Efficient and Automated Generation of Orthogonal Atmospheres for the Characterization of Low-Cost Gas Sensor Systems in Air Quality Monitoring

Georgi Tancev, *Member, IEEE*, Andreas Ackermann, Gérald Schaller, and Céline Pascale

***Abstract*— This article presents a compact continuous-flow automaton for the metrological characterization of an array of low-cost gas sensor systems (up to 17 devices) used in air quality monitoring. The automaton readily generates homogeneous gas mixtures of carbon monoxide (CO), nitrogen monoxide (NO) and dioxide (NO_2), and ozone (O_3) in the parts-per-billion range (ppb, nmol mol^{-1}) that are stable, that is, dispersion of 1 ppb (NO, NO_2, O_3) and 9 ppb (CO), with response times of 2 min (CO, NO, O_3) and 30 min (NO_2). The resulting mixtures, which are traceable to the International System of Units (SI) due to the usage of calibrated high-grade reference instruments, can be humidified [0%–60% relative humidity (RH)] and, in addition, the devices under test (DUTs) can be heated (to temperatures between 5 °C and 30 °C) to systematically simulate different atmospheric environments. The application of fractional factorial designs makes the protocol efficient and leads to orthogonal variables. With the presented installation, ten low-cost gas sensor systems are calibrated and an uncertainty estimation is performed. The average values of the relative standard uncertainties across all DUTs are estimated as 52% (CO), 61% (NO), 22% (NO_2), and 21% (O_3).**

***Index Terms*— Air quality monitoring, calibration, design of experiments, gas sensor, low cost, metrology.**

I. Introduction

AIR pollution is harmful to human health and leads to premature death [1], [2], [3]. Meanwhile, studies have shown that air quality can vary greatly across space and time within cities [4], but high-grade measurement instruments are scarcely distributed due to their high costs. Therefore, individual exposure generally remains unknown. Thus, there has been a wide interest in high-resolution air quality monitoring with low-cost gas sensors, for example, for carbon monoxide (CO), nitrogen monoxide (NO) and dioxide (NO_2), and ozone (O_3), in the last years [5], [6], [7], [8], [9], [10]. These sensors, however, suffer from cross-sensitivities [11], interferences with environmental factors like relative humidity (RH) [12] and temperature (T) [11], unit-to-unit variability [11], and aging [13].

To counter some of these effects and to meet data quality objectives such as the ones defined in the Directive 2008/50/EC [14], different sensors are usually merged into the so-called low-cost sensor systems (known as sensor fusion) so that cross-sensitivities and interferences can be calibrated. Examining that many variables, each with several levels, in a laboratory setting is undeniably laborious and, as a consequence, rarely affordable. Consequently, low-cost sensors are usually characterized or calibrated in field campaigns (i.e., by placing them next to high-grade environmental monitors) [15], [16] using machine-learning algorithms such as neural networks [5] or random forests [8].

However, it can be shown that field calibration only captures the local pollutant distribution (both with respect to space and time) [17], [18], [19]. The main problem lies in the field data used for calibration, as the underlying pollutant distribution is not representative of every location and season (i.e., a sampling bias), leading to relocation problems and concept drift. No causality can be established in field studies, as all variables of interest are generally correlated (i.e., not orthogonal), as they are generated by the same processes (e.g., combustion).

To avoid these problems, the distinct factors must be orthogonal [20]. To achieve this, several factors must be varied between individual experiments, that is, complex gas mixtures must be generated under varying environmental conditions, which is anything but trivial (e.g., due to potential chemical reactions). At the same time, the generation of reference conditions should also be affordable and traceable to the International System of Units (SI) [21]. While producing arbitrary gas mixtures is well established (e.g., in chemical manufacturing), the generation of SI-traceable mixtures for metrological characterization of measurement instruments is more challenging, that is, having mixtures of well-known measurement uncertainty. This is due to potential chemical or physicochemical reactions when generation is performed using methods such as dynamic dilution. For instance, NO and O_3 react to NO_2, and NO_2 is adsorbed strongly on certain surfaces. Another important aspect is that calibration must be affordable, particularly for low-cost products. Hence, the amount of manual labor should be reduced to a minimum.

Meanwhile, automated gas generation and mixing with diffusion/permeation tubes and dynamic dilution followed by humidification within exposure chambers have been explored [22], [23], whereas both existing works use calibrated instruments as reference methods. Although high-volume

Manuscript received 29 June 2022; revised 18 July 2022; accepted 1 August 2022. Date of publication 16 August 2022; date of current version 8 September 2022. This work was funded by Innosuisse under Grant 36779.1 IP-ENG. The Associate Editor coordinating the review process was Dr. Rosenda Valdes Arencibia. *(Corresponding author: Georgi Tancev.)*

The authors are with the Swiss Federal Institute of Metrology (METAS), 3003 Bern, Switzerland (e-mail: georgi.tancev@metas.ch).

Digital Object Identifier 10.1109/TIM.2022.3198747

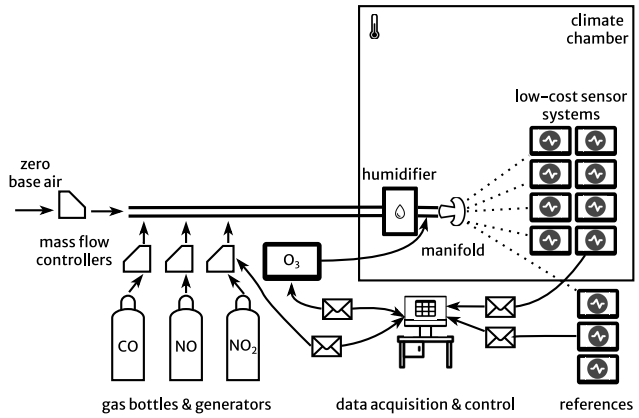


Fig. 1. Schema of the setup.

installations enable parallel sampling of many different sensors or devices under test (DUTs) regardless of their opening, they also come with several limitations, such as large response time, risk of inhomogeneity, and large consumption of resources.

Thus far, a compact continuous-flow installation with a short response time and low risk of inhomogeneity as well as an efficient protocol in which all relevant parameters are varied simultaneously in a systematic manner has not yet been presented. In doing so, the orthogonality of the synthetic atmospheres must be guaranteed, which is crucial for multidimensional calibration models with interactions. This motivates the usage of experimental designs and fractional factorial designs, in particular, leading to an efficient protocol with a minimum number of experiments [20], [24].

This work builds on the existing knowledge and demonstrates how the metrological characterization (and traceable calibration) of low-cost gas sensor systems can be made affordable. In particular, mixtures of CO, NO, NO_2, and O_3 are dynamically generated and humidified, whereas DUTs can be tempered within a climate chamber; high-grade reference instruments serve as a reference method. Moreover, it will be demonstrated how these mixtures can be delivered to the DUTs in parallel, how it is guaranteed that the number of experiments is minimal, and how the resulting atmospheres (i.e., combinations of pollutants and environmental factors) become orthogonal. A web application for the design of orthogonal experiments is made available. Most importantly, no flooding of a high-volume chamber is required, thereby leading to a comparatively responsive system with low consumption of resources and homogeneous mixtures so that a variety of synthetic atmospheres can be simulated in a few experiments. In a final step, the installation is used to characterize ten low-cost gas sensor systems of a collaborator.

II. Materials and Methods

A. Hardware

1) Gas Generation: A schema of the setup is shown in Fig. 1 and was operated according to ISO/IEC 17025 and ISO/IEC 6145 [25]. All instruments were interconnected by perfluoroalkoxy alkane tubings and metal fittings (6.35 mm,

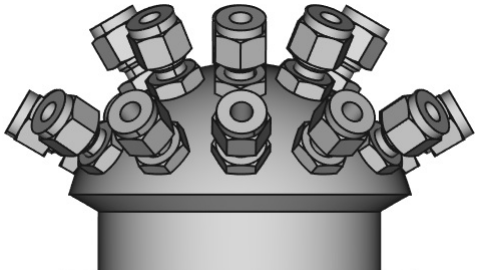


Fig. 2. Schema of the manifold (fabricated from polytetrafluoroethylene).

Swagelok Company, Solon, OH, USA). The (dry zero) base air stream was fixed to 7.0 L min^{-1}, whereas the flow was regulated by a calibrated mass flow controller (SFC5400, Sensirion AG, Stäfa, Switzerland). Amount fractions of CO, NO, and NO_2 in the parts-per-billion (ppb, nmol mol^{-1}) range were generated from gas cylinders (20×10^3 to 100×10^3 ppb (CO and NO in nitrogen, NO_2 in synthetic air), Messer Schweiz AG, Switzerland), also regulated via calibrated mass flow controllers (SFC5400, Sensirion AG, Switzerland). The mixing region was designed to yield instantaneous blending and the design is described in [26]. The gas mixture was humidified using a water tempering system (RC6 & RC22, MGW Lauda, Germany) in combination with nafion tubings (FC125-240-10MP, Perma Pure LLC, Lakewood, NJ, USA), whereas the resulting RH was measured and regulated according to the desired set point. O_3 was injected (after humidification) from a dynamic gas generator (Sonimix4001, LNI Swissgas SA, Switzerland) with a fixed flow of 3.3 L min^{-1} but variable amount fractions.

The mixture was then delivered to the reference instruments and DUTs via some custom-made manifold (depicted in Fig. 2), fabricated from polytetrafluoroethylene to minimize physicochemical interactions. The design of the manifold enables parallel characterization of up to 17 DUTs and guarantees that every device that is connected measures at the same location. One of the openings was not connected to avoid overpressure. Finally, yet importantly, all DUTs were placed inside a climate chamber to temper the constituents.

2) References: Since reactions between the different compounds are possible (NO and O_3, in particular [27]), the whole process was monitored by reference instruments that were periodically calibrated with national standards; these measurements served also as the reference method for calibration of the DUTs. Tubings from manifold to reference instruments and DUTs were of equal length to account for potential reactions. In addition, the tubings outside the climate chamber were insulated to prevent condensation. CO ($u^1 = 1.0\%$) and RH ($u = 2.0\%$) were measured by a cavity ring-down analyzer (G2401, Picarro Inc., Santa Clara, CA, USA), NO ($u = 1.2\%$) and NO_2 ($u = 1.2\%$) by a chemiluminescence analyzer (42*i*, Thermo Fisher Scientific, Waltham, MA, USA), and O_3 ($u = 0.6\%$) by an ultraviolet photometric analyzer (49*C*, Thermo Fisher Scientific). T ($u = 0.2\%$) within the climate chamber was measured by a thermometer (Almemo2890-9/Pt100, Ahlborn,

[1]Throughout the article, standard uncertainties ($k = 1$) are used.

TABLE I
FRACTIONAL FACTORIAL DESIGN FOR SIX VARIABLES

Experiment	CO ppb	NO ppb	NO_2 ppb	O_3 ppb	RH %	T °C
1	850	15	150	150	5	5
2	50	150	150	150	5	5
3	50	15	15	150	5	5
4	850	150	15	150	5	5
5	850	15	150	15	60	5
6	50	150	150	15	60	5
7	50	15	15	15	60	5
8	850	150	15	15	60	5
9	50	15	150	15	5	30
10	850	150	150	15	5	30
11	850	15	15	15	5	30
12	50	150	15	15	5	30
13	50	15	150	150	60	30
14	850	150	150	150	60	30
15	850	15	15	150	60	30
16	50	150	15	150	60	30

Ilmenau, Germany). The sampling rate was set to 4 min^{-1}, but the time series was then resampled to average values of 30 min (i.e., a mean from 120 samples).

3) DUTs: Ten low-cost gas sensor systems (protoypes, LNI Swissgas SA, Versoix, Switzerland) were characterized with the setup, repeating the experimental protocol in Table I three times. No assessment of the sensors or systems was performed before these experiments. The low-cost sensor systems consisted of electrochemical sensors for CO, NO, NO_2, and NO_2/O_3 (B4 series, Alphasense Ltd., Braintree, U.K.) [28], [29], [30], [31], as well as sensors for RH and T. All measurement values were only available in mV. The original sampling frequency was 4 min^{-1}, but the resulting time series was resampled to average values of 30 min (i.e., a mean from 120 samples). The low-cost sensor systems were equipped with fans and sucked actively.

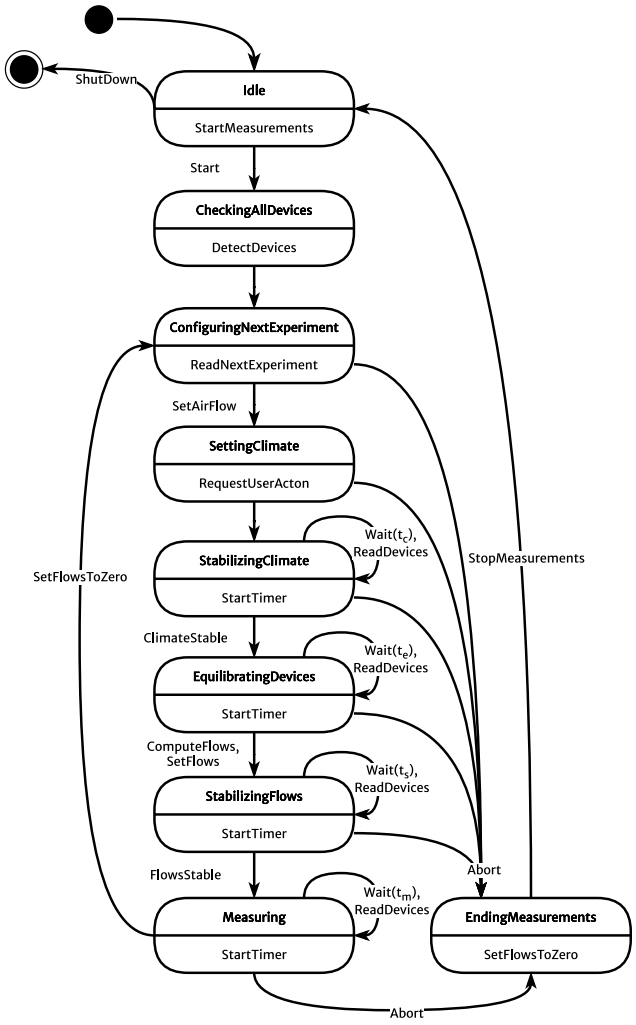


Fig. 3. State machine.

B. Software

1) Design of Experiments: Although arbitrary atmospheric settings can be simulated, the default sequence of experiments performed, that is, combinations of amount fractions and environmental conditions were based on experimental designs [20]. In these designs, factors of interest are orthogonal, and the number of experiments to be performed is minimized, thereby leading to an efficient protocol. As an example, Table I illustrates a fractional factorial design for six variables (written as 2^{6-2}_{IV}), which also served as a protocol for the characterization of the low-cost sensor systems in this study. Each row corresponds to one atmosphere with a unique combination of pollutants and environmental factors. The evaluation of the performance in all atmospheres should result in a representative estimate that is largely independent of location or season.

Moreover, a publicly available web app named design-R, based on the Python library pyDOE [32], was developed with which designs can be created and downloaded. Design-R is available under the domain design-r.herokuapp.com.

2) Calculations: For the automated computation of flows and amount fractions, n mass balances (1) and (2), with input flow rate $Q_{i,\text{in}}$, input amount fraction $x_{i,\text{in}}$, output flow rate $Q_{i,\text{out}}$, output amount fraction $x_{i,\text{out}}$, and flow rate of the base air stream Q_{air}, were solved with a Python script on-the-fly

$$Q_{i,\text{in}}x_{i,\text{in}} = Q_{i,\text{out}}x_{i,\text{out}} \tag{1}$$

$$Q_{i,\text{out}} = Q_{\text{out}} = Q_{\text{air}} + \sum_{i=1}^{n} Q_{i,\text{in}}. \tag{2}$$

3) State Machine: Software written in LabVIEW (v2020, NI, TX, USA) communicated with all instruments (see Fig. 1). Its function was to send commands to the mass flow controllers and generators (to set flows/amount fractions) and to request measurement values from all reference instruments. Moreover, all measured values were recorded in one single (TDMS) file at a rate of 4 min^{-1}. The state machine of the automaton is depicted in Fig. 3. Initially, the automaton starts the measurement process by checking all instruments. Once the test passes, it requests an experimental design (like the one in Table I) as input. Next, it starts the base airflow and requests the user to set the climate conditions manually, that is, adjust T and RH. If the user confirms that these have been set, the automaton waits until the set points are met (t_c). In general, the climate

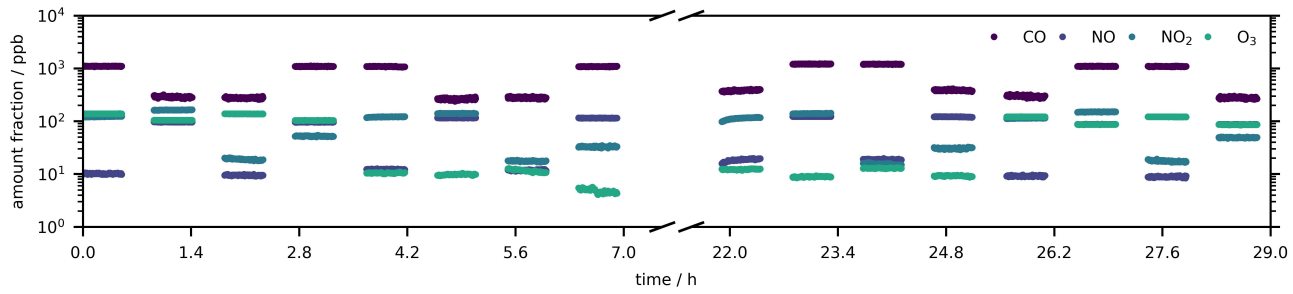


Fig. 4. Visualization of the reference device measurements in a typical sequence of experiments; only the retained part of the measurements is shown.

chamber heats up/cools down within 1–2 h and the humidifier does so within 20–30 min. However, even if the set points are met, the DUTs require some time to reach equilibrium with the environment. Therefore, climate stabilization is followed by waiting for the devices to reach equilibrium with the environment (t_e). In a subsequent step, the flows and amount fractions are computed, and another timer starts for the gas mixture to stabilize (t_s). Finally, the actual measuring process is initiated (t_m). After that, all flows are stopped and the next experiment is prepared.

C. Uncertainty Estimation

The experimental design protocol (see Table I) was performed three times within three weeks. Within each of the three repetitions (runs), estimates of the calibration parameters were obtained. An extended calibration model was developed (i.e., predicting the amount fraction of each pollutant with all available sensors). More specifically, the collection of all $n = 16$ experiments leads to averages of $p = 6$ sensor signals $S \in \mathbb{R}^{n \times (p+1)}$ which can be mapped to the mean of the $q = 4$ reference signals $R \in \mathbb{R}^{n \times q}$ with parameters $\beta \in \mathbb{R}^{(p+1) \times q}$ (3). The additional dimension is for the intercept

$$R = S\beta. \tag{3}$$

These parameters are found via linear regression using the Python library scikit-learn [33], whereas each run leads to a set of calibration parameters. Before performing the fit, each variable (raw sensor signal) in matrix S was scaled to the range of $[0, 1]$. The quality of the models was assessed with the coefficient of determination R^2 that essentially evaluates the amount of information contained in the sensor signals. (In other words, it evaluates the suitability of the basis S for the prediction of R.)

The uncertainty of the predictions, $u_{\text{pred.}}$, was estimated according to the law of uncertainty propagation [34] and consisted of two relevant terms. The first term is the average relative root mean squared error (RMSE), $\bar{e}_{\hat{r}}$. It was computed by squaring and averaging the relative residuals $e_{\hat{r}}$ between predictions $\hat{r}$ and reference values r over all runs and all experiments, given by (4) and (5)

$$e_{\hat{r}}^2 = \left(\frac{\hat{r} - r}{r} \times 100\right)^2 \tag{4}$$

TABLE II
CORRELATIONS BETWEEN REFERENCE MEASUREMENTS

	CO	NO	NO_2	O_3	RH	T
CO	1.00					
NO	0.02	1.00				
NO_2	0.00	0.23	1.00			
O_3	−0.06	−0.32	0.04	1.00		
RH	−0.08	−0.08	−0.00	−0.04	1.00	
T	0.07	0.01	−0.05	−0.07	−0.03	1.00

$$\bar{e}_{\hat{r}} = \sqrt{\frac{\sum e_{\hat{r}}^2}{n}}. \tag{5}$$

However, a residual was only included in the sum, if the predicted reference value was greater or equal to 10 ppb. The second term is the uncertainty of the corresponding reference instrument, $u_{\text{ref.}}$. The resulting formula is given by (6). Note that this leads to an uncertainty estimate that is representative of the whole range of generated amount fractions (see Table I)

$$u_{\text{pred.}} = \sqrt{\bar{e}_{\hat{r}}^2 + u_{\text{ref.}}^2} \tag{6}$$

To assess the added value of sensor fusion, simple calibration models were also fit (with unscaled variables).

III. RESULTS AND DISCUSSION

A. Characterization of the Installation

Fig. 4 and Fig. 5 depicts a typical sequence of experiments (the one shown in Table I) in which the four gases as well as RH and T are varied simultaneously, whereas only the retained part of each step is shown. The sequence is split into two working days (with overnight heating of the climate chamber in between). The most important observation is that the set points are not met, which is not a problem per se due to the availability of references. On the one hand, the levels of CO are too high since no air purifier was used (the baseline was around 200 ppb). On the other hand, NO and O_3 react to NO_2. However, this reaction slightly breaks the desired orthogonality. To assess the extent of introduced relationships between the reference signals, a correlation matrix was computed (see Table II). It shows that some correlation is indeed introduced, but the values remain rather low and in any case lower than what can be found in the field [17]. To circumvent this issue and to remove the need for reference instruments, chemical reactions could be characterized and

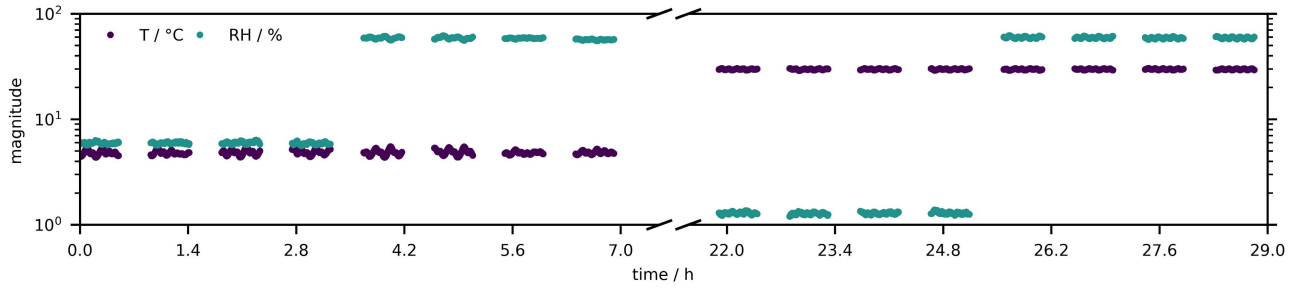


Fig. 5. Visualization of the reference device measurements in a typical sequence of experiments; only the retained part of the measurements is shown.

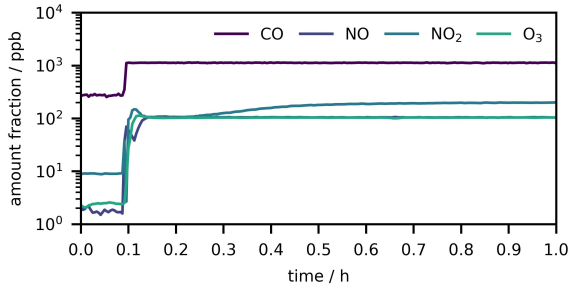


Fig. 6. Time trace of reference device measurements (one single step).

then added to the mass balances for computing the inputs in (1) and (2), or alternatively, a black-box model (mapping from composition and environmental factors to flow rates) could be developed.

Each step lasts exactly 60 min and the response time of the system is rather short, that is, 2–3 min (excluding NO_2), as shown in Fig. 6. The longer response time for NO_2, that is, 20–30 min, originates from its stickiness, as all adsorption sites on the tubings need to be occupied first. Over the course of a measurement campaign, however, this time eventually decreases as all adsorption sites will be occupied. Note that this was also the motivation to discard the first 20–30 min of each step, as it is part of the flow stabilization state (see Fig. 3).

Within the retained part of each step, the reference signals and environmental factors are quite stable (σ_{intra}) as shown in Table III. On average, the values are below 9 ppb for CO and around 1 ppb for NO, NO_2, and O_3, suggesting a homogeneous gas mixture. This is also supported by fluid dynamics, that is, the Reynolds number [$\text{Re} = (\rho u L/\mu)$]. At 20 °C, the density (ρ) and dynamic viscosity (μ) of air are 1.2041 kg m^{-3} and 18.13×10^{-6} kg $\text{m}^{-1}\text{s}^{-1}$, respectively. The inner diameter of the tubings (L) is 4.6×10^{-3} m. The lower bound of the total flow is 0.172×10^{-3} m^3 s^{-1}, leading to a flow velocity (u) of 10.3 m s^{-1}. Hence, the Reynolds number takes a value of 3162, which is above the critical Reynolds number (2250) for pipes and tubes [35], that is, the transition from laminar to turbulent flow. Since the flow is turbulent in this case, it can be expected that the resulting gas mixture is indeed homogeneous. The variability along different runs (σ_{inter}) is slightly higher than σ_{intra}, but still relatively low on average (see Table III). No condensation of water could be observed, which is in part due to the chosen upper level of RH (i.e., 60%).

Nonetheless, there are still a few options on how to improve the installation. In the current state, changes in RH and T are not automatized and have to be adjusted manually. Since the changes between different levels of RH and T are only few, this is not much of a problem, but it could be interesting to automatize that as well in the future to minimize slack time or to operate during the night. (That would allow doing more experiments.) Moreover, the response time for NO_2 could be drastically lowered by using some nonstick coating (e.g., SilcoNert2000) [26], which would enable performing more experiments per unit time.

The setup is currently intended to simulate orthogonal atmospheres using fractional factorial designs with two levels per factor. As a consequence, nonlinearity cannot be captured. This constraint can be relaxed by using mixed-level fractional factorial designs [36], in which orthogonality and balance are traded against efficiency (i.e., the total number of experiments) and additional levels per factor. This would enable an even more representative estimate of the measurement uncertainty, but more experiments would have to be performed. It should be clarified, however, that the installation can assess the steady state but no transient behavior. In particular, it could be shown that humidity transients cause fluctuations in the sensor signals [12]. Hence, results from the laboratory must be validated in field experiments, as the atmosphere in the field is far more complex.

B. Characterization of the Low-Cost Sensor Systems

In the following, the results from the experiments with the low-cost sensor systems are presented. Fig. 7 shows the distribution of the calibration coefficients from the multiple multivariate linear regressions from the three runs and all ten devices (each black dot/bar, that is, $\mu \pm \sigma$, represents one device). Since the variables have been scaled before fitting, the unit of all parameters corresponds to the unit of the output and the magnitude is proportional to the importance of the sensor signal in the chosen intervals (see Table I).

For the prediction of the CO levels, the relevance of RH and T is of a similar magnitude as the sensor signal itself. In other words, without knowing these variables, no reliable prediction of the CO pollution levels is possible. As a consequence, sensor fusion seems strongly recommended, which is also true

TABLE III

STANDARD DEVIATION OF REFERENCE MEASUREMENTS WITHIN EXPERIMENTS (σ_{INTRA}, $n = 120$) AND BETWEEN RUNS (σ_{INTER}, $n = 3$)

	σ_{intra}						σ_{inter}					
Experiment	CO ppb	NO ppb	NO_2 ppb	O_3 ppb	RH %	T °C	CO ppb	NO ppb	NO_2 ppb	O_3 ppb	RH %	T °C
1	7.62	0.17	1.30	0.27	0.12	0.26	54.34	1.35	4.80	3.11	0.23	0.05
2	8.53	0.40	1.12	0.17	0.11	0.19	49.07	0.40	4.09	2.81	0.20	0.05
3	7.11	0.18	0.59	0.23	0.14	0.26	47.28	1.92	0.37	2.81	0.23	0.06
4	8.03	0.44	0.72	0.36	0.12	0.27	29.87	0.72	1.18	2.75	0.19	0.06
5	10.73	0.19	1.22	0.17	1.23	0.29	24.64	2.09	1.14	1.13	1.16	0.08
6	8.78	0.49	0.91	0.22	1.39	0.29	23.08	2.87	1.05	2.42	0.76	0.05
7	8.08	0.21	0.31	0.63	0.47	0.13	10.29	0.42	0.34	1.19	0.60	0.06
8	7.30	0.44	0.67	0.42	0.59	0.15	17.63	2.89	1.42	7.75	1.72	0.05
9	11.52	0.94	5.22	0.18	0.03	0.31	21.73	0.79	1.96	0.71	0.02	0.03
10	8.45	0.41	1.09	0.15	0.03	0.36	32.12	4.30	2.06	8.49	0.02	0.02
11	7.87	0.23	0.53	0.14	0.03	0.32	30.51	2.51	2.09	1.51	0.05	0.01
12	8.95	0.99	0.79	0.17	0.03	0.34	28.97	4.59	1.52	3.34	0.03	0.05
13	10.34	0.19	1.10	0.21	1.14	0.36	1.73	0.21	1.40	0.38	0.38	0.09
14	7.31	0.43	1.04	0.15	1.03	0.31	11.66	2.06	2.69	0.39	0.95	0.01
15	8.38	0.20	0.51	0.31	1.01	0.31	3.04	0.25	0.24	0.61	1.59	0.03
16	8.38	0.38	0.68	0.43	1.03	0.31	6.68	2.49	0.82	0.53	1.08	0.03
Average	8.58	0.39	1.11	0.26	0.53	0.28	24.54	1.86	1.70	2.50	0.58	0.05

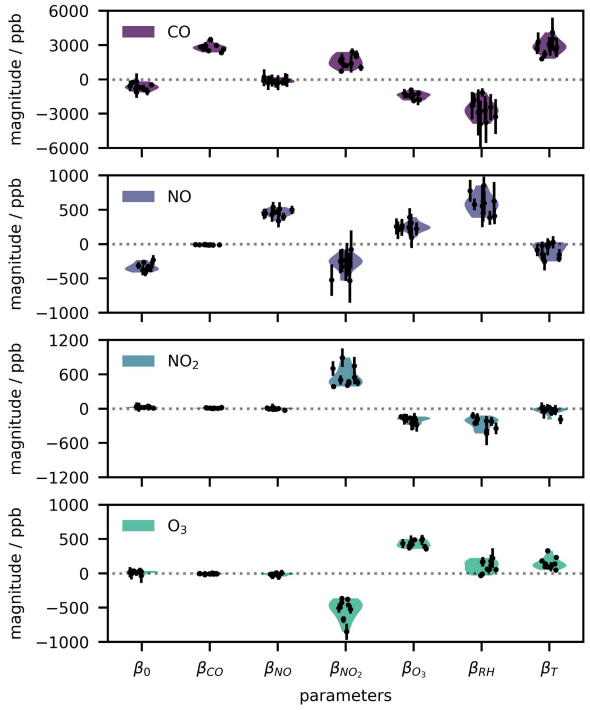


Fig. 7. Parameter distribution of extended linear calibration models across all three runs (dot/bar: $\mu \pm \sigma$) from all DUTs (area: unit-to-unit variability).

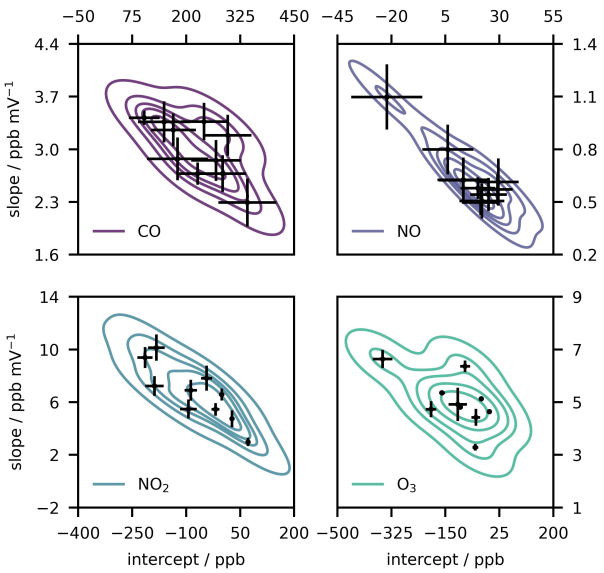


Fig. 8. Parameter distribution of simple linear calibration models across all three runs (dot/bar: $\mu \pm \sigma$) from all DUTs (isoclines: unit-to-unit variability).

for the other pollutants. In addition, all sensors appear to see NO_2 and O_3 as both sensor signals seem to be required for all calibration models. In general, the calibration coefficients vary from run to run, suggesting low repeatability with significant unit-to-unit variability.

Fig. 8 shows the distribution of the coefficients using a simple calibration model (using nonscaled input variables this time). In this plot, the dispersion of the coefficients across units and runs becomes even more apparent. Ideally, every low-cost sensor system should be calibrated individually to obtain accurate measurements. For the other three pollutants, similar observations can be made. However, the repeatability for NO_2 and O_3, which are two of the most relevant pollutants, is much better according to Figs. 7 and 8.

The R^2 and RMSE scores of both models (Figs. 9 and 10) suggest once again that calibrating interfering effects is essential for low measurement uncertainty. For the extended calibration model, the average RMSE values are 95 ppb (CO), 25 ppb (NO), 10 ppb (NO_2), and 5 ppb (O_3); the average R^2 values are 0.94 (CO), 0.68 (NO), 0.96 (NO_2), and 0.99 (O_3). For the simple calibration model, the average RMSE values are 367 ppb (CO), 44 ppb (NO), 41 ppb (NO_2), and 39 ppb (O_3); the average R^2 values are 0.19 (CO), 0.08 (NO), 0.39 (NO_2), and 0.45 (O_3). Fig. 11 shows low-pass filtered sample

predictions (first half of Table I) from one of the devices after an extended calibration.

The high R scores suggest that the available basis vectors [columns of S in (3)] are sufficient for the collected dataset. However, if center points were available, the same basis vectors might lead to lower R^2 scores because it might turn out that the true relationship was nonlinear. Supposing that the true function was indeed not linear but convex, the obtained models would overestimate the amount fraction (with the opposite being true for a concave function).

For a high-end product, this would be certainly not acceptable; for low-cost gas sensor systems, such an error can be considered acceptable as it is most likely smaller than the error that would result if the cross-sensitivities and interferences were not characterized at all. An equivalent Box–Behnken design with three levels per factor would require 64 experiments instead of 16 [20], thereby increasing the workload substantially. Mixed-level fractional factorial designs [36] would be a potential solution.

Both scores are significantly higher in the case of an extended calibration model. More precisely, the additional sensors contain valuable information for the reliable prediction of pollution levels. Furthermore, the models for NO_2 and O_3 appear to be much more reliable than the ones for NO according to the RMSE, which is related to the measurement uncertainty (6). For CO, there appears to be a large spread in the quality. It is also important to stress that these scores assess the model as a whole and not just the gas sensor for the corresponding pollutant. For example, malfunctioning sensors for RH and T would also lead to low performance because the RH and T sensor signals contain relevant information.

Overall, the observations are largely consistent with earlier findings using the same gas sensors in field campaigns [9], [10]. Mueller *et al.* [11] also observed the relevance of RH and T, high unit-to-unit variability, and time-varying sensor behavior for NO_2 and O_3 calibration models. However, they achieved RMSEs of 5–7 ppb (NO_2) and 2–5 ppb (O_3), respectively. Zimmerman *et al.* [8] reported mean absolute errors of 40 ppb (CO) and 5 ppb (NO_2, O_3) using multiple multivariate linear regressions. (These absolute errors could be even lowered with random forest models, hinting at some nonlinearity.) Bigi *et al.* [37] reported RMSE values below 5 ppb (NO, NO_2), which is rather impressive given how low the estimated performance of the NO sensor is.

Evidently, a higher model performance would be traced back to a better engineered low-cost sensor system, models, or methodology (e.g., data preprocessing and filtering). Nonetheless, the cause of higher performance in field studies might also have other causes. Outdoors and pollutants are correlated because they stem from the same chemical processes [17]. Consequently, one good gas sensor such as the one for NO_2 or O_3 could also predict more than one pollutant due to the existing correlations. In this manner, a NO_2 sensor could also accurately "measure" (or rather, infer) the pollution levels of NO. In the laboratory, this is inhibited by the orthogonal factors. At the same time, selection bias, for example, due to preliminary screening in the laboratory, might also be an issue.

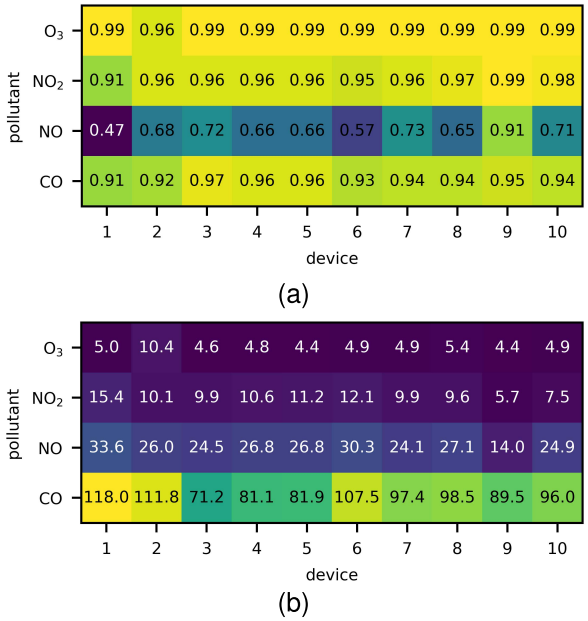


Fig. 9. (a) Coefficient of determination (R^2) and (b) absolute RMSE (in ppb) of the predictions for each device and pollutant using an extended model.

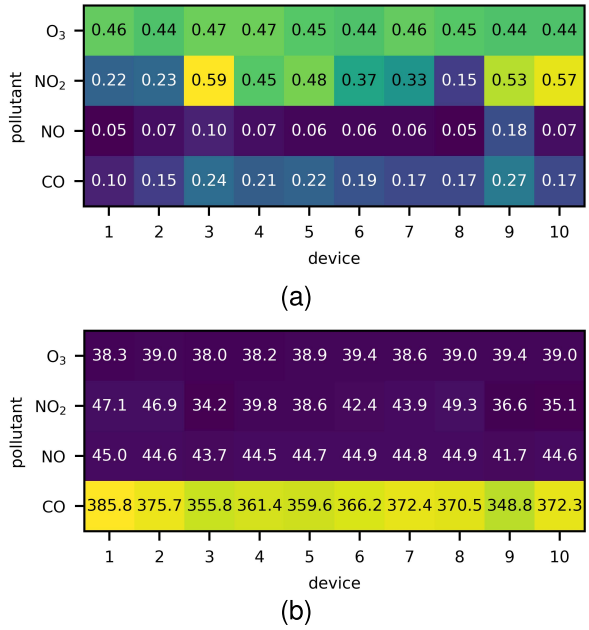


Fig. 10. (a) Coefficient of determination (R^2) and (b) absolute RMSE (in ppb) of the predictions for each device and pollutant using a simple model.

As for the estimated standard measurement uncertainty, Fig. 12 shows its distribution for each pollutant across all devices. The data quality objectives ($k = 1$) for indicative measurements are 12.5% for CO, NO, and NO_2 as well as 15% for O_3 [14]. The standard measurement uncertainties averaged over all devices were estimated as 52% (CO), 61% (NO), 22% (NO_2), and 21% (O_3). The spread among the CO models is the largest, with three below 50%, two above, and the rest somewhere around that value. The predictions of the NO levels have the lowest quality on average. Interestingly, the estimated

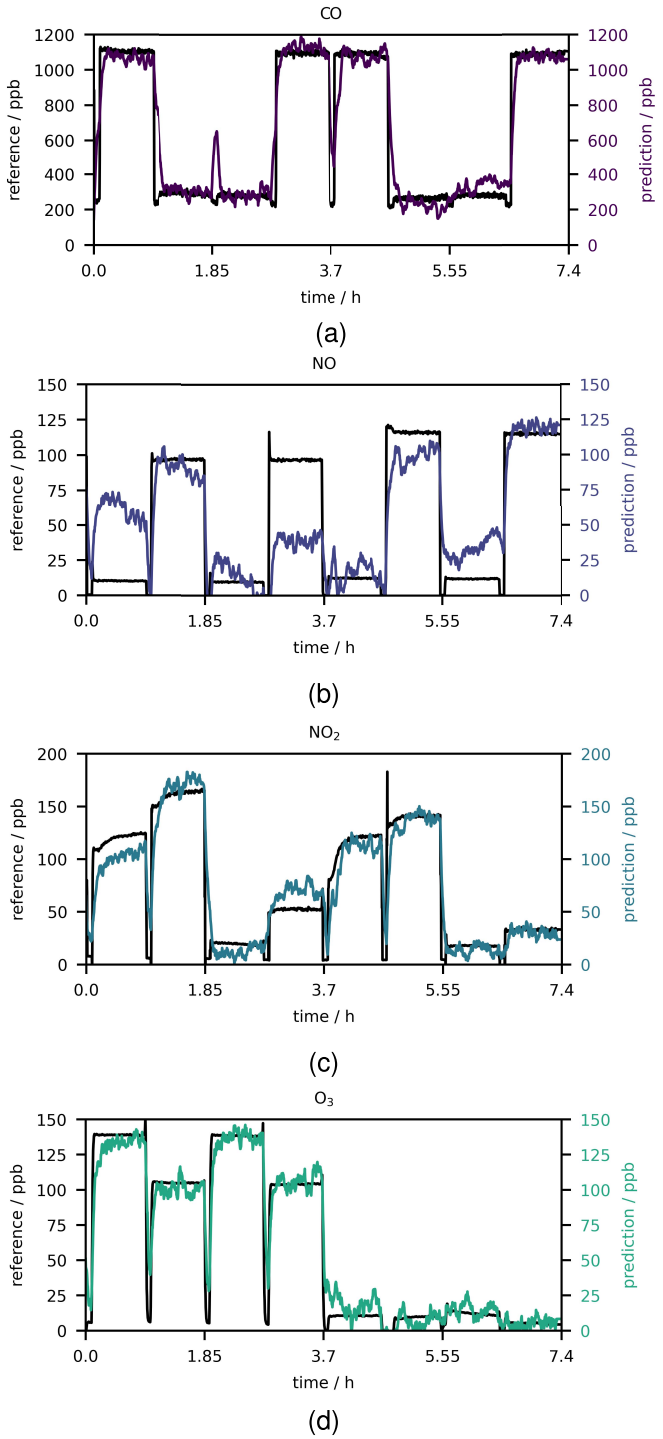


Fig. 11. Low-pass-filtered predictions of (a) CO, (b) NO, (c) NO_2, and (d) O_3 from device 3 after an extended calibration.

uncertainty for NO_2 and O_3 is the lowest and only misses the data quality objective by a few percent. It is important to stress that low and high pollution levels are taken into account in the estimates. Alternatively, the residuals could be normalized by hourly (or daily) limit values (e.g., 100 ppb for NO_2 [14]), thereby potentially fulfilling the data quality objectives.

One other commonly used low-cost technology to estimate pollution levels are passive samplers, which accumulate the mass of pollutants over time via diffusion, thereby providing one average measurement value per location per time interval (e.g., week or month) [38], [39], [40]. After sampling, the concentration is determined via manual chemical analysis, whereas the standard uncertainties can range from 8% to 74% (depending on pollution levels and sampling duration) [41]. Therefore, properly calibrated low-cost gas sensor systems could be an alternative to passive samplers, in particular, for NO_2 and O_3.

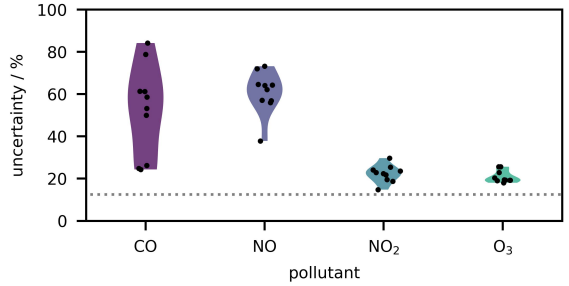


Fig. 12. Standard uncertainties of the measurements across all devices. The dotted line refers to the data quality objective for CO/NO/NO_2 (i.e., 12.5%).

However, these gas sensors are consumables, guaranteeing long-term performance, for example, as part of a wireless sensor network in a smart city, is an unsolved problem. For instance, a recent study suggests the exchange of sensors after six months [13], and consequently, the whole system would need a recalibration. Deploying the systems in batches would also facilitate the recalibration process since they could be recalibrated in batches again. A more modern approach would be to apply a calibration strategy for the whole wireless sensor network [9], [19], [42]. Nonetheless, such an approach should also be SI-traceable but no best practice exists yet.

IV. Conclusion and Outlook

This work presented a compact continuous-flow automaton that efficiently generates orthogonal atmospheres, that is, combinations of pollutants and environmental factors so that the resulting reference variables are hardly correlated. Moreover, up to 17 low-cost gas sensor systems can be characterized in parallel within two working days. The focus of the protocol was to create reliable, yet inexpensive calibration models that are representative, that is, models that perform well regardless of the location at which a low-cost sensor system is deployed.

The automaton generates gas mixtures almost instantaneously with response times of 2–30 min, whereas this period could be shortened in the future by using an appropriate coating, thereby enabling to perform more experiments per run. Moreover, the reference gas mixtures are SI-traceable due to the usage of calibrated reference instruments. Moreover, they are stable as well as homogeneous, with standard deviations ranging from 1 ppb (NO, NO_2, O_3) to 9 ppb (CO) within single experiments. In the current state, the orthogonal mixtures are generated based on fractional factorial designs with two levels per factor, thus nonlinearity cannot be captured. In a subsequent step, however, chemical reactions should be included in the mass balances and mixed-level fractional factorial designs should be explored.

The importance to quantify and correct the effects of cross-sensitivities and interferences with environmental factors could be confirmed in a laboratory setting. In some instances, the interferences with environmental factors were of similar magnitude as the pollutant sensor signal itself. Moreover, unit-to-unit variability and lack of repeatability, especially for CO and NO sensors, were observed to be nonnegligible. Finally, yet importantly, standard measurement uncertainties were estimated as 52% (CO), 61% (NO), 22% (NO_2), and 21% (O_3). These values do not yet meet the data quality objectives for indicative measurements. For NO_2 and O_3, however, the uncertainties are lower or at least as high as the ones from passive samplers. Hence, it could be reasonable to replace passive samplers with well-calibrated low-cost sensors to increase spatial and temporal resolutions. In the next step, it would be important to validate the estimated performance in field experiments, that is, by collocation next to high-grade reference instruments at different sites.

ACKNOWLEDGMENT

The authors would like to thank Stefan Russi for designing the manifold; Noe Steiner for setting up the electronics; Daniel Schwaller, Tobias Bühlmann, Suzana Velikić, and Remo Senn for calibrating all instruments; and LNI Swissgas SA for lending the ozone generator.

AUTHOR CONTRIBUTIONS

Georgi Tancev devised the main conceptual idea, wrote the Python codes, carried out the experiments, analyzed the data, created the figures, and took the lead in writing the manuscript. Andreas Ackermann conceptualized and assembled the setup. Gérald Schaller developed the LabVIEW code. Céline Pascale provided critical feedback.

CONFLICT OF INTEREST

None declared.

REFERENCES

[1] J. R. Balmes and M. D. Eisner, "Indoor and outdoor air pollution," in *Murray and Nadel's Textbook of Respiratory Medicine*, 6th ed. Amsterdam, The Netherlands: Elsevier, 2016, pp. 1331–1342. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/B9781455733835000749

[2] A. J. Alberg, M. V. Brock, and J. M. Samet, "Epidemiology of lung cancer," in *Murray Nadel's Textbook Respiratory Medicine*, 6th ed. Amsterdam, The Netherlands: Elsevier, 2016, pp. 927–939, doi: 10.1016/B978-1-4557-3383-5.00052-X.

[3] S. Khomenko *et al.*, "Premature mortality due to air pollution in European cities: A health impact assessment," *Lancet Planet. Health*, vol. 5196, no. 20, pp. 1–14, 2021.

[4] J. S. Apte *et al.*, "High-resolution air pollution mapping with Google street view cars: Exploiting big data," *Environ. Sci. Technol.*, vol. 51, no. 12, pp. 6999–7008, 2017.

[5] S. De Vito, M. Piga, L. Martinotto, and G. Di Francia, "CO, NO_2 and NO_x urban pollution monitoring with on-field calibrated electronic nose by automatic Bayesian regularization," *Sens. Actuators B, Chem.*, vol. 143, no. 1, pp. 182–191, Dec. 2009.

[6] M. I. Mead *et al.*, "The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks," *Atmos. Environ.*, vol. 70, pp. 186–203, May 2013, doi: 10.1016/j.atmosenv.2012.11.060.

[7] A. C. Lewis *et al.*, "Evaluating the performance of low cost chemical sensors for air pollution research," *Faraday Discuss.*, vol. 189, pp. 85–103, Jul. 2016.

[8] N. Zimmerman *et al.*, "A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring," *Atmos. Meas. Techn.*, vol. 11, no. 1, pp. 291–313, 2018.

[9] B. Maag, Z. Zhou, and L. Thiele, "A survey on sensor calibration in air pollution monitoring deployments," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4857–4870, Dec. 2018.

[10] F. Karagulian *et al.*, "Review of the performance of low-cost sensors for air quality monitoring," *Atmosphere*, vol. 10, no. 9, p. 506, Aug. 2019.

[11] M. Mueller, J. Meyer, and C. Hueglin, "Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of Zurich," *Atmos. Meas. Techn.*, vol. 10, no. 10, pp. 3783–3799, 2017.

[12] A. K. Farquhar, G. S. Henshaw, and D. E. Williams, "Understanding and correcting unwanted influences on the signal from electrochemical gas sensors," *ACS Sensors*, vol. 6, no. 3, pp. 1295–1304, Mar. 2021.

[13] J. Li, A. Hauryliuk, C. Malings, S. R. Eilenberg, R. Subramanian, and A. A. Presto, "Characterizing the aging of alphasense NO_2 sensors in long-term field deployments," *ACS Sensors*, vol. 6, no. 8, pp. 2952–2959, Aug. 2021.

[14] EUR-Lex. *Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on Ambient Air Quality and Cleaner Air for Europe*. Accessed: Jan. 1, 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32008L0050

[15] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola, "Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide," *Sens. Actuators B, Chem.*, vol. 215, pp. 249–257, Aug. 2015.

[16] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola, "Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO_2," *Sens. Actuators B, Chem.*, vol. 238, pp. 706–715, Jan. 2017, doi: 10.1016/j.snb.2016.07.036.

[17] G. Tancev and C. Pascale, "The relocation problem of field calibrated low-cost sensor systems in air quality monitoring: A sampling bias," *Sensors*, vol. 20, no. 21, p. 6198, Oct. 2020.

[18] S. De Vito, E. Esposito, N. Castell, P. Schneider, and A. Bartonova, "On the robustness of field calibration for smart air quality monitors," *Sens. Actuators B, Chem.*, vol. 310, May 2020, Art. no. 127869, doi: 10.1016/j.snb.2020.127869.

[19] G. Tancev and F. G. Toro, "Stochastic online calibration of low-cost gas sensor networks with mobile references," *IEEE Access*, vol. 10, pp. 13901–13910, 2022.

[20] D. C. Montgomery, *Design and Analysis of Experiments*, 8th ed. Hoboken, NJ, USA: Wiley, 2013.

[21] T. J. Quinn and J. Kovalevsky, "Measurement and society," *Comp. Rendus Physique*, vol. 5, no. 8, pp. 791–797, 2004.

[22] N. A. Martin, J. J. Helmore, S. White, I. L. B. Snook, A. Parish, and L. S. Gates, "Measurement of nitrogen dioxide diffusive sampling rates for palmes diffusion tubes using a controlled atmosphere test facility (CATFAC)," *Atmos. Environ.*, vol. 94, pp. 529–537, Sep. 2014, doi: 10.1016/j.atmosenv.2014.05.064.

[23] L. Spinelle, M. Gerboles, M. Aleixandre, and F. Bonavitacola, "Evaluation of metal oxides sensors for the monitoring of O_3 in ambient air at Ppb level," *Chem. Eng. Trans.*, vol. 54, pp. 319–324, Sep. 2016.

[24] J. F. Box, "R. A. Fisher and the design of experiments, 1922–1926," *Amer. Statistician*, vol. 34, no. 1, pp. 1–7, 1980.

[25] METAS. *International Recognition: CIPM MRA—ISO/IEC 17025*. Accessed: Jun. 1, 2022. [Online]. Available: https://www.metas.ch/metas/en/home/metas/managemetnsystem/internationale-anerkennung.html

[26] C. Pascale, M. Guillevic, A. Ackermann, D. Leuenberger, and B. Niederhauser, "Two generators to produce Si-traceable reference gas mixtures for reactive compounds at atmospheric levels," *Meas. Sci. Technol.*, vol. 28, no. 12, Dec. 2017, Art. no. 124002.

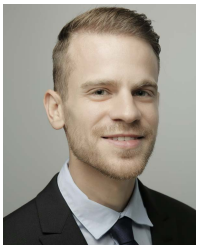
[27] H. H. Lippmann, B. Jesser, and U. Schurath, "The rate constant of NO + $O_3 \rightarrow NO_2 + O_2$ in the temperature range of 283–443 K," *Int. J. Chem. Kinetics*, vol. 12, no. 8, pp. 547–554, Aug. 1980.

[28] Alphasense. *Technical Specification CO-B_4*. Accessed: Jun. 1, 2022. [Online]. Available: http://www.alphasense.com/WEB1213/wp-content/uploads/2019/09/CO-B4.pdf

[29] *Technical Specification NO-B_4*. Accessed: Jun. 1, 2022. [Online]. Available: http://www.alphasense.com/WEB1213/wp-content/uploads/2019/09/NO-B4.pdf

[30] *Technical Specification NO2-B43F*. Accessed: Jun. 1, 2022. [Online]. Available: http://www.alphasense.com/WEB1213/wp-content/uploads/2019/09/NO2-B43F.pdf

[31] *Technical Specification OX-B431*. Accessed: Jun. 1, 2022. [Online]. Available: http://www.alphasense.com/WEB1213/wp-content/uploads/2019/09/OX-B431.pdf
[32] J.-M. Martinez, Y. Collette, M. Baudin, and M. Christopoulou. (2012). *pyDOE*. [Online]. Available: https://pythonhosted.org/pyDOE/
[33] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, Jul. 2017.
[34] *Evaluation of Measurement Data-Guide to the Expression of Uncertainty in Measurement*, Int. Org. Standardization, Geneva, Switzerland, 2008. [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM1002008E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6
[35] D. Barkley, B. Song, V. Mukund, G. Lemoult, M. Avila, and B. Hof, "The rise of fully turbulent flow," *Nature*, vol. 526, no. 7574, pp. 550–553, Oct. 2015.
[36] Y. Guo, J. R. Simpson, and J. J. Pignatiello, "Construction of efficient mixed-level fractional factorial designs," *J. Quality Technol.*, vol. 39, no. 3, pp. 241–257, Jul. 2007.
[37] A. Bigi, M. Mueller, S. K. Grange, G. Ghermandi, and C. Hueglin, "Performance of NO, NO_2 low cost sensors and three calibration approaches within a real world application," *Atmos. Meas. Techn.*, vol. 11, no. 6, pp. 3717–3735, 2018.
[38] G. Ouyang and J. Pawliszyn, "Configurations and calibration methods for passive sampling techniques," *J. Chromatography A*, vol. 1168, nos. 1–2, pp. 226–235, Oct. 2007.
[39] B. Pekey and Ü. Özaslan, "Spatial distribution of SO_2, NO_2, and O_3 Concentrations in an industrial city of Turkey using a passive sampling method," *Clean Soil, Air, Water*, vol. 41, no. 5, pp. 423–428, May 2013.
[40] L. Rosario, M. Pietro, and S. P. Francesco, "Comparative analyses of urban air quality monitoring systems: Passive sampling and continuous monitoring stations," *Energy Proc.*, vol. 101, pp. 321–328, Nov. 2016.
[41] M. Kirchner *et al.*, "Field intercomparison of diffusive samplers for measuring ammonia," *J. Environ. Monitor.*, vol. 1, no. 3, pp. 259–265, 1999.
[42] J. M. Barcelo-Ordinas, M. Doudou, J. Garcia-Vidal, and N. Badache, "Self-calibration methods for uncontrolled environments in sensor networks: A reference survey," *Ad Hoc Netw.*, vol. 88, pp. 142–159, May 2019.



Georgi Tancev (Member, IEEE) was born in Strumica, North Macedonia, in 1990, and moved to Switzerland in 1995. He received the B.S. and M.S. degrees in chemical and bioengineering from the Swiss Federal Institute of Technology (ETH) Zürich, Zürich, Switzerland, in 2015 and 2017, respectively.

He currently works as a Research Scientist with the Swiss Federal Institute of Metrology (METAS), Bern, Switzerland, focusing on trustworthy and reliable air quality monitoring with low-cost sensors. Previously, he worked on process modeling and optimization at Novartis, Basel, Switzerland, as well as in medical device development and clinical data science at the University Children's Hospital Basel, Basel. His research interests include mathematical modeling, numerical simulation, optimization, and machine learning.

Andreas Ackermann was born in Fribourg, Switzerland, in 1968.

In 1986, he finished his apprenticeship as a Laboratory Chemist, which was followed by a diploma for passing the professional examination as Laboratory Technician. He is currently working in this role as part of the Gas Analysis Laboratory, Swiss Federal Institute of Metrology (METAS), Bern. Before joining METAS, he used to work in research and development in the field of novel polymers at Novartis (formerly CibaVision), Basel, Switzerland, for 12 years.

Gérald Schaller received the B.S. degree in electrical engineering from the University of Applied Sciences and Arts of Western Switzerland, Delémont, Switzerland, in 1990.

After graduation, he worked as a freelance Software Engineer. One of these first jobs was to automate the time scale at the Swiss Federal Institute of Metrology (METAS), Bern (with software written in C), Switzerland, which is how he entered the field of measurement science. Since 2000, he has been working permanently as a Software Engineer at METAS and plays a crucial part in many projects across several metrological domains, generally aiming to automate measurement systems with LabVIEW.

Céline Pascale was born in Geneva, Switzerland, in 1987. She received the M.S. degree in chemical and biochemical engineering from the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland, in 2011.

Since 2011, she has been working with the Swiss Federal Institute of Metrology (METAS), Bern, Switzerland, where she is currently the Head of the Gas Analysis Laboratory. In this role, she is responsible for developing and maintaining a metrological infrastructure for gas component amount fractions in the areas of air quality as well as climate and breath analysis. She leads European partnership programs and is part of several technical committees, such as the Consultative Committee for Quantity of Matter—Gas Analysis Working Group (CCQM-GAWG).