Sensors and Actuators B 215 (2015) 249–257

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

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

Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide☆

Laurent Spinelle[a,*], Michel Gerboles[a,**], Maria Gabriella Villani[b], Manuel Aleixandre[c], Fausto Bonavitacola[d]

[a] European Commission, Joint Research Centre (JRC), Institute for Environment and Sustainability (IES), Air and Climate Unit, Via Enrico Fermi 2749, 21027 Ispra, VA, Italy
[b] ENEA, Agenzia Nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile, Ispra, VA, Italy
[c] Institute for Applied Physics, CSIC, Madrid, Spain
[d] Phoenix Sistemi & Automazione s.a.g.l., Muralto, TI, Switzerland

ARTICLE INFO

Article history:
Received 25 October 2014
Received in revised form 17 March 2015
Accepted 19 March 2015
Available online 28 March 2015

Keywords:
Gas sensors
Validation
Measurement uncertainty
Multivariate linear regression
Neural network
Air Quality Directive

ABSTRACT

The performances of several field calibration methods for low-cost sensors, including linear/multi linear regression and supervised learning techniques are compared. A cluster of ozone, nitrogen dioxide, nitrogen monoxide, carbon monoxide and carbon dioxide sensors was operated. The sensors were either of metal oxide or electrochemical type or based on miniaturized infra-red cell. For each method, a two-week calibration was carried out at a semi-rural site against reference measurements. Subsequently, the accuracy of the predicted values was evaluated for about five months using a few indicators and techniques: orthogonal regression, target diagram, measurement uncertainty and drifts over time of sensor predictions. The study assessed if the sensors were could reach the Data Quality Objective (DQOs) of the European Air Quality Directive for indicative methods (between 25 and 30% of uncertainty for O_3 and NO_2). In this study it appears that O_3 may be calibrated using simple regression techniques while for NO_2 a better agreement between sensors and reference measurements was reached using supervised learning techniques. The hourly O_3 DQO was met while it was unlikely that NO_2 hourly one could be met. This was likely caused by the low NO_2 levels correlated with high O_3 levels that are typical of semi-rural site where the measurements of this study took place.

1. Introduction

Compared to the reference methods defined in the Air Quality Directive [1], the use of low-cost gas sensors for monitoring ambient air pollution would reduce air pollution monitoring costs and would also allow larger spatial coverage especially in remote areas where monitoring with traditional facilities is cumbersome. However, the calibration of low-cost sensors for monitoring air quality remains a challenge. The selectivity and stability of sensors are generally found problematic [2–4]. Consequently, more sophisticated algorithms for quantifying air pollution are being developed. Among the studied methods, the temperature cycle operation was shown to limit cross sensitivities and ageing of sensors [5] under laboratory conditions. This method is also relevant for the identification of organic compounds [6]. Kamionka et al. reported the use of several metal oxides (MO_x) sensors operated at different heating temperature [7,8]. Theses multi-sensors were either calibrated against standard gas mixtures or using artificial neural network under field conditions. The latter method resulted in mixed results either satisfactory for short periods or generally weak for longer data series. Neural network calibration has been mainly implemented for the identification of organic compounds and smell [9,10] or for monitoring compounds such as CO or CH_4 at high levels [11]. Few attempts were made to use neural network for the calibration of sensors for monitoring in the low nmol/mol range [12,13]. One of these studies looked at neural network calibration for benzene at nmol/mol levels [14]. However, the majority of the studies mentioned above used the sole MO_x-type sensors which are known to suffer from a lack of stability and long response time

☆ Selected papers presented at EUROSENSORS 2014, the XXVIII edition of the conference series, Brescia, Italy, September 7–10, 2014.

* Corresponding author. Tel.: +39 0332 789572.
** Corresponding author. Tel.: +39 0332 785652.
E-mail addresses: laurent.spinelle@jrc.ec.europa.eu (L. Spinelle), michel.gerboles@jrc.ec.europa.eu (M. Gerboles), mariagabriella.villani@enea.it (M.G. Villani), manuel.aleixandre@gmail.com (M. Aleixandre), fausto.bonavitacola@ingpec.eu (F. Bonavitacola).

http://dx.doi.org/10.1016/j.snb.2015.03.031

Table 1
List of clustered sensors.

Manufacturer	Sensor models	Pollutant	Number of sensors
αSense	O3B4	O_3	1
Citytech	O3_3E1F	O_3	2
αSense	NO2B4	NO_2	2
Citytech	NO2_3E50	NO_2	2
	NO_3E100	NO	2
SGX-Sensotech	MICS-2710	NO_2	2
	MICS-4514-NO2	NO_2	2
CairPol	CairClip NO2	NO_2	2
Figaro	TGS-5042	CO	2
SGX Sensortech	MICS-4514-CO	CO	2
Edinburgh Sensors	Gascard NG	CO_2	1
ELT Sensors	S-100	CO_2	2

[15]. A recent study describes a new real-time field calibration by comparing mobile sensor responses with reference measurements of existing reference monitoring stations [16].

Recently, within the EURAMET MACPoll project [17], the performance of single commercial sensors has been evaluated [18–21] according to a precise protocol [22]. This study produced large datasets of measurements for several compounds under laboratory conditions and field campaigns. Such datasets were not previously available in literature, especially considering the number of controlled parameters (NO_x, O_3, CO, SO_2, CO_2, temperature, relative humidity, wind and pressure).

In this study, an analysis of the performance of different calibration models over a great number of O_3 and NO_2 sensors tested in the same conditions is performed. The performances of these methods were compared taking as indicator their resulting measurement uncertainty. It was then evaluated if the uncertainty could meet the Data Quality Objective (DQO) of the European Air Quality Directive [1].

2. Material and methods

Experiments were carried out in collaboration with the European Reference Laboratory for Air Pollution (ERLAP) at the EMEP station of the Joint Research Centre (45°48.881′ N, 8°38.165′ E). The station is located in a semi-rural area at the NW edge of the Po valley (Italy) and is equipped with meteorological sensors (temperature, relative humidity, wind and pressure) and reference gas analysers for NO_x, O_3, CO, CO_2 and SO_2. These reference measurements were used for data validation, comparison and data treatment of sensor responses.

Based on the evaluation and validation of low-costs sensors [18–21], several sensors were chosen to be grouped in a clustered system able to detect O_3, NO/NO_2, CO and CO_2. The best performing sensors showing the shortest response time, the highest sensitivity and the best repeatability were selected.

2.1. Low-cost sensors

The cluster consisted of 5 NO_2 sensors and 2 CO sensors, both electrochemical and metal oxide type, 1 NO and 2 O_3 electrochemical sensors and 2 infrared CO_2 sensors (see Table 1). For NO_2, MO_x and electrochemical sensors were used in order to benefit from the different inherent cross-sensitivities of both types of sensors. The list of tested sensors is presented in Table 1 with manufacturer and models information. All sensors were connected through NI DAQ boards (NI USB 6009 and NI USB 6018 from National Instruments, USA) to our LabVIEW in-house designed DAQ software. The periodicity of data acquisition was 100 Hz and measurements were averaged every minute without filtering. No data treatment was applied during data acquisition. None of the sensors were calibrated before delivery. The sensors were enclosed into aluminium protective boxes and the evaluation boards were covered with Teflon tape to protect the electronic and to avoid contamination of the sensor.

Two CairClip sensors, model NO2 ANA [23] were supplied by CAIRPOL (La Roche Blanche – France). CairClip is an integrated system that includes an amperometric sensor, a dynamic air sampling, a patented filter, and an electronic circuit which allows a direct real time display of the measured value and complete status with internal data logging. Reliability of the measurement is achieved by limiting the effect of humidity variations by using a gas specific inlet filter combined with dynamic air sampling system.

Citytech sensors (Life Safety Germany GmbH, City Technology, Bonn, Germany) consist of 3 Electrodes amperometric sensors with organic electrolyte. Two O_3 sensors (model O3_3E1F [24]), two NO_2 sensors (model NO2_3E50 [25]) and two NO sensors (NO_3E100 [26]) were tested. Each sensor was mounted on a Citytech evaluation board that converts the raw sensor signal voltage, with the possibility to vary the bias potential, using various load, feedback resistors and different levels of current amplification. The board was configured to give an output of 1V-100 nA with damping 10.

αSense sensors were supplied by αSense Ltd (Essex – United Kingdom). One O_3 sensor (model O3B4 – 4 electrodes [27]) and two NO_2 sensors (NO2B4 – 4 electrodes [28]) were tested. The B4 type sensor is a 4 electrodes electrochemical sensor designed for nmol/mol gas levels. As well as the normal Working, Reference and Counter electrodes, B4 sensors include a 4th auxiliary electrode, which is used to correct for zero current changes. Each sensor gave two signals, the 2nd one being the background signal of the auxiliary electrode that has to be subtracted to the sensor raw response of the working electrode. Each sensor was mounted on a αSense test boards (αSense 4-electrodes Individual Sensor Board (ISB) [29]).

Two models of SGX Sensortech (Neuchâtel – Switzerland) sensors were tested in this study: the MICS 2710 for NO_2 [30] and the MICS 4514 which is a combined NO_2 and CO sensor [31]. Both of these sensors consist of Metal Oxide semiconductor sensors. While the MICS 2710 is used to detect NO_2, the MICS 4514 can detect NO_2 and CO with two different signal outputs. Three MICS-EK1 gas sensor evaluation kits [32] were used. The MICS 4514 were directly soldered on the EK1 adapter by the manufacturer. Based on the manufacturer datasheet, the evaluation kits were operated in manual mode on low heating for the NO_2 sensors (43 mW corresponding to a R_{LOAD} of 1 kΩ) and high heating for the CO sensors (76 mW correspond to a R_{LOAD} of 256 kΩ).

TGS 5042-A00 sensors are manufactured by Figaro (Illinois – USA). It consists in a battery like electrochemical sensor [33]. Two TGS 5042-A00 were mounted on two evaluation modules COM5042 able to convert the sensor output current into a voltage [34].

The carbon dioxide module S-100 is manufactured by TCC ELT (Environment Leading Technology, South Korea) and is based on the NDIR (Non-dispersive Infrared) technology [35].

The OEM Gascard® NG infrared gas sensor (0–1000 μmol/mol) is manufactured by Edinburgh Sensors (Lancashire – UK). It is based on dual wavelength NDIR technology with automatic temperature and pressure corrections using real-time environmental condition measurements. The CO_2 sensor uses an active sampling with a 1 l/min pump.

2.2. Reference measurements

The measuring campaign was performed at the JRC – Ispra station from January 2014 until July 2014. Within this period, only 5 months (from March to July) have been considered as valid and were taken into account in this study. As described in [18–21], the mobile laboratory was equipped with routine analysers, meteorological and low cost sensors:

- For meteorological parameters: ambient temperature, ambient relative humidity, ambient pressure and a 10 m mast for wind speed and wind direction.
- For O_3, a UV Photometric Analyser Thermo Environment 49C, a chemiluminescence Nitrogen Oxides Analyser Thermo 42C for NO_2/NO/NO_x, a Non-dispersive Infrared Gas-Filter Correlation Spectroscopy Horiba APMA 370 for CO, a UV Fluorescent Analyser Thermo 43C TL for SO_2. For CO_2, we used a differential Non-dispersive Infrared Gas Analyser Li-cor 6262.

The gas analysers were calibrated in laboratory before the field tests and then they were checked every month. Field checks were carried out using filtered zero air and span value. This one consisted of low concentration gas cylinder certified by the Joint Research Centre which is accredited for these analyses. The gas cylinders used included concentration levels of 50, 100 and 200 nmol/mol for NO/NOx, 50 nmol/mol for SO_2, 1.3 μmol/mol for CO and 369 μmol/mol for CO_2 (uncertified). An ozone generator Thermo Environment 49 CPPS II model, delivering 100 nmol/mol of ozone, was used for the calibration checks of the ozone analyser. The highest observed calibration drift during field tests consisted of 2.5% for NO/NO_2 and O_3, 4.5% for CO, 2% for SO_2 and 1.5% for CO_2. These drifts are consistent with the uncertainty of the working standards used on field (about 3%) and the data quality objective of reference measurements (15%) given in the European Directive for air quality. Therefore, no corrections of measurements were made apart from the discarding values during maintenance and calibration checks.

3. Calibration methods of the sensors and choice of variables

Three calibration methods were tested: simple linear regression (LR), multivariate linear regression (MLR) established within MACPoll [18–21] and artificial neural networks (ANN) with raw, standardized (scaled) and calibrated sensor responses.

3.1. Linear regression (LR)

For each sensor a calibration function was established by assuming the linearity of the sensor responses with reference measurement for each pollutant. Ordinary linear regression was used with the minimization of square residuals of the sensor responses versus reference measurements. The calibration functions were of the type $R_s = a \cdot X + b$ where R_s represents the sensor responses and X is the corresponding reference measurements of air pollutant. Finally, the measuring function, the converse equation $X = (R_s - b)/a$ was applied to all sensor responses in order to predict air pollutant levels.

Among our dataset, the cases corresponding to the initial two weeks of valid measurements were used for calibration (about 336 hourly values). The remaining data (about 90% of the total dataset) were used for validation of the measuring functions.

3.2. Multivariate linear regression (MLR)

The calibration was carried out using the least square method taking into consideration more than one explanatory variables Y_i. Models were established during the MACPoll studies (see Table 2). Coefficients a, b, c, d and e represent calibration parameters extracted from the multi-linear regression; NO_2, O_3 and NO stand for the reference measurements. RH, T and H_2O are respectively relative humidity, temperature and absolute humidity. As for the LR, the calibration functions consisted of equations of the type $R_s = f(X,Y_i)$, where $f(X,Y_i)$ is a function of multiple reference measurements. The resulting measuring function, $X = f(R_s,Y_i)$, was applied to each sensor. The same pattern of calibration/validation sets as for linear regression was used for the multi linear regression.

Table 2
MLR models of single sensor.

Sensor's model	Multivariate linear model
O3B4	$O_3 = \frac{Rs - bNO_2 - cNO_2 \cdot H_2O - d}{a}$
O3_3E1F	$O_3 = \frac{Rs - bNO_2 - c}{a}$
NO2B4	$NO_2 = \frac{Rs - bO_3 - cT - dRH - e}{a}$
NO2_3E50	$NO_2 = \frac{Rs - bO_3 - cT - dRH - e}{a}$
MICS-2710	$NO_2 = \frac{Rs - bO_3 - cT - d}{a}$
MICS-4514	$NO_2 = \frac{Rs - bO_3 - cNO - dT - e}{a}$
CairClip NO2	$NO_2 = \frac{Rs - bO_3 - c}{a}$

3.3. Artificial neural network (ANN)

Artificial neural networks (ANN) are very sophisticated modelling techniques able to model extremely complex functions well suited for the calibration of a cluster of sensors. In this study, two types of ANN architectures were considered: radial based functions and multilayer perceptron (MLP). The former did not produce good results and is not presented hereafter. The latter is the most popular network architecture used today, due originally to Rumelhart and McClelland [36]. It consists of artificial units that receive a number of inputs (either from original data, or from the output of other units in the neural network) and typically one hidden layer with hidden units. The weighted sum of the inputs is formed to compose the activation of the unit. The activation signal is passed through an activation function to produce the output of the unit. With a defined number of layers and number of units in each layer, the network's weights and thresholds must be set in order to minimize the prediction error made by the network. This is the role of the training algorithms which uses iterative techniques called backpropagation. We used the BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm, the most recommended techniques for training [37] to automatically adjust the weights and thresholds in order to minimize this error. The error of a particular configuration of the network can be determined by running all the training cases through the network, comparing the actual output generated with the desired or target outputs. The differences are combined by an error function which gives the network error as a sum squared error, where the individual errors of output units on each case are squared and summed together.

Among the whole dataset, the first week of valid measurements was used for training (about 168 hourly values). This train period contains input and output variables assuring ten times as many cases as connections in the networks. When any sensor or reference measurements were missing the whole cases were discarded. The 2nd week of the measuring campaign was used as test dataset in order to limit the over-learning of the backpropagation algorithm. It was used to check progress against an independent dataset. As training progresses, the training error naturally dropped. However, when the selection error stopped dropping, or indeed started to rise indicating that the network was starting to over fit the data, the train was stopped. In this case, the number of hidden units was decreased.

The rest of the dataset (about 85% of data) was used as a validation set to ensure that the results on the testing and training set are real, and not artefacts of the training process. Moreover we iteratively conduct a number of experiments with each configuration, retaining the best networks in terms of error of the testing set since the validation dataset remains unknown at calibration time. In this way we also avoid being fooled if training locates a local minimum. Once we had experimentally determined an effective configuration

Table 3
Lists of possible and selected inputs for the different ANN.

	Possible parameters and sensors	Selected inputs after sensitivity analysis	Selected architectures of ANN
O_3	O_3: O3_3E1F; NO_2: MICS-2710 and MICS-4514-NO2, NO2_3E50; CO: MICS-4514-CO and TGS-5042; CO_2 sensors avoided because of correlation of O_3 and CO_2; absolute humidity	O3_3E1F, MICS-2710 and TGS-5042	Number of networks selected: 100 Number of hidden layer: 3 to 10 Hidden activation: exp, logistic, tanh Ouput activation: exp, identity, sine
NO_2	O_3: O3_3E1F; NO_2: NO2_3E50, MICS-2710, MICS-4514-NO2 and NO2B4_107; NO: NO_3E100; CO sensors avoided because of correlation of NO_2 and CO; CO_2: CO2_712; absolute humidity	NO2_3E50, MICS-4514-NO2, O3_3E1F, MICS-2710, MICS-4514-CO, NO2B4_107, absolute humidity	Number of networks selected: 100 Number of hidden layer: 4 to 12 Hidden activation: exp, logistic, tanh, sine Ouput activation: exp, identity, logistic, sine, tanh

for our networks, we resampled and generated new networks with that configuration.

In order to select input variables for the ANN, initially all sensors that were both correlated with the air pollutant of interest and independent between each other were selected. Between sensor pairs of the same brand model, we have selected the one giving the highest correlation with the pollutant of interest.

Then a sensitivity analysis was performed in order to discard sensors which were not significant for the NN architectures. The sensitivity analysis used the Sums of Squares residuals (SSR) of the model, computing the ratios of SSR for the full NN models out of SSR when the respective sensor was eliminated from the neural net. The parameters that were not found significant were discarded one at a time and the training was repeated until all parameters were found significant. As far as possible, we tried to avoid selecting meteorological sensor and reference measurements as input of ANNs in order to only rely on low-cost sensors.

Three studies have been performed by changing the input data: raw, standardized and calibrated by MLR sensors data. For the standardized values, the numeric data were scaled applying a z transformation with means of zero and standard deviation of 1. The output of ANN consisted in a set of at most 100 networks within 10,000 tested networks with different MLP architectures (see Table 3).

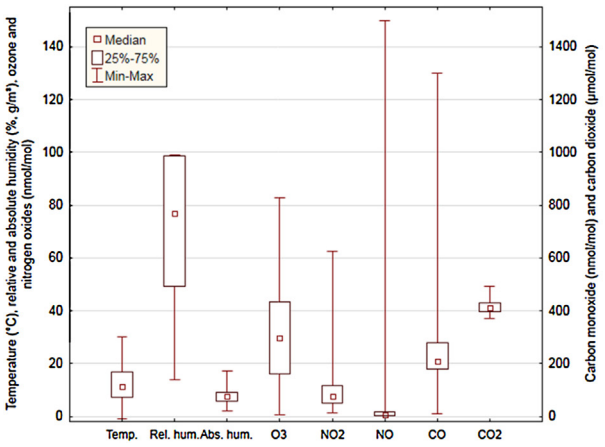
3.4. Evaluation of calibration method

The evaluation of sensor performances took into account hourly values. It was carried out using only values predicted by each calibration method. For each one, regression and difference-based analysis were conducted to evaluate their performance. These included the calculation of the coefficient of determination (R^2), comparing the slope and intercept of the regression line with objective values of 1 and 0 respectively. The mean bias error (MBE) and the root mean squared error (RMSE) standardized with the standard deviation of the reference measurements were used to draw a target diagram [38].

To assess the performance of each calibration method at individual air pollutant levels, we have also calculated the measurement uncertainty using orthogonal regression of the estimated outputs against reference data. This uncertainty was compared to the DQO for indicative method that corresponds to a relative expanded uncertainty of 30% for O_3 and 25% for NO_2 at the limit value set by the European Directive. The estimation method of the uncertainty, which corresponds to the relative expanded uncertainty U_r, was carried out using Eq. (1) where b_0 and b_1 are the slope and intercept of the orthogonal regression and RSS the sum of square of residuals is calculated using Eq. (2). The details of calculation of the orthogonal regression

Table 4
Performances of linear and multi-linear regression for calibration of single sensors. Results were observed on the validation set (the quoted values represent standard uncertainties).

Sensors	Calibration R^2	Validation R^2	Slope $\pm u$	Intercept $\pm u$	n
O3B4_20 (LR)	0.023	0.212	4.83 ± 0.05	−142.99 ± 1.97	2111
O3_3E1F (LR)	0.845	0.667	1.84 ± 0.02	−26.31 ± 0.86	2037
O3_3E1F (LR)	0.878	0.813	1.65 ± 0.02	−20.29 ± 0.62	1996
O3B4_20 (MLR)	0.503	0.479	2.33 ± 0.03	−37.97 ± 1.19	2070
O3_3E1F (MLR)	0.852	0.584	2.10 ± 0.03	−30.15 ± 1.05	1996
O3_3E1F (MLR)	0.945	0.824	1.49 ± 0.01	−12.03 ± 0.55	1955
NO2Cair1 (LR)	0.465	0.004	−60.94 ± 0.09	438.99 ± 0.74	2091
NO2Cair2 (LR)	0.240	0.036	−29.14 ± 0.12	229.86 ± 0.97	2091
NO2B4_107 (LR)	0.108	0.009	−73.28 ± 0.15	522.71 ± 1.23	2089
NO2B4_113 (LR)	0.230	0.002	−139.52 ± 0.13	991.24 ± 1.08	2089
NO2_3E50 (LR)	0.001	0.068	−247.53 ± 1.37	1727.00 ± 11.04	2055
NO2_3E50 (LR)	0.002	0.051	280.68 ± 1.36	−1977.94 ± 10.95	2053
MICS-2710 (LR)	0.206	0.131	8.13 ± 0.07	−53.83 ± 0.52	2089
MICS-2710 (LR)	0.200	0.126	8.48 ± 0.07	−56.45 ± 0.54	2089
MICS-4514-NO_2 (LR)	0.168	0.016	56.37 ± 0.16	−378.51 ± 1.24	2111
MICS-4514-NO_2 (LR)	0.269	0.203	6.18 ± 0.06	−40.00 ± 0.48	2089
NO2Cair1 (MLR)	0.745	0.021	16.17 ± 0.06	−97.84 ± 0.47	2091
NO2Cair2 (MLR)	0.585	0.004	32.50 ± 0.06	−205.92 ± 0.44	2091
NO2B4_107 (MLR)	0.351	0.026	17.07 ± 0.07	−108.36 ± 0.53	2048
NO2B4_113 (MLR)	0.679	0.086	6.96 ± 0.05	−33.59 ± 0.40	2048
NO2_3E50 (MLR)	0.768	0.078	10.97 ± 0.07	−78.28 ± 0.57	2014
NO2_3E50 (MLR)	0.562	0.062	16.01 ± 0.09	−108.18 ± 0.72	2012
MICS-2710 (MLR)	0.745	0.057	−477.45 ± 2.42	3176.44 ± 19.08	2098
MICS-2710 (MLR)	0.744	0.063	31407.64 ± 166.23	−207890.74 ± 1308.95	2098
MICS-4514-NO_2 (MLR)	0.525	0.010	577.84 ± 1.27	−3970.53 ± 10.01	2098
MICS-4514-NO_2 (MLR)	0.786	0.016	−852.21 ± 2.36	5692.90 ± 18.55	2098



	T	RH	AR	P	O_3	NO_2	NO	CO	CO_2
T	1.00								
RH	-0.60	1.00							
AH	0.35	0.51	1.00						
P	-0.06	-0.05	-0.12	1.00					
O_3	0.72	-0.72	-0.08	-0.15	1.00				
NO_2	-0.38	0.17	-0.17	0.36	-0.50	1.00			
NO	-0.16	0.04	-0.11	0.19	-0.32	0.52	1.00		
CO	-0.45	0.22	-0.25	0.45	-0.49	0.80	0.40	1.00	
CO_2	-0.61	0.62	0.05	0.28	-0.81	0.43	0.20	0.49	1.00

T: temperature, RH: relative humidity, AH: absolute humidity, P atmospheric pressure

Fig. 1. Box whisker plots and correlation table pattern in the reference measurements at the JRC-EMEP station.

can be found in the Guide for the demonstration of equivalence [39].

$$U_r(y_i) = \frac{2\left(\sqrt{\frac{\mathrm{RSS}}{(n-2)} - u^2(x_i) + [b_0 + (b_1 - 1)x_i]^2}\right)}{y_i} \quad (1)$$

$$\mathrm{RSS} = \sum (y_i - b_0 - b_1 x_i)^2 \quad (2)$$

Finally the drift over time of each calibration methods was plotted in order to evidence general trends. To ease the detection of possible patterns by filtering noise, the daily residuals were plotted between reference measurements and sensor predictions rather than the hourly ones.

4. Results

4.1. Presentation of dataset

The dataset was analyzed using descriptive statistics. Special attention was paid to outliers, data magnitude and variability in order to produce a full valid and complete dataset. The JRC EMEP station being a semi-rural site in a humid region, it shows high relative humidity, low air pollutant levels for NO and NO_2 and relatively high O_3 levels (see Fig. 1). The high value of CO and NO_2, respectively around 1.3 µmol/mol and 150 nmol/mol, are due to the provisional location of the mobile laboratory near to a railroad crossing.

An important aspect of the dataset is the lack of independence between parameters. Usually, O_3 is highly correlated with temperature and anti-correlated with relative humidity and CO_2 and to a lower extent with CO and NO_2. As a consequence, it will be difficult to estimate O_3 correctly using temperature, relative humidity and CO_2 as estimators. Unfortunately, it is well known that temperature

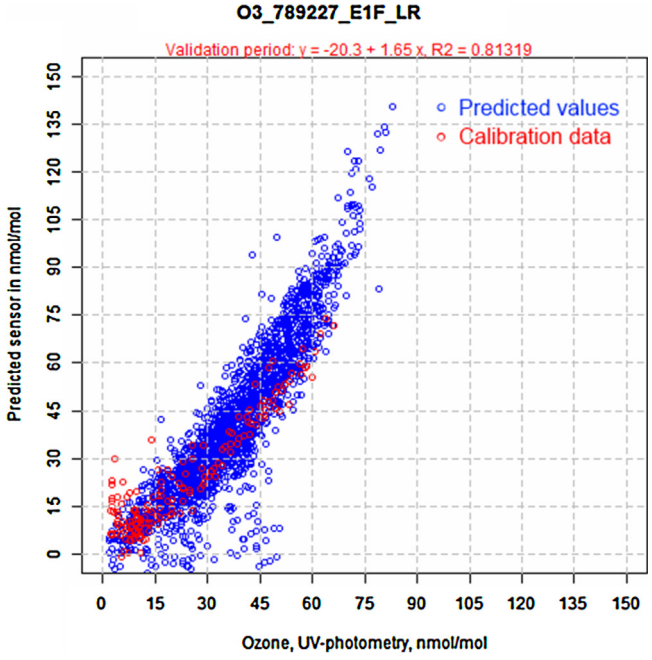


Fig. 2. Scatterplot of O3_3E1F calibrated sensor data using the linear regression against reference measurements. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

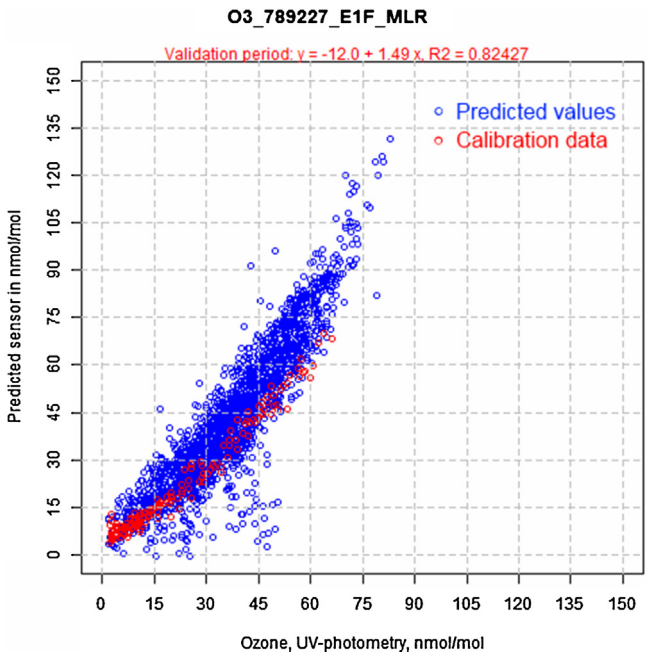


Fig. 3. Scatterplot of O3_3E1F calibrated sensor data using the MLR against reference measurements. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

Table 5
Performances of ANN calibrations. Results were observed on the validation set. The quoted values represent the standard uncertainties and n is the number of data used in the calculation.

Sensors	R^2	Slope $\pm u$	Intercept $\pm u$	n
O3 (ANN raw)	0.915	1.12 ± 0.01	−2.77 ± 0.29	1996
O3 (ANN std)	0.910	1.10 ± 0.01	−2.05 ± 0.29	2020
O3 (ANN MLR)	0.857	1.02 ± 0.01	1.36 ± 0.35	1979
NO_2 (ANN raw)	0.596	0.64 ± 0.01	5.81 ± 0.08	2064
NO_2 (ANN std)	0.560	0.70 ± 0.01	5.34 ± 0.09	2064
NO_2 (ANN MLR)	0.553	0.79 ± 0.01	6.50 ± 0.10	2064

and humidity are important factors affecting sensor responses. An alternative consists in using absolute humidity instead of temperature and relative humidity since absolute humidity is not correlated with O_3 nor CO_2. There is always a possibility that any statistical model able to correctly predict O_3 will in fact predict CO_2 and benefit of the high correlation of CO_2 to predict O_3. The same type of doubt exists for NO_2 and CO since they are highly correlated.

4.2. Results of calibration methods

Table 4 gives all the parameters for both linear regression (LR) and multi-linear regression (MLR) parameters for every single sensor. The regression analysis was performed using data of the first two weeks as calibration period. For each sensor, the measuring equation ($X=(R_S-B)/a$ or $X=f(R_S,Y_i)$) was applied to the validation dataset.

Results have shown that sensors from the same type are slightly different. However, they mainly stay within the same range. This shows that identical sensors tend to perform in a similar way even if some variance can be observed.

For example, Fig. 2 gives the scatterplot of the LR predicted sensor values versus the O_3 reference measurements for the 2nd O3_3E1F sensor. Red dots represent the values used during the calibration process and the blue ones represent the predicted data based on the validation data set. This sensor was selected because it showed the best correlation factor ($R^2=0.88$) during the calibration period. The scatterplot shows that the strength of association slightly decreased during the validation period similar to the calibration with $R^2=0.81$ compared to $R^2=0.88$. During the MACPoll project, it was observed that the O3_3E1F sensor was

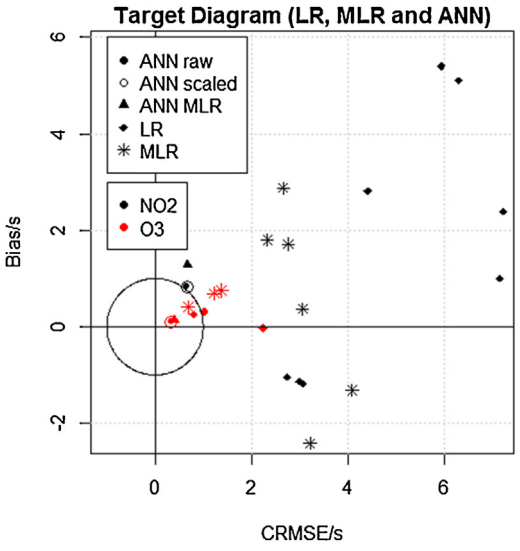


Fig. 4. Target diagram for ANNs with raw, scaled and modelled inputs, LR and MLR for NO_2 and O_3 calibration methods. A few NO_2 LR and MLR sensor fell outside the limits with values higher than 7. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

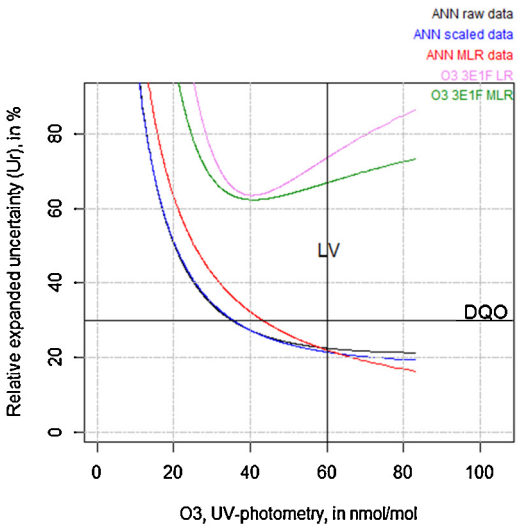


Fig. 5. U_r of the different calibration models versus reference data of O_3. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

not affected by temperature or humidity but it suffered from a cross-sensitivity to NO_2. As needed in the MLR model (see Table 2), the needed inputs, both gaseous and meteorological, have been selected within the reference measurements to maximize the benefits of the calibration. The constants a, b and c of the sensor models were fitted during the first two weeks of valid measurements. Subsequently, the equation was applied to the validation dataset. Fig. 3 gives the scatterplot of the orthogonal regression of the calibrated sensor data using the MLR method against the reference measurements. In this particular case, the use of NO_2 reference values improved R^2 from 0.88 to 0.95 for the calibration period.

Table 3 gives the lists of all inputs before the sensitivity analysis and the selected set of inputs. Table 5 gives the regression parameters fitted during the validation period. Results are given for the three types of input data: raw, standardized (std) and calibrated with MLR data (MLR). For the 3 types of data, the same list of inputs has been respected in order to be able to compare them. The difference observed in the number of data used for the calculation is mainly due to the manual validation of data performed in order to remove artefacts and wrong values.

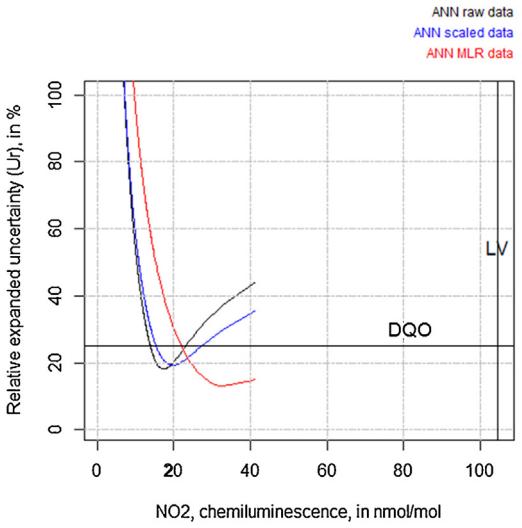


Fig. 6. U_r of the different calibration models versus reference data of NO_2. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

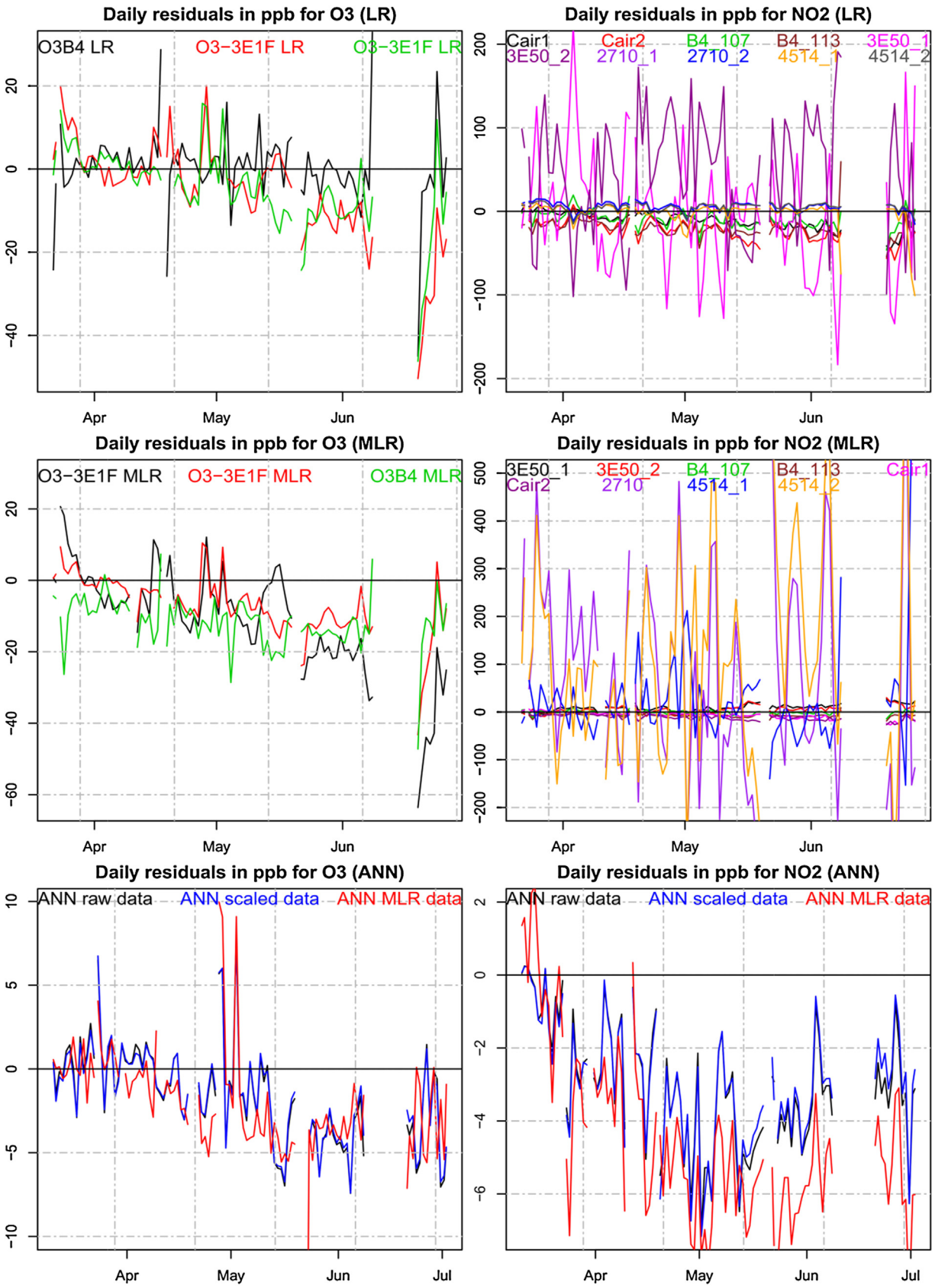


Fig. 7. Drifts of calibration methods for single sensors (at left for linear and multi-liner regression) and clusters of sensors (at right for artificial neural networks – ANN). (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

Fig. 4 gives the target diagram for LR, MLR and ANNs calibration methods for both gaseous species. This type of diagram [38] was used to evaluate the time trends of the sensor predictions (E) and the reference measurements (M). This kind of plot represents on the X-axis the centred root mean square error (CRMSE) normalized by the standard deviation of reference measurements (σ_0), which is an indicator of the modelled random error. The Y-axis represents the correlation of the coefficient R normalized by σ_0, which symbolizes the systematic bias. The distance between each point and the origin represents the root mean square error (RMSE). Finally, the circle area corresponds to the area of acceptance and stands for points where the model random error is equivalent to the variance of the observations. Data inside this circle indicate a positive correlation between modelled and observed values.

Based on Eq. (1), the relative expanded uncertainty (U_r) was plotted against O_3 and NO_2. Fig. 5 shows U_r versus O_3 reference measurements for the best O3_3E1F calibrated by LR, MLR and by ANN raw data, ANN scaled data or ANN MLR data. Fig. 6 shows U_r for NO_2. Only the plots of ANNs methods appear since the ones for LR and MLR fell outside the y-axis. Finally, Fig. 7 gives the times series of the O_3 and NO_2 residuals between reference measurements and sensor predictions using LR, MLR and ANNs calibration methods.

5. Discussion

Considering the best O_3 sensors, the coefficients of determination of the calibration dataset were high for LR calibration and slightly higher for the MLR method, both methods resulting in a high R^2 for the validation dataset. However in both cases, the slope (about 2) and intercept of the orthogonal regression were respectively different from 1 and 0. All these indicators are noticeably much better for the ANNs methods: higher R^2 for the validation dataset, slope nearer from 1 and intercept of a few nmol/mol.

Similarly, the target diagram shows that the ANNs result both in a lower bias (shown on the y-axis) and lower unbiased RMSE (called centred root-mean-square error, CRMSE, shown on the x-axis) than LR and MLR. Moreover, LR and MLR symbols fall generally outside the target circle, called efficiency score, evidencing RMSE up to 2 fold higher than the standard deviation of reference measurements.

The use of ANN for calibration purpose appears to be the most efficient in terms of uncertainties. In fact, the O_3 DQO is only reached with ANNs for concentrations higher than 35 nmol/mol, which is a lower level than the limited value of the Directive. Moreover, the ANNs based on modelled inputs performed slightly worse at low level and reach the 30% DQO at about 45 nmol/mol. The ANNs based on raw and standardized inputs showed identical U_r plots. Consequently, the easier ANNs inputs based on raw or standardized ANNs should be preferred. The U_r plots for LR and MLR show higher values exceeding the DQO. These plots also show a positive trend towards high O_3, showing the effect of large slopes of the orthogonal regression (see Table 4). Fig. 7 gives evidences of a slight drift of the calibration methods over time of about 5 nmol/mol over nearly 4 months for ANNs, and about 20 nmol/mol for LR and MLR. While the ANNs with the raw, scaled and MLR input results in similar drifts and constant noise, MLR showed slightly higher drift and noise than LR.

It should be noted that in Figs. 2 and 3 we have observed a slight overestimation of the predicted values. Actually, the main error is due to the extrapolation of data higher than the maximum value observed in the calibration dataset. Both methods are suffering from a lack of sensitivity regarding interfering effect such as temperature and relative humidity.

For NO_2, none of the sensors has given a high R^2 for LR methods apart from one CairClip NO_2 with the calibration dataset. Higher values where reached for MLR up to 0.75 for some sensors both of MO_x and electrochemical types. Unfortunately the R^2 was very low resulting in slop and intercept far from 1 and 0 respectively. It is likely that NO_2 sensors at a semi-rural site are affected by the low NO_2 levels and high correlation between O_3 and NO_2 to which the sensor are generally sensitive. These observations are corroborated by the target diagram which shows that none of the NO_2 sensors are within the efficiency score, with LR and MLR calibration being much higher than the ANNs methods. However, within ANNs, as for O_3 the raw and scaled inputs resulted in lower biases than ANN-MLR.

No high NO_2 was measured during the measuring campaign, making it difficult to correctly apply the calibration methods. For LR and MLR, U_r was too high to be visible within Fig. 6. However, using ANNs, an interesting U_r of around 20% was reached, this value increasing for NO_2 higher than 20 nmol/mol for the raw and scaled inputs. This behaviour was not observed with the ANN-MLR which appeared to remain rather constant. One shall remember that implementing the ANN-MLR requires a set of 7 sensors, of which 2 NO_2 MO_x and 2 NO_2 electrochemical sensors, 1 O_3 electrochemical sensor, 1 CO electrochemical sensor and absolute humidity (therefore temperature and relative humidity sensor). Moreover, all gas sensors were previously calibrated using correction models (Table 2) including reference measurements for O_3.

Finally, Fig. 7 shows that LR and MLR methods appear to be without drift over time. Nevertheless, enormous noise in particular for MO_x sensors can be observed. As for O_3, the ANNs methods appears to suffer from a low drift of about 4 nmol/mol in about 4 months for raw and scaled inputs and 6 nmol/mol for MLR inputs.

Based on the requirement of the European Air Quality Directive for indicative methods, calibration would have been performed only if the uncertainty had exceeded the Data Quality Objective (DQOs). For O_3 the DQO corresponds to an uncertainty of 30% at the limit value of 60 nmol/mol, which means 18 nmol/mol. For NO_2, the DQO is 25 nmol/mol, which represents 25% of uncertainty at the limit value of 100 nmol/mol. For both O_3 and NO_2, ANN method shows a maximum drift on residuals of 6 nmol/mol, three times lower than the DQO of O_3.

6. Conclusions

Based on the measurement uncertainty estimated by orthogonal regressions of the sensor outputs versus reference data, the most suitable calibration method appeared to be ANN using raw or scaled sensor inputs. Simple LR and MLR have shown to produce the highest measurement uncertainty. While ANN with MLR inputs needed reference data for calibration of most sensors, ANN with raw/scaled data, using only 3 sensors of different types (1 O_3 chemical, 1 NO_2 resistive sensor and 1 CO electrochemical sensor), were able to solve the main interferences of the O_3 sensor.

In general, it was shown that the ANN method increased the strength of association between estimated and reference data (higher R^2 and lower CRMSE). Moreover, it also allowed the decrease of the bias to reference data, with the slope and intercept of orthogonal regression being respectively nearer to 1 and 0.

It is likely that by combining different type of sensors, like electrochemical O_3 and NO_2 MO_x sensors for example, the ANN can solve the cross sensitivity issues from which suffers the major part of sensors. We have also observed that the humidity/temperature dependence was also corrected, without the needs of such measurements. We suppose that it is linked with the difference of influence of these parameters on both types of sensors. Finally, we showed that using a cluster of sensors for calibration purpose, the data quality objectives of the European Directive for indicative methods could be met for O_3 (uncertainty, U_r, of 30%) at semi-rural stations. Formal conclusions on the possibility to meet the

DQO for NO_2 should be evaluated at higher levels typical for urban environments.

Acknowledgements

The authors wish to acknowledge the collaboration of our JRC colleagues C. Grüning and G. Manca for their contribution with CO_2 measurements and F. Lagler, N.R. Jensen and A. Dell'Acqua for carrying out air pollution measurements. This study was carried out within the EMRP Joint Research Project ENV01 MACPoll. The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

References

[1] Directive 2008/50/EC of the European Parliament and the Council of 21 May 2008 on ambient air quality and cleaner air for Europe.
[2] H. Nakagawa, S. Okazaki, S. Asakura, K. Fukuda, H. Akimoto, S. Takahashi, S. Shigemori, An automated car ventilation system, Sens. Actuators B: Chem. 65 (June (1–3)) (2000) 133–137.
[3] M.C. Carotta, G. Martinelli, L. Crema, M. Gallana, M. Merli, G. Ghiotti, E. Traversa, Array of thick film sensors for atmospheric pollutant monitoring, Sens. Actuators B: Chem. 68 (1) (2000) 1–8.
[4] C. Pijolat, C. Pupier, M. Sauvan, G. Tournier, R. Lalauze, Gas detection for automotive pollution control, Sens. Actuators B: Chem. 59 (October (2–3)) (1999) 195–202.
[5] M. Lösch, M. Baumbach, A. Schütze, Ozone detection in the ppb-range with improved stability and reduced cross sensitivity", Sens. Actuators B: Chem. 130 (1) (2008) 367–373.
[6] A. Gramm, A. Schütze, High performance solvent vapor identification with a two sensor array using temperature cycling and pattern classification, Sens. Actuators B: Chem. 95 (October (1–3)) (2003) 58–65.
[7] M. Kamionka, P. Breuil, C. Pijolat, Calibration of a multivariate gas sensing device for atmospheric pollution measurement, Sens. Actuators B: Chem. 118 (October (1–2)) (2006) 323–327.
[8] M. Kamionka, P. Breuil, C. Pijolat, Atmospheric pollution measurement with a multi-materials sensing device, Mater. Sci. Eng. C 26 (March (2–3)) (2006) 290–296.
[9] D.B. Reza Nadafi, S.N. Nejad, M. Kabganian, F. Barazandeh, Neural network calibration of a semiconductor metal oxide micro smell sensor, in: 2010 Symposium on Design Test Integration and Packaging of MEMS/MOEMS (DTIP), 2010, pp. 154–157.
[10] M. Baumbach, A. Sossong, H. Delprat, K. Soulantica, A. Schuetze, A. Maisonnat, B. Chaudret, New micro machined gas sensors combined with intelligent signal processing allowing fast gas identification after power-up, Proc. Sens. 2 (2005) 91–96.
[11] G. Huyberechts, P. Szecówka, J. Roggen, B.W. Licznerski, Simultaneous quantification of carbon monoxide and methane in humid air using a sensor array and an artificial neural network, Sens. Actuators B: Chem. 45 (December (2)) (1997) 123–130.
[12] S. De Vito, G. Di Francia, L. Martinotto, Neural calibration of portable multisensor device for urban atmospheric pollution measurement, in: Sensors and Microsystems, World Scientific, 2008, pp. 283–290.
[13] K. Brudzewski, S. Osowski, Gas analysis system composed of a solid-state sensor array and hybrid neural network structure", Sens. Actuators B: Chem. 55 (1) (1999) 38–46.
[14] S. De Vito, E. Massera, M. Piga, L. Martinotto, G. Di Francia, On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario, Sens. Actuators B: Chem. 129 (February (2)) (2008) 750–757.
[15] L. Spinelle, M. Aleixandre, M. Gerboles, Comparison between the performances of commercially available low cost sensors for the monitoring of O_3 and NO_2 in ambient air, Unpublished.
[16] D. Hasenfratz, O. Saukh, L. Thiele, On-the-fly calibration of low-cost gas sensors, in: G.P. Picco, W. Heinzelman (Eds.), Wireless Sensor Networks, Springer, Berlin, Heidelberg, 2012, pp. 228–244.
[17] http://macpoll.eu
[18] L. Spinelle, M. Gerboles, M. Aleixandre, Report of Laboratory and In-situ Validation of Micro-sensor for Monitoring Ambient O12: CairClipO3/NO2 of CAIRPOL (F), Publications Office of the European Union, Luxembourg, 2013, EUR 26373.
[19] L. Spinelle, M. Gerboles, M. Aleixandre, Report of Laboratory and In-situ Validation of Micro-sensor for Monitoring Ambient Air Pollution – NO9: CairClipNO2 of CAIRPOL (F), Publications Office of the European Union, Luxembourg, 2013, EUR 26394.
[20] L. Spinelle, M. Gerboles, M. Aleixandre, Report of Laboratory and In-situ Validation of Micro-sensor for Monitoring Ambient Air – Ozone Micro-sensors, αSense, Model B4 O3 sensors, Publications Office of the European Union, Luxembourg, 2013, EUR 26681.
[21] L. Spinelle, M. Gerboles, M. Aleixandre, Report of laboratory and in-situ validation of micro-sensor for monitoring ambient air – NO2: NO2-B4 sensor of αSense (UK), unpublished.
[22] L. Spinelle, M. Gerboles, M. Aleixandre, Protocol of evaluation and calibration of low-cost gas sensors for the monitoring of air pollution, Publications Office of the European Union, Luxembourg, 2013, EUR 26112.
[23] Technical Data Sheet CairClip NO2 (preliminary version). http://www.cairpol.com/index.php?option=com_content&view=article&id=72&Itemid=151&lang=en
[24] Ozone Sensor, Citytech O3 3E 1 F, Rev. 11/2011, Data sheet, http://www.citytech.com/PDF-Datasheets/o33e1f.pdf
[25] Nitrogen Dioxide Sensor, Citytech NO2 3E 50, Rev. 11/2011, Data sheet, http://www.citytech.com/PDF-Datasheets/no23e50.pdf
[26] Nitric Oxide Sensor, Citytech NO 3E 100, Rev. 11/2011, Data sheet, http://www.citytech.com/PDF-Datasheets/no3e100.pdf
[27] O3B4 Ozone Sensor 4-Electrode, data sheet, http://www.alphasense.com/WEB1213/wp-content/uploads/2013/11/O3B4.pdf
[28] NO2B4 Nitrogen Dioxide, 4_Electrode, Data sheet, http://www.alphasense.com/WEB1213/wp-content/uploads/2013/11/NO2B4.pdf
[29] http://www.alphasense.com/WEB1213/wp-content/uploads/2014/02/ISB.pdf
[30] http://www.cdiweb.com/datasheets/e2v/mics-2710.pdf
[31] http://www.sgxsensortech.com/content/uploads/2014/08/MICS-%E2%80%93-4514-Carbon-Monoxide-+-Nitrogen-Dioxide-dual-sensor.pdf
[32] http://www.sgxsensortech.com/content/uploads/2014/07/MICS-EK1-Evaluation-Kit.pdf
[33] http://www.figarosensor.com/products/5042pdf.pdf
[34] http://www.figaro.co.jp/en/product/docs/com5042_product_information_rev03.pdf
[35] http://tccelt.co.kr/2012/kor/pdf/S-100/DS_S-100_Rev2.7.pdf
[36] D.E. Rumelhart, J.L. McClelland, P.D.P. Research Group (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, MIT Press, Cambridge, MA, USA, 1986.
[37] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, Inc, New York, NY, USA, 1995.
[38] J.K. Jolliff, J.C. Kindle, I. Shulman, B. Penta, M.A.M. Friedrichs, R. Helber, R.A. Arnone, Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, J. Mar. Syst. 76 (February (1–2)) (2009) 64–82.
[39] Guide to the demonstration of equivalence of ambient air monitoring methods, report by an EC Working Group on Guidance.

Biographies

Laurent Spinelle is graduated of a Ph.D. in materials for electronics. During his study, he acquired a wide experience in the analysis and characterization of inorganic materials. He was entrusted with the development of a process making possible the selective separation of two oxidizing gaseous species. From the elaboration of a gas sensor microsystem, he is now involved in the development of a validation protocol for commercial sensors in the Institute for Environment and Sustainability at the Joint Research Centre. This work gave him the opportunity to increase his knowledge in Environmental Science and acquire skills in statistical analysis and uncertainties.

Michel Gerboles studied Analytical Chemistry at the University of Bordeaux (FR). In 1990, he joined the European Reference Laboratory for Air Pollution at the Joint Research Centre of the European Commission. His initial activities were focused on the traceability of measurement techniques for inorganic air pollutants, primary reference methods for air quality monitoring and the organization of proficiency testing/intercomparison exercises in support to the European policy on air quality. His activities moved to the speciation of PM (ions and heavy metals) and the evaluation of indicative measurement methods like diffusive sampling. His current research interests include the validation of low-cost gas sensors, and the development of automatic methods for the validation and extraction of quality indicators from large ambient air monitoring datasets.

Maria Gabriella Villani received her Laurea in Physics in 1998 at Padua University (Italy), Master Degree in Geography–Meteorology, 2002, at Indiana University, Bloomington Indiana (USA), and her Ph.D. in Atmospheric Sciences, 2006, at Carlo Bo University, Urbino (Italy). She worked as research assistant at the National Research Council, Institute of Atmospheric Sciences and Climate, and, as a Postdoc, at the Joint Research Centre, Institute of Sustainability, in the Climate Change Unit. Currently, she is a researcher in ENEA, Italian National Agency for New Technologies, Energy and Sustainable Development, UTTEI-SISP Unit.

Manuel Aleixandre received his B.Sc. in physics from the Universidad Autonoma of Madrid, Spain, in 1999 and his Ph.D. degree from the Universidad Nacional de Educacion a Distancia in 2007. His research interests include neural networks, fuzzy logic, statistics and multivariate data analysis, calibrations, gas sensors based in fibre optic sensors, instrumentation and development of electronic noses for several applications.

Fausto Bonavitacola is an expert in data acquisition, control and simulation devices, sensors and measure instrumentation. He holds a degree in "Electronic engineering" with a specialization in "devices and circuits". He has been involved for the last 15 years as an IT expert and/or an analyst/programmer in many projects and/or research activities carried out in various laboratories at the Joint Research Centre in Ispra.