$-0.15 < MNB < 0.15$	$CV < 0.15$ for NO_2

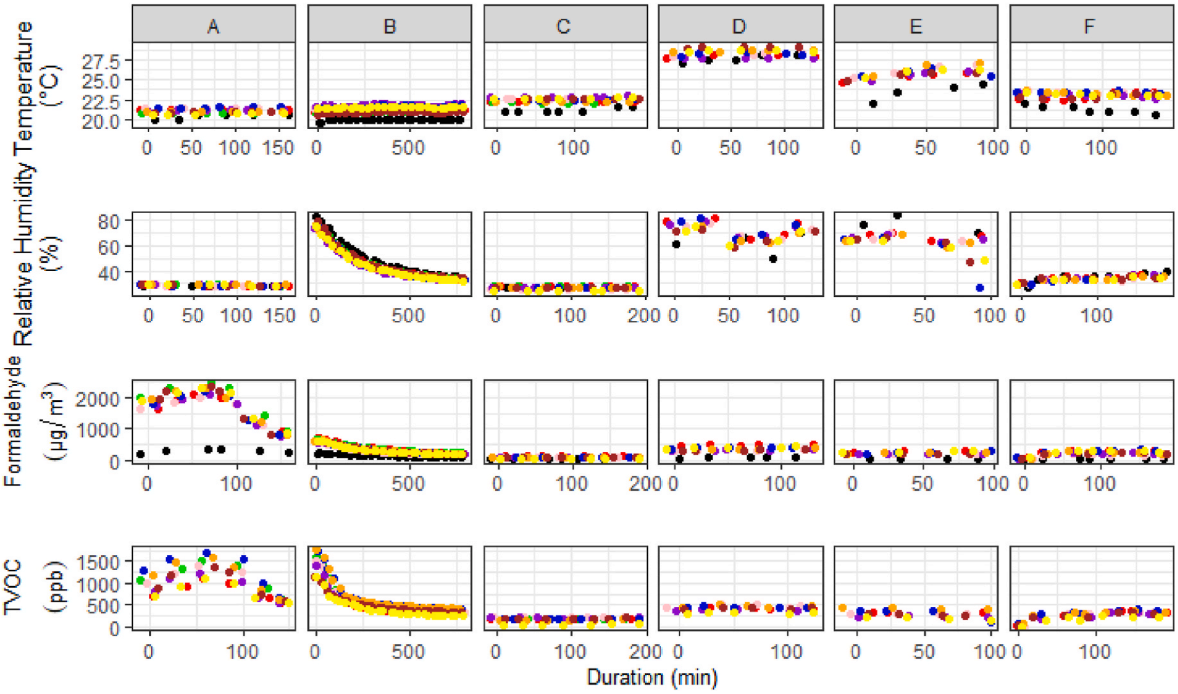


Fig. 2. Out-of-the-box response of the eight sensors to the six exposure tests. Every facet plots the results of each test, and the points are colored based on the sensor. Dots in black represent the reference instrument.

Table 4

Fitting of parameters using the different estimations. A p-value less than 0.05 was considered statistically significant. The middle part of the table shows the errors with typical formulations, and the lower part shows the evaluation according to EPA recommendations. OLS is just shown here as a comparison point, but its use is not recommended for calibration given the experimental design.

	OLS	p-value	ML	p-value	REML	p-value
Intercept	4.934	0.000	−7.83029	0	−7.958	0
FA	0.748	4.73E-05	0.176965	0.423	0.152	0.491
Temperature	−4.976	3.74E-06	6.721079	0	6.870	0
RH	0.661	0.015	−0.37359	0.078	−0.377	0.072
TVOC	2.576	0.190	2.227658	0.019	2.303	0.016
Residual standard error	0.211		0.48		0.63	
R-squared	0.7751					
AIC	−5.76		−128.9467		−127.0673	
RMSE	0.203		0.543		0.540	
MAE	0.146		0.509		0.512	
ρ	0.65		0.972		0.972	
CV	0.20	Tier II/IV	0.1486	Tier III	0.1482	Tier III
MNB	4.29e-18		0.112		0.099	

be predicted and may be perfectly predicted. However, its size will be overpredicted when the corrections are used to predict the same river in summer. Therefore, even though in Table 4, Fig. 4, and Fig. 5 the OLS predicts the measurements by the reference equipment with the smallest errors, the model is not recommended. Calibration is needed so that the sensors can be used to predict unseen data, and if the model is overfitted to the calibration dataset, each time that a measurement is taken outside of the dataset conditions, the sensor will not be reliable. ML and REML, as they consider autocorrelations, will not produce the best fitting of the measured data but will be the models that estimate best the calibration so that the sensor works best when used with new data.

Considering the OLS, based on p-values, TVOC is not a significant parameter, but formaldehyde, RH, and temperature are significant. However, when considering ML and REML, TVOC becomes significant instead of formaldehyde. In the last cases, some VOCs causing cross-sensitivities are produced, e.g., the test with wooden materials, making cross sensitivities so important that TVOC becomes an explanatory variable. According to their manufacturers, the reference formaldehyde sensor has known cross-sensitivity with possible-present VOC such as limonene, styrene, propionaldehyde, n-Nonyaldehyde, benzaldehyde, and acetaldehyde, among others, while the Dart formaldehyde sensor has cross sensitivities with ethanol phenol, ethylene among others and all these could have been degassed from our test. ML and REML weight the data based on the autocorrelations; therefore, TVOC cross-sensitivities with formaldehyde gain importance.

Most published works use temperature and relative humidity in linear fits to increase the fitting (e.g., Crilley et al., [41]). However, for ML and REML, RH is not a significant parameter as these sensors are already compensated for it in the out-the-box measurements [33].

Fig. 3 shows the autocorrelation of the residuals of the ML model. The model’s residuals are highly autocorrelated, and this figure shows the importance of considering and describing autocorrelation as it may affect the reliability of the models if not accounted for. By using autocorrelation, the one-step forecast error of ML and REML will be much smaller. The variance of the one-step forecast error is proportional to 1 minus the squared value of the autocorrelation in lag 1 [37]. When lag1 autocorrelation is high, the uncertainty of short-term predictions is highly reduced.

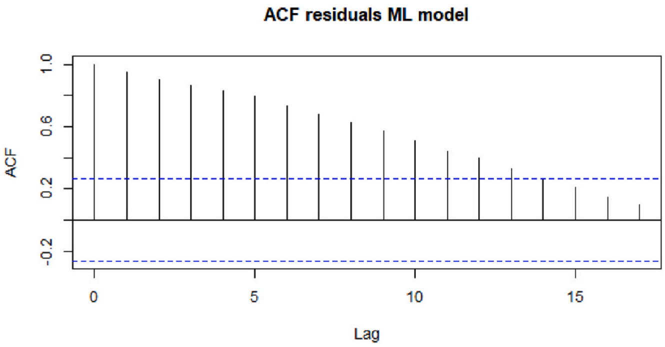


Fig. 3. ACF of the residuals for the ML model.

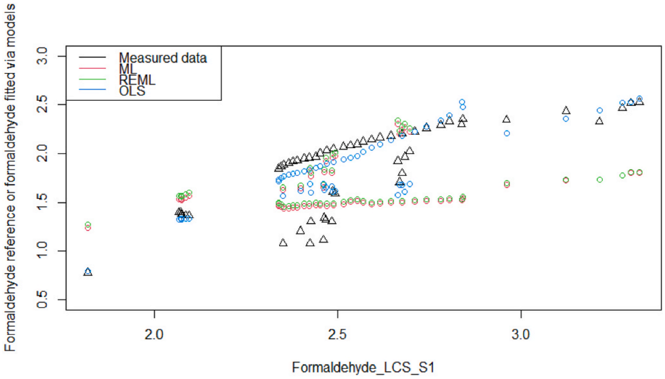


Fig. 4. Fitting of LCS measurements using all measured variables as explanatory variables.

Fig. 4 shows the fitting results when using all the measured parameters. The OLS fitting follows the values on the top part of the graph very well. Most of these values correspond to the same test, test B (defined in Table 2). OLS will not consider the autocorrelation of the values, and as this test is the test with more points, it has a larger effect on the fitting parameters. The suggested ML and REML will account for the autocorrelation, and thus the fitting of these points is much worse. When using REML or ML, the model considers only the "new information" from a test, and thus it predicts individual points in test B poorly, but it predicts much better the ones in the lower part of the graph as they have new information. Overall, the ML and REML show a more balance fit to the data.

Fig. 5 shows the predictions facetted by the test. Tests A, B, and C are the longest tests and thus very well predicted by OLS; however, tests D-F have a similar error for all three methods (test defined in Table 2). ML and REML overpredict results for tests C–F and underpredict in A and B. OLS under and overpredicts in tests A and B, underpredicts in tests C and D, and overpredicts in tests E and F. The fact that results are under and over-predicted makes the calibration more unreliable due to randomness.

Using stepwise regression, a regression model was built that minimizes the AIC value for ML and REML models. A simpler model was developed using the AIC to measure the loss of information while removing variables. The model considering only formaldehyde and temperature as the explanatory variables is selected in this case. The AIC penalizes adding more variables; thus, only the variables that are better predictors are maintained. In this case, TVOC has the lowest AIC value, and it is the parameter where less information will be lost when being removed.

Additionally, multicollinearity is checked. Multicollinearity happens when two or more predictor variables are highly correlated. TVOC and formaldehyde are strongly correlated. Hence when removing multicollinear predictors, the remaining predictor will still contain most of the information [42]. Keep in mind that resulting AIC values with different estimates should not be compared.

3.3. *Calibration using formaldehyde and temperature*

Table 5 shows the results for the model considering only formaldehyde and temperature. Formaldehyde and temperature are significant for all three types of models. OLS results are included in Table 5 to compare to other existing literature but will not be further discussed.

According to EPA's recommendations, the sensors are evaluated as Tier III, supplementary monitoring for all the sensors with these models.

Fig. 6 is not substantially different from Fig. 4. The prediction of the larger values is better with these models. This is probably a consequence of removing TVOC from the model, which gives formaldehyde measurements more weight. The same can be concluded by looking at the smaller MAE and RMSE with fewer parameters.

Both Figs. 6 and 7 show that the LCS predicts the trends of formaldehyde concentration reliably. However, Fig. 7 shows that despite being the error smaller in REML and ML, the models still have a systematic bias. Tests A and B were underpredicted, test C was very well predicted, and D, E, and F were overpredicted. Measurements A and B were performed with no heater and only a humidifier for the latter. During the D and E, the radiator was on, and in F, none was on. Therewas a second difference; these measurements were taken one year apart. In the mid-time, the sensors were exposed to different concentrations of formaldehyde during several measurements, and what we see may be drift due to the loss of baseline or accumulation of material on the oxidizing membrane. The measurement principle relies on a two-electrode electrochemical type, operating by the diffusion principle. Clogging of the membrane may incur wrong measurements or over predictions.

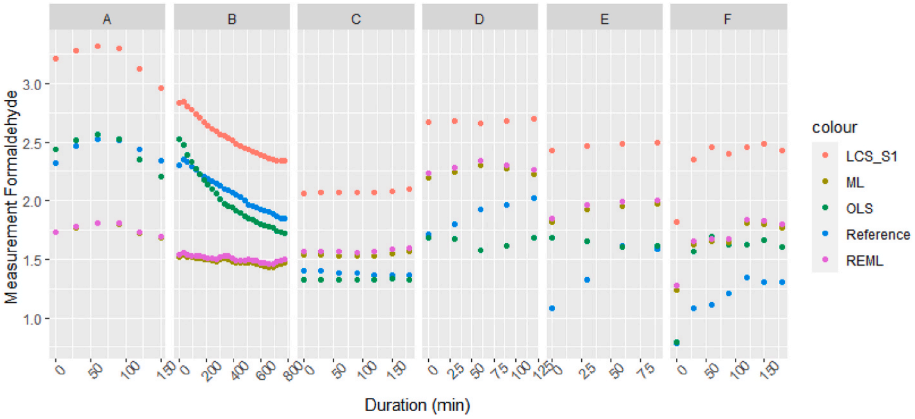


Fig. 5. Prediction of calibrated values using OLS, ML, and REML methods. Results facetted by test.

Table 5
Estimation of parameters using the different estimations.

	OLS	p-value	ML	p-value	REML	p-value
Intercept	4.21	5.7E-04	−6.18	1.0E-04	−6.33	1.00E-04
FA	1.03	1.8E-14	0.63	0	0.62	0
Temperature	−3.69	3.7E-05	4.62	2.0E-04	4.77	1.00E-04
Residual standard error	0.25		0.42		0.53	
R-squared	0.70					
AIC	−159.7		−125.9		−123.7	
RMSE	0.245		0.37		0.37	
MAE	0.193		0.32		0.33	
CV	0.19	Tier III	0.15	Tier III	0.15	Tier III
MNB	9.66e-17		−0.11		−0.09	

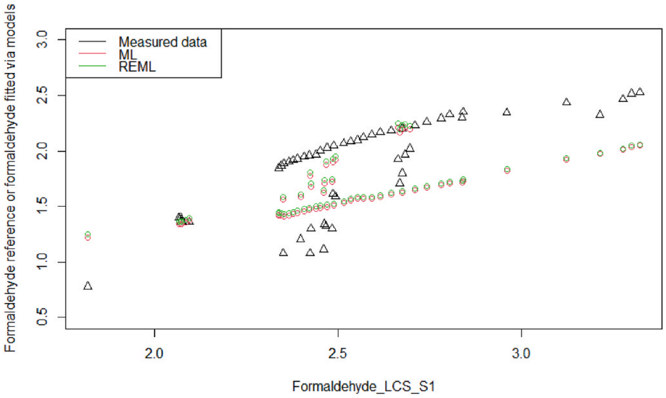


Fig. 6. Fitting of LCS measurements using formaldehyde and temperature as explanatory variables.

It is also important to note that three years after using the sensors, two of the eight sensors stopped working, two years before the sensor's five-year expected lifetime [35]. According to EPA's recommendations, these sensors should only be used for supplementary and not for regulatory monitoring. However, as they respond well to the changes in trends, they can be used to control personal exposure.

3.4. Summary of the essential parameters to make a good calibration

To evaluate sensors, it is common in the literature to develop linear models that correlate measurements with a laboratory-grade sensor and an LCS, and by evaluating the $R^{2,}$ the goodness of the fit or the sensor is concluded.

However, at least three elements should be considered before developing such correlations, 1) the experimental design, 2) the autocorrelations, and 3) the selection of the best model.

3.4.1. Experimental design and limitations of the presented tests

It is essential to make an experimental design that ensures causal and proper dependencies in the data [43]. The following facts affected the selection of the model that could be used for calibration:

- The length of the measurements was not equal for all the tests. Some tests went overnight, and others lasted only a couple of hours. Difference test lengths would affect a linear model by giving overweight to the longer test. For test B, the idea behind having so long measurements was to see if, after exposure, the sensors returned to the initial baseline or if they suffered from drift.
- The same tests were run at slightly different humidity and temperatures, but these were not constant during the test or from test to test. We intended to test at "dry-wet" and "cold-warm" conditions as Demanega did [36]. Neither the radiator nor the humidifier maintained constant conditions, and conditions averaged around the set points. Having constant factors would help establish or not the effect of a single factor. Dependent variables could be measured at different levels, intervals, or ratios, affecting the level of precision attainable here; these variables could only be measured and not controlled [43].
- The selected tests were known from the literature to produce formaldehyde [44]; however, too little attention was set to the different TVOCs that were simultaneously produced, resulting in cross-sensitivities for both formaldehyde and the TVOC sensors. For example, formaline is a known source of formaldehyde, but it contains methanol, to which the WZ modules and the Graywolf sensor have a strong cross-sensitivity. Cross-sensitivity is a prevailing challenge for sensors that measure gaseous pollutants [21].
- If a high autocorrelation is seen for the measurements, then this autocorrelation has to be taken into account to give reliable estimates for highly correlated measurements. As a rule of thumb, conventional OLS methodologies can be used if the autocorrelation is less than, say, 0.3.

3.4.2. Importance of describing the correlation

OLS assumes independent observations, and if this is not fulfilled, then the outlined measures that describe the systematic variation in the data should be used.

In general, LCSs measure with a high sampling rate, i.e., with a small-time distance between the individual data points. In such cases, the calibration error at two consecutive measurements is often highly correlated. Using measurement campaigns with more frequent sampling than needed would often yield overestimated R^2 [40] due to autocorrelations in the time series. In the case of frequent sampling, the errors

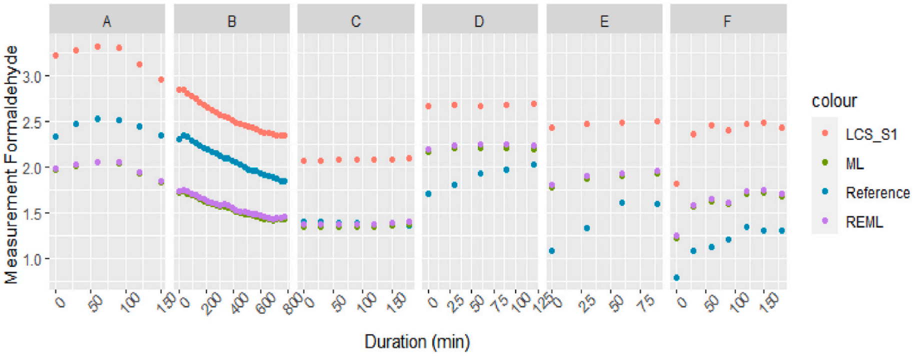


Fig. 7. Prediction of calibrated values using ML and REML methods. Results facetted by test.

from the one-time point are often highly correlated with the sampling error at a neighboring time point. Compared to more scarcely sampling points, the extra points arising from the frequent sampling over a long time do not provide extra information proportional to the relative number of samples. Using several test exposures at different environmental conditions and with different sources rather than very frequently sampled measurements are recommended to develop calibration models.The LCS collected data every 5 min, and the measurements were averaged every 30 min to be compared to the laboratory-grade sensor. Measurements taken continuously will almost always be autocorrelated and result in models with autocorrelated residuals.

One condition of linear regression is that the residuals are independent and identically distributed. However, in calibration, when the sampling is done as here, continuously, this assumption is then often violated. Measurements are taken often with little variations of the source, thus with measurements very correlated to the previous. Consequently, generalized least squares estimation must be considered so that the explanatory variables fully describe the autocorrelation. ML considers the autocorrelation, but REML has better compensations for the estimated and scaling parameters (related to the variance).

3.4.3. Selection of best model

In our analysis, the best model is selected based on a test for significant parameters and a comparison between different model candidates using the AIC criterion. In general, the ML approach provides a robust framework for model selection even in the case of autocorrelated errors. However, the REML approach is preferred for the final parameter estimation since it provides more unbiased estimates of the variances.

3.4.4. Performance of the tested formaldehyde LCS

Both LCS and the reference sensor present very similar dynamic responses, which means that LCS could be used to detect concentration changes. However, there were quantitative discrepancies even after the calibration. These discrepancies over-and under-estimated the values obtained by the reference sensor. The LCS and the reference sensor suffer from cross-sensitivities to some VOC released by the tested sources. However, the cross-sensitivity reaction is different for the different chemical compounds and sensors, making the evaluation of the precise values more complicated.

The present study suggests that these sensors have the potential to be used in indoor environments as Supplemental Monitoring. According to EPA's recommendation as Tier III, a sensor can be used to improve the characterization of concentration gradients [22]. This would mean triggering the proper responses for the control, but this control would need to focus on trends better than values.

4. Conclusions

LCSs use is becoming widespread in the market. However, these sensors are often delivered from the provider with limited information regarding use and performance, reliability, and response to aging or drift. This article analyzes the performance of eight IAQ stations with the same formaldehyde sensor type via comparison with reference equipment. Tested sensors were pre-calibrated from the factory at purchase, and the drift is removed via a 12 h calibration before the experiments.

The tests were run in a mini chamber as a collection of measurements of formaldehyde. The lengths of the tests were heterogeneous based on the estimated duration of the exposure. Some tests were run in cold, warm, dry, and wet conditions controlled with a domestic radiator and humidifier. The experimental design did not ensure that tests data were not autocorrelated.

Given the autocorrelation of the measurements, Ordinary Least Squares Estimations should not be used. In this article, there are two alternative methods for evaluating the calibration: Maximum Likelihood and Restricted Maximum Likelihood Estimation. ML considers the autocorrelation, but REML has better compensations for the estimated and scaling parameters (related to the variance). This article has created a procedure for estimating a weighting according to the autocorrelation based on the first-order Markov. Then a simple method using this weighting was created. Finally, the AIC criterion was used to select the most significant parameters, and for the calibration of the formaldehyde sensor, formaldehyde and temperature were estimated as significant parameters.

According to EPA's recommendations, these models evaluate the sensors as Tier III supplementary monitoring. These results are presented for only one of the eight sensors. Out of the eight, one sensor stopped working during the calibration tests (a second one stopped recently after continuous use), and the remaining six presented similar performance.

The main message is that when sensors collect data continuously with a very high-frequency interval, there is often little difference between measurements, which are often highly autocorrelated. OLS cannot be used in this case and different models considering the autocorrelation are necessary. This paper exactly presents such new methods that can use data that are autocorrelated. The practical implication is that these models allow handling heterogeneous test sampling. This means they can use data where in some tests, many samples are taken on the same day, and some tests where much fewer tests are taken, and then measurements after some days without data sampling.

The LCS and the reference sensor suffer from cross-sensitivities to some VOC released by the tested sources. Even if there were discrepancies where the LCS over-and under-estimated the values obtained by the reference sensor, they both presented very similar dynamic responses, indicating that LCS could be used to detect concentration changes. The present study suggests that these sensors have the potential to be used in indoor environments as Tier III supplemental Monitoring (according to EPA's recommendation), especially for triggering appropriate controls.

CRediT authorship contribution statement

Maria Justo Alonso: Writing – original and reviewed draft, Conceptualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Henrik Madsen:** Writing original and reviewed draft, Methodology, Formal analysis, Conceptualization. **Peng Liu:** Writing – original and reviewed draft, Formal analysis. **Rikke Bramming Jørgensen:** Writing – original and reviewed draft, Formal analysis, Supervision. **Thomas Berg Jørgensen:** Investigation. **Even Johan Christiansen:** Software, Resources, Investigation. **Olav Aleksander Myrvang:** Software, Resources, Investigation. **Diane Bastien:** Writing –original draft, Resources. **Hans Martin Mathisen:** Writing – original and reviewed draft, Supervision.

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.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge the support from the Research Council of Norway (project no. 257660) and several partners through the Research Centre on Zero Emission Neighborhoods in Smart Cities (FME ZEN).

References

[1] A. Moum, Å. Hauge, J. Thomsen, Four Norwegian Zero Emission Pilot Buildings – Building Process and User Evaluation, November. 2017.

[2] H. Maula, V. Hongisto, V. Naatula, A. Haapakangas, H. Koskela, The effect of low ventilation rate with elevated bioeffluent concentration on work performance, perceived indoor air quality, and health symptoms, Indoor Air 27 (6) (Nov. 2017) 1141–1153.

[3] M. Collins, S. Dempsey, Residential energy efficiency retrofits: potential unintended consequences, J. Environ. Plann. Manag. (2019).

[4] W.J. Fisk, B.C. Singer, W.R. Chan, Association of residential energy efficiency retrofits with indoor environmental quality, comfort, and health: a review of empirical data, Build. Environ. (2020).

[5] T. Salthammer, Formaldehyde sources, formaldehyde concentrations and air exchange rates in European housings, Build. Environ. (2019).

[6] GAO, Formaldehyde in textiles, Thread Threat Formaldehyde Text (August) (2011) 1–56.

[7] I. Lang, T. Bruckner, G. Triebig, Formaldehyde and chemosensory irritation in humans: a controlled human exposure study, Regul. Toxicol. Pharmacol. (2008).

[8] IARC, Chemical agents and related occupations - formaldehyde, IARC Monogr. Eval. Carcinog. Risks Hum. 100F (2012) 401–435.

[9] P. Wolkoff, Indoor air pollutants in office environments: assessment of comfort, health, and performance, Int. J. Hyg Environ. Health (2013).

[10] Arbeidstilsynet, Klima Og Luftkvalitet På Arbeidsplassen, 444, 2012.

[11] Lovdata, Lov om arbeidsmiljø, arbeidstid og stillingsvern mv. (arbeidsmiljøloven), Arbeidsmiljøloven (2016).

[12] FHI, Luftkvalitetskriterier: virkninger av luftforurensning på helse (2013).

[13] G.G. Mandayo, et al., System to control indoor air quality in energy efficient buildings, Urban Clim. (2015).

[14] A. Polidori, V. Papapostolou, B. Feenstra, H. Zhang, Field Evaluation of Low-Cost Air Quality Sensors Field Setup and Testing Protocol, January, 2017.

[15] R. Williams, A. Kaufman, T. Hanley, J. Rice, S. Garvey, Evaluation of field-deployed low cost PM sensors, U. S. Environ. Prot. Agency (2014).

[16] K.E. Kelly, et al., Ambient and laboratory evaluation of a low-cost particulate matter sensor, Environ. Pollut. (2017).

[17] L. Spinelle, M. Gerboles, M.G. Villani, M. Aleixandre, F. Bonavitacola, Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2, Sensor. Actuator. B Chem. (2017).

[18] L. Spinelle, M. Aleixandre, M. Gerboles, Protocol of Evaluation and Calibration of Low-Cost Gas Sensors for the Monitoring of Air Pollution, 2013.

[19] B.C. Singer, W.W. Delp, Response of consumer and research grade indoor air quality monitors to residential sources of fine particles, Indoor Air (2018).

[20] H.Y. Liu, P. Schneider, R. Haugen, M. Vogt, Performance assessment of a low-cost PM 2.5 sensor for a near four-month period in Oslo, Norway, Atmosphere (Basel). (2019).

[21] Y. Kang, L. Aye, T.D. Ngo, J. Zhou, Performance evaluation of low-cost air quality sensors: a review, Sci. Total Environ. 818 (Apr. 2022) 151769.

[22] R. Williams, et al., Air Sensor Guidebook (1) (2014) 1–5. *Epa/600/R-14/159*.

[23] J.T.S. Martin Eling, Hato Schmeiser, Low-cost Outdoor Air Quality Monitoring and Sensor Calibration: A Survey and Critical Analysis, Arxiv, 2021.

[24] M. Müller, et al., Integration and calibration of non-dispersive infrared (NDIR) CO2 low-cost sensors and their operation in a sensor network covering Switzerland, Atmos. Meas. Tech. (2020).

[25] H. Baltruschat, N.A. Anastasijevic, M. Beltowska-Brzezinska, G. Hambitzer, J. Heitbaum, Electrochemical detection of organic gases. The development of a formaldehyde sensor, Berichte der Bunsengesellschaft/Physical Chem. Chem. Phys. (1990).

[26] R. Piedrahita, et al., The next generation of low-cost personal air quality sensors for quantitative exposure monitoring, Atmos. Meas. Tech. (2014).

[27] M.R. Giordano, et al., From low-cost sensors to high-quality data: a summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors, J. Aerosol Sci. (2021).

[28] GrayWolf Sensing Solution c), "Formaldehyde Multimode Monitor FM-801.".

[29] Pegasor, Pegasor AQTM indoor air quality monitor operating manual, TLS Times Lit. Suppl. 5911 (2016) 29.

[30] M. Justo Alonso, R.B. Jørgensen, H. Madsen, H.M. Mathisen, Performance assessment of low-cost sensors under representative indoor air conditions, in: Indoor Air 2022 Conference, 2022.

[31] A.I. Ladyada, Adafruit SGP30 Arduino Library." [Online]. Available: https://adafruit.github.io/Adafruit_SGP30/html/index.html.

[32] Sensirion, Sensirion/Embedded-Sht Public, 2018 [Online]. Available: https://github.com/Sensirion/embedded-sht/blob/master/shtc1/shtc1.c.

[33] Marco Höhener, Multi-Gas, Humidity and Temperature Module SVM30, NRND, Sensirion, 2019 [Online]. Available: https://www.sensirion.com/en/environmental-sensors/gas-sensors/multi-gas-humidity-temperature-module-svm30/. (Accessed 15 December 2020).

[34] C.T. R, R: A Language and Environment for Statistical Computing," Vienna, Austria, 2020.

[35] Dart, Dart Sensors WZ-S Formaldehyde Module Operation Manual, 2019.

[36] I. Demanega, I. Mujan, B.C. Singer, A.S. Anđelković, F. Babich, D. Licina, Performance assessment of low-cost environmental monitors and single sensors under variable indoor air quality and thermal conditions, Build. Environ. (2021).

[37] H. Madsen, Time series analysis, Chapman & Hall (2007).

[38] H. Madsen, P. Thyregod, Introduction to General and Generalized Linear Models, 2010.

[39] M. Taddy, Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions, 2019.

[40] C. Malings, et al., Fine particle mass monitoring with low-cost sensors: corrections and long-term performance evaluation, Aerosol Sci. Technol. (2020).

[41] L.R. Crilley, et al., Effect of aerosol composition on the performance of low-cost optical particle counter correction factors, Atmos. Meas. Tech. (2020).

[42] Z. Zhang, Variable selection with stepwise and best subset approaches, Ann. Transl. Med. (2016).

[43] S. Bell, Experimental design, in: International Encyclopedia of Human Geography, 2009.

[44] M.Z.M. Salem, M. Böhm, Understanding of formaldehyde emissions from solid wood: an overview, Bioresources (2013).