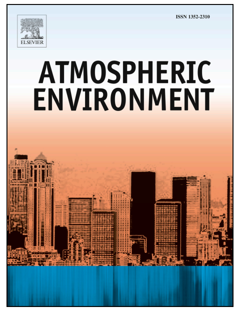
Accepted Manuscript



Assessment of air quality microsensors *versus* reference methods: The EuNetAir Joint Exercise – Part II

C. Borrego, J. Ginja, M. Coutinho, C. Ribeiro, K. Karatzas, Th Sioumis, N. Katsifarakis, K. Konstantinidis, S. De Vito, E. Esposito, M. Salvato, P. Smith, N. André, P. Gérard, L.A. Francis, N. Castell, P. Schneider, M. Viana, M.C. Minguillón, W. Reimringer, R.P. Otjes, O. von Sicard, R. Pohle, B. Elen, D. Suriano, V. Pfister, M. Prato, S. Dipinto, M. Penza

PII: S1352-2310(18)30543-0

DOI: 10.1016/j.atmosenv.2018.08.028

Reference: AEA 16196

To appear in: *Atmospheric Environment*

Received Date: 27 December 2017

Revised Date: 6 August 2018

Accepted Date: 12 August 2018

Please cite this article as: Borrego, C., Ginja, J., Coutinho, M., Ribeiro, C., Karatzas, K., Sioumis, T., Katsifarakis, N., Konstantinidis, K., De Vito, S., Esposito, E., Salvato, M., Smith, P., André, N., Gérard, P., Francis, L.A., Castell, N., Schneider, P., Viana, M., Minguillón, M.C., Reimringer, W., Otjes, R.P., von Sicard, O., Pohle, R., Elen, B., Suriano, D., Pfister, V., Prato, M., Dipinto, S., Penza, M., Assessment of air quality microsensors *versus* reference methods: The EuNetAir Joint Exercise – Part II, *Atmospheric Environment* (2018), doi: 10.1016/j.atmosenv.2018.08.028.

Assessment of Air Quality Microsensors *Versus* Reference Methods: the EuNetAir Joint Exercise – part II

C. Borrego[a,b], J. Ginja[a], M. Coutinho[a], C. Ribeiro[a], K. Karatzas[c], Th. Sioumis[c], N. Katsifarakis[c], K. Konstantinidis[c], S. De Vito[d], E. Esposito[d], M. Salvato[d], P. Smith[e] *, N. André[f], P. Gérard[f], L. A. Francis[f], N. Castell[g], P. Schneider[g], M. Viana[h], M.C. Minguillón[h], W. Reimringer[i], R.P. Otjes[j], O. von Sicard[k], R. Pohle[k], B. Elen[l], D. Suriano[m], V. Pfister[m], M. Prato[m], S. Dipinto[m], M. Penza[m]

[a] *IDAD – Institute of Environment and Development, Campus Universitário, 3810-193 Aveiro, Portugal*

[b] CESAM & *Department of Environment and Planning, University of Aveiro, 3810-193 Aveiro, Portugal*

[c] *Department of Mechanical Engineering, Aristotle University, 54124 Thessaloniki, Greece*

[d] *Smart Networks and Photovoltaic Division, ENEA, C.R. Portici, 80055 Portici (NA), Italy*

[e] *Department of Chemistry, University of Cambridge, UK* now at IQFR-CSIC, Calle de Serrano 119, Madrid 28004, Spain*

[f] *Institute of Information and Communication Technologies, Université Catholique de Louvain, Belgium*

[g] *NILU Norwegian Institute for Air Research, Instituttveien 18, 2027 Kjeller, Norway*

[h] *IDAEA-CSIC, Spanish National Research Council, Jordi Girona 18, 08034 Barcelona, Spain*

[i] *3S – Sensors, Signal Processing, Systems GmbH, 66121 Saarbruecken, Germany*

[j] *ECN – Energy Research Center of the Netherlands, Petten, Netherlands*

[k] *Siemens AG, Corporate Technology, Germany*

[l] *VITO – Vlaamse Instelling voor Technologisch Onderzoek, Mol, Belgium*

[m] *ENEA, Laboratory of Functional Materials and Technologies for Sustainable Applications, 72100 Brindisi, Italy*

Abstract

The EuNetAir Joint Exercise focused on the evaluation and assessment of environmental gaseous, particulate matter (PM) and meteorological microsensors versus standard air quality reference methods through an experimental urban air quality monitoring campaign. This work presents the second part of the results, including evaluation of parameter dependencies, measurement uncertainty of sensors and the use of machine learning approaches to improve the abilities and limitations of sensors. The results confirm that the microsensor platforms, supported by post processing and data modelling tools, have considerable potential in new strategies for air quality control. In terms of pollutants, improved correlations were obtained between sensors and reference methods through calibration with machine learning techniques for CO (r^2=0.13-0.83), NO_2 (r^2=0.24-0.93), O_3 (r^2=0.22-0.84), PM10 (r^2=0.54-0.83), PM2.5 (r^2=0.33-0.40) and SO_2 (r^2=0.49-0.84). Additionally, the analysis performed suggests the possibility of compliance with the data quality objectives (DQO) defined by the European Air Quality Directive (2008/50/EC) for indicative measurements.

Keywords: Air quality monitoring; Reference methods; Low-cost microsensors; Experimental campaign; Measurement uncertainty; Machine learning

1. Introduction

Air pollution is a very significant environmental and social issue. At the same time, it is a complex problem posing multiple challenges in terms of management and mitigation of harmful pollutants. Air pollutants have numerous impacts on health, ecosystems, built environment and climate; they may be transported or formed over long distances, and they may affect large areas. Air pollution continues to affect the health of Europeans, particularly in urban areas. It also has considerable economic impacts; cutting lives short, increasing medical costs and reducing productivity through working days lost across the economy (EEA, 2017; WHO, 2018). According to The World Health Organization (WHO), in 2016, 91% of the world population were living in places where the WHO air quality guidelines limits were not met. Additionally, outdoor air pollution in both cities and rural areas was estimated to cause 4.2 million premature deaths worldwide in 2016 (WHO, 2018).

Europe's most problematic pollutants in terms of health are PM, NO_2 and ground-level O_3. In 2015, about 7 % of the EU-28 urban population was exposed to PM2.5 levels above the EU's annual limit value. Considering the stricter WHO guidelines, approximately 82 % were exposed to levels exceeding the limit values. Exposure to PM2.5 caused the premature death of estimated 428 000 people in 41 European countries in 2014. Regarding NO_2, around 9 % of the EU-28 urban population was exposed to levels above the EU's annual limit value and WHO guidelines in 2015. Exposure to NO_2 caused the premature death of an estimated 78 000 people in 41 European countries in 2014. For O_3 levels, 30 % of the EU-28 urban population was exposed to concentrations above the EU's target value in 2015. Considering the stricter WHO guidelines, approximately 95 % were exposed to levels exceeding the limit value. Exposure to O_3 caused the premature death of an estimated 14400 people in Europe in 2014 (EEA, 2017).

A wide range of adverse effects of ambient air pollution on health has been well documented in multiple studies (Pascal et al., 2013; Wu et al. 2016). By working to reduce air pollution levels, countries can lower the burden of stroke, heart disease, lung cancer, and both chronic and acute respiratory illness, with long-term benefits to the population (WHO, 2018).

However, there is significant inequality in exposure to air pollution and related health risks: air pollution combines with other aspects of the social and physical environment to create a disproportionate disease burden in less affluent parts of society. The WHO guidelines address all regions of the world and provide uniform targets for air quality that would protect the vast majority of individuals from the adverse effects on health. More than 80% of the population in the WHO European Region (including the European Union, EU) lives in cities with levels of PM exceeding WHO Air Quality Guidelines. Since even at relatively low concentrations the burden of air pollution on health is significant, effective management of air quality that aims to achieve WHO Air Quality Guidelines levels is necessary to reduce these risks to a minimum. Exposure to air pollutants is largely beyond the control of individuals, requiring action by public authorities at the national, regional and international levels. A multisector approach, engaging transport, housing, energy production and industry is needed to develop and effectively implement long-term policies that reduce the risks of air pollution to health (WHO, 2013).

The evaluation of the status of air quality (AQ) is based on ambient air measurements, in conjunction with data on anthropogenic emissions and their trends. Holistic solutions must be found that involve technological development, and structural and behavioural changes. Air quality policies have delivered, and continue to deliver, many improvements. However, substantial challenges remain and considerable impacts on human health and on the environment persist (EEA, 2017).

The increasing availability of low cost sensors employing various monitoring principles creates the need to identify which are the most appropriate to be further refined and calibrated. For this purpose, it is necessary to estimate the overall performance of a large number of collocated sensors (Kotsev et al., 2016). This not only calls for the application of standard time series analysis and comparison methods, but also the incorporation of overall measurement profile and behaviour, so that a sensor network may generate reliable AQ data. Calibration methods have been based on sensor intercomparison, as well as on overall sensor network calibration (Jiao et al., 2016) and self-calibration (Fishbain and Moreno-Centeno, 2016). Recent studies indicate that machine learning technologies may significantly improve the performance of air quality sensor nodes reducing the impact of cross-sensitivity issues (Spinelle et al., 2015; De Vito et al., 2018). Most of them however, have been carried out on single systems, developed by the same company or research institutions, limiting the understanding and potential about their general applicability.

In the first part of this work, the overall results of an intercomparison of AQ microsensors with reference methods during an AQ monitoring campaign in Aveiro, Portugal, were presented (Borrego et al., 2016). The overall performance of the diverse sensors in terms of their statistical metrics and measurement profile indicated significant differences in the results. In terms of pollutants, the following results were observed: O_3 (r^2: 0.12-0.77), CO (r^2: 0.53-0.87) and NO_2 (r^2: 0.02-0.89) with some promising results, but equally sensors showing no correlation with the reference method. For PM (r^2: 0.07-0.36) and SO_2 (r^2: 0.09-0.20) the results showed a poor performance with low correlation coefficients between the reference and microsensor measurements.

The purpose of this study is to present the second part of the results of the intercomparison campaign in Aveiro for two weeks in October 2014, complementing the analysis performed by Borrego et al. (2016). More specifically, it is intended to (a) understand parameter dependencies, (b) measurement uncertainty of sensors, (c) the use of machine learning approaches to improve the abilities and limitations of sensors, contributing to their calibration and further development.

The paper is organized into the following sections: Section 2 gives a description of the experimental campaign and methodology; Section 3 presents the results obtained with the different data analysis strategies; finally, Section 4 provides the conclusions.

2. Experimental Design

2.1. Characterization of the study site

In this exercise, the AQ microsensor systems were installed side-by-side on the IDAD Air Quality Mobile Laboratory (LabQAr), supplied with standard equipment and reference analysers for CO (Infrared photometry), NOx (Chemiluminescence), O_3 (Ultraviolet photometry), SO_2 (Ultraviolet fluorescence), particulate matter PM10 / PM2.5 (Beta-ray absorption), and meteorological variables (Vaisala WXT520). During the exercise, LabQAr was parked on Avenue Santa Joana, near the Cathedral of Aveiro, in an urban traffic location in Aveiro city centre. The sensors were mainly installed between 2.5 and 3 m above ground on the roof of the mobile laboratory, with the reference meteorological measurements at ~ 5 m on a telescopic mast (Fig. 1).

Fig. 1. Set-up of the AQ mobile station and micro-sensors during the 1st EuNetAir campaign.

2.2. Comparison of technical requirements of the sensor nodes

Aside from the performance of the sensor nodes with regard to comparability with reference instruments, discussed in Borrego et al. (2016), the selection of a specific sensor node should also consider a number of technical parameters as well as data quality and uncertainty.

These include; size, power, connectivity requirements, and number of pollutants monitored (Table 1).

Table 1. Technical requirements and features of the sensor nodes evaluated during the Aveiro intercomparison campaign.

Sensor node	**AQMesh**	**SNAQ**	**Prototype Module**	**NanoEnvi**	**ECN airbox**	**EveryAware SensorBox**	**OdorCheckerOutdoor**	**AUTh-ISAG**	**Prototype Module (with IAQcore)**	**Air-sensor box**
Operated by	**IDAEA-CSIC + AQMesh**	**Cambridge Univ.**	**UCL/CCS**	**NILU + Envira**	**ECN**	**VITO**	**3S**	**AUTh**	**Siemens AG**	**ENEA**
Parameters measured	NO, NO_2, CO, O_3	NO, NO_2, CO, CO_2, O_3, VOC, PM10	T, RH	CO, NO_2, O_3, T, RH	PM2.5, PM10, NO_2	NOx, CO, O_3, VOC, gasoline/diesel exhaust fumes, T, RH	T, RH, VOC	T, RH, p, NO_2, O_3	CO/VOC (nonspecific but calibrated to CO)	CO, NO_2, O_3, SO_2, PM10, T, RH
Time resolution (and ability to modify it, Y/N)	15 min (Y)	20 sec (Y)	1 sec (Y)	5 min; Yes (5, 10, 15, 30 and 60 min)	10 min (Y)	1 s (N)	3 min (Y)	5 min (Y)	10 s (Y)	15 min (Y)
Able to operate while connected to power (Y/N)	No; no power required	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
May operate on battery (Y/N, and duration if Yes)	Yes; <24 months	Yes; >1 month	Yes (external batterypack)	Yes; 2 days	Yes	Yes (external batterypack)	No	Yes , <10 hours, increased with extended battery	Yes (<5h, powered by connected notebook)	No
Able to store data internally (Y/N, max duration with default time resolution)	Yes	Yes	No	Yes; >1 year	Yes; one month	Yes; 6 months	Yes, > 2 months	Yes , > 1 year	No	Yes; > 1 year
Online data transmission possible (Y/N)	Yes	Yes	Yes	Yes	Yes	Yes (with smartphone and 'Airprobe' app)	Yes, modem under development	Yes	No	Yes
If online data transmission possible, how? Wifi, SIM card, etc.	SIM card	SIM card	Zigbee optional	SIM card, Ethernet, Zigbee	SIM card	Bluetooth communication with smartphone	SIM w/ GPRS/UMTS	SIM card, Xbee wireless module	---	SIM card, Wi-Fi, Ethernet
Able to be deployed outdoors (waterproof) (Y/N)	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes
Frequency of maintenance required (inlet cleaning, etc.)	No maintenance	No maintenance	No maintenance	Biannual calibration	Yearly	No maintenance	~ 2 months	No maintenance	No maintenance	Yearly
Other features	Standard sun roof	Meteo variables measured as well	Highly modular platform	Solar battery available	Other parameters also possible	Open source hardware and software	Modular setup: other sensors attachable (EC, wind, …)	Open source + commercial hardware and software	Commercial hardware (IAQcore)	Modular platform with other optional gas sensors

Among the sensor nodes deployed in this work, the number of parameters monitored ranged between one and seven, also covering meteorological variables. Time resolution ranged from 1-second up to 15-minute averages, suggesting different potential applications. For example, the 1-second time resolution of the EveryAware sensor box is well suited to personal exposure, while the 15-minute averages produced by the AQMesh and ENEA nodes would be more representative of ambient pollutant concentrations. However, this does not preclude the different sensor nodes being deployed for the same application, as was the case in this exercise.

One of the goals for widespread use of sensor technologies for air quality monitoring is to maximise spatial data coverage (Castell et al., 2013; Schneider et al., 2017), by deploying dense networks of sensors, e.g. across urban areas. However, this is not feasible if the nodes present limitations with regard to connectivity or power requirements. As shown in the Table 1 most of the nodes (7 out of 8) are able to operate on battery, minimising the need to provide mains power at the measurement locations. Also, all of them are able to transmit data, either directly to “cloud” platforms or through apps on mobile phones. Only one of the nodes requires Wi-Fi access, which is a potentially limiting factor for large scale deployment across urban areas. Finally, 7 out of the 8 nodes may be deployed outdoors (only one of them requiring additional protection to be built), being resistant to weather conditions. The latter requirement would be irrelevant when dealing with indoor air quality monitoring.

For outdoor air quality monitoring, the frequency of maintenance is a relevant parameter. Whereas most of the nodes require no or limited (yearly) maintenance, one node requires maintenance or calibration every 2-6 months. In terms of operation and data availability, frequent maintenance may be an issue for sensor node selection.

2.3. Calibration strategies

Some of the installed AQ microsensors output results in concentration units while others provide voltage or frequency data. Therefore, a pre-processing of raw data was necessary to proceed to concentration units. All sensor nodes (with the exception of the Siemens node) have hence been pre-calibrated. Each team was responsible for their own unit conversion, including different calibrations and conversion strategies, depending on the sensors used. Additional information is presented in Supplementary Material (Table S1).

The gas sensors used for the AUTh-ISAG node were off-the-shelf metal-oxide sensors that did not undergo a specialized calibration procedure by their manufacturer. In this case it was decided to apply a simple signal correction procedure in order to calibrate the readings received, assuming a linear calibration approach as suggested by Balzano and Nowak (2008).

The NanoEnvi node did not undergo a specialized calibration procedure by their manufacturer. The data was not post-processed to correct for temperature and humidity effects or cross-interference with other gases. For the analysis, only the negative concentration values were removed. Negative values were only registered for the NO_2 sensors and represented about 20% of the total data.

The two ECN-Airbox were calibrated before the Aveiro campaign carrying out co-located measurements with reference equipment at an official monitoring station in Amsterdam (NL49014 GGD Vondelpark). The ECN sensors were developed to minimize cross contamination and meteorology interference. For the NO_2 sensor this was established by introduction of a differential measurement technique enforced by a pre-processing step prior to the sensor by switching frequently to zero ambient airflow (NO_2 removed). Moreover, the sample flow was stabilized by a patented RH delaying cartridge.

The AQMesh v 4.0 pods used in this experiment reported NO, NO_2, CO and O_3 concentrations which are the result of a two-stage process, following the AQMesh standard operating procedure (SOP). In the first stage, the AQMesh algorithm (a fixed mathematical formula which does not use machine learning) is applied to the raw data in counts, which are converted to ambient concentration units, along with compensation for various environmental effects upon the sensors, providing precision of measurement. The AQMesh SOP requires that pods are deployed 2 weeks prior to the actual measurements for stabilization and application of scaling using reference data. Given that the logistics did not allow doing so in this field campaign, the calibration via scaling was done in this study using the reference data from the trial period.

The EveryAware SensorBox (EA SB) is a portable, low-cost measurement device, which allows measurements of the personal exposure to traffic pollution. The EA SB combines a number of low-cost electrochemical and metal oxide sensors to measure concentrations of CO, NO_2, gasoline exhaust and diesel exhaust. Furthermore, additional sensors have been added to allow correction for meteorological influences (T and RH sensor) and for cross sensitivities (O_3 and VOC sensor).

The ENEA-Air-Sensor Box used in the Aveiro campaign consisted of commercial low-cost electrochemical sensors for gas (NO_2, O_3, CO, SO_2) detection and commercial cost-effective optical particle counter (OPC) for particulate matter (PM10) detection, including miniaturized sensors for meteorology parameters (T and RH). Before the Aveiro campaign, ENEA calibrated the prototype Airbox by co-located measurements with reference analysers in an official air quality monitoring station at JRC, located in Ispra, Italy.

During the Aveiro Intercomparison Exercise, two SNAQ (Sensor Networks for Air Quality) boxes were deployed (hereafter referred to as CAM). Both units utilised Alphasense electrochemical cell (ECC, model B4/BH) for species NO, NO_2, O_3, CO, SO_2 and Total VOC. Measurements of CO_2 (SenseAir K30) and particulate matter (University of Hertfordshire CAIR) were also undertaken. The CAM boxes employ Gill WindSonic 2-D sonic anemometers to assist in source attribution. Before deployment, each sensor box was fitted with new Alphasense electrochemical cells. Throughout the campaign and subsequent data analysis, only the raw signal data from the sensors was used, correcting for temperature and humidity effects as per Popoola et al., (2016). In addition to the factory calibration of the ECC sensors, a second calibration was employed based on a comparison of both CAM sensor boxes with reference instruments of the Department of Chemistry in Cambridge.

2.4. Data analysis and quality control

2.4.1 Meteorological measurements

One of the most important benefits of AQ monitoring networks is the ability to pinpoint pollution sources, and account for regional (> 50 km), meso-scale (500 m to 50 km) and micro-scale (< 500 m) influences. This requires suitable meteorological data to support the measurements. Due to the influence of urban topography and traffic at the micro-scale it is beneficial to have meteorological measurements in the same place as the AQ sensors (Popoola et al., 2013). During the Aveiro campaign, principal meteorological variables were measured by the IDAD LabQAr van using a Vaisala WXT 520 weather station (Borrego et al., 2016). The CAM_10 and CAM_11 boxes measured wind speed, wind direction, temperature and humidity, allowing comparison of meteorological variables and source apportionment.

2.4.2 Measurement uncertainty

The European Air Quality Directive (EU, 2008) defines the Data Quality Objective (DQO) that monitoring methods need to comply with to be used as indicative measurement for regulatory purposes. The DQO is a measure of the acceptable uncertainty for indicative measurements. According to the Directive, allowed uncertainties are 50% for PM10 and PM2.5, 30% for O_3 and 25% for CO, NOx, NO_2 and SO_2.

To assess the performance of each sensor and of the sensor platform as a whole, the measurement of uncertainty has been calculated following the methodology described in JCGM (2008) and Spinelle et al. (2015). The relative expanded uncertainty was estimated using Equation 1, where x_i indicates the reference measurement, y_i the candidate method (sensor), b_0 and b_1 are the slope and intercept of the orthogonal regression, respectively, RSS is the sum of squares of the residuals (Equation 2), and u is the uncertainty of the reference instrument. Further details on the calculation of the expanded uncertainty can be found in the Guide for the demonstration of equivalence (EC WG, 2010).

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

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

2.4.3 Multidimensional data visualization

In order to investigate the behaviour of the AQ nodes in terms of the monitored parameters, it was decided to visualize all meteorological and gaseous data in a way that can reveal dependencies and similarities of patterns; information that can be of value for the node validation and calibration. For this reason, the T-distributed Stochastic Neighbour Embedding (t-SNE) method was employed. This is a relatively new nonlinear mapping technique that is capable of preserving both the local and global structure of a high dimensional dataset (van der Maaten and Hinton, 2008). As multiple parameters (like air pollutant concentrations and meteorological conditions) are produced from the operation of the sensor boxes, it is impossible to simultaneously visualize them and thus investigate possible relationships. The t-SNE method is capable of visualizing this high dimensional feature space in a lower dimensional (2-D or 3-D) space. The main characteristic of the method is that the groups or clusters of features (here consisting of AQ nodes and reference measurements) appearing in t-SNE reflect similarities, thus the closer the attributes are to each other forming a group, the more they can be considered as similar or as belonging to the same cluster. Although the method itself requires multiple iterations and tests and should be only be used for data exploration purposes, ideally, measurements for the same parameter (e.g. NO_2) should be close to each other, regardless of the measuring unit (sensor) producing them.

2.4.4 Multivariate calibration

Air quality multisensor device suffer from several limitations that have their basis in the technological nature of the transducers they rely on. Cross sensitivities make the sensor response depend not only on the target gas, but also on the concentrations of so called interferent species. Environmental parameters like temperature, relative humidity and pressure have similar influence on the sensor to target gas response curve. A calibration that does not take into account these parameter values is prone to failure. Laboratory based calibration procedures use a limited number of combinations of target gas and interferent concentrations. The difficulties in replicating the exact conditions that a calibrating node will encounter in its operating life represent a significant limit to these procedures. In particular, the number of different configurations of target gases and interferent concentrations together with environmental conditions may undergo a “combinatorial” expansion. In order to overcome these limitations, several researchers (De Vito et al., 2008; Kamionka et al., 2006) proposed the use of field measurements taken with a gas multisensor device, as well as a collocated reference analyser, to build a data-driven, multivariate calibration procedure with the aid of neural networks (Webb, 2005). Recently, the use of machine learning approaches has become common practice in the field, primarily for the performance and cost benefits that can be obtained with respect to classic approaches (Vidnerová and Neruda, 2016). Those pioneering results were confirmed by Spinelle et al. (2015), in a series of multinode studies, highlighting the significant benefits of this approach when dealing with real world deployments to the point of partially reaching the DQO set by the EU directive for indicative measurements.

Another driver of sensor node performance limitation is the dynamic behaviour of the sensors. It is usually characterized by a limited responsiveness, thus minimising their ability to deal with rapid transients of pollutants concentrations that may occur in pervasive near-to-road or mobile deployments. In addition, the responsiveness of a single sensor to different gases may differ. To tackle this issue, Esposito et al. (2016) have shown the effectiveness of machine learning approaches, improving the dynamic and overall performances of fast sampling nodes.

The manufacturing variability is a significant limitation to the scalability of each calibration procedure. As previously mentioned, calibration operation for each different node may be required, since their response behaviour to target gases as well as interferents may differ significantly as also shown by Castell et al. (2017). Drift effects, and specifically those related with ageing and poisoning may affect the long term performance of the sensor node requiring relatively frequent recalibration actions (Tsujita et al., 2005; S. Marco et al., 2012). Data driven approaches have been proposed for the improvement of long term performances, but the problem remains open (S. Marco et al., 2012; De Vito et al., 2012).

The univariate calibration approach was implemented with the aid of a simple linear regression (LR). For each sensor, a calibration function was established by assuming the linearity of the sensor response with the reference measurement for each pollutant. Orthogonal linear regression with the minimization of square residuals of the sensor response versus reference measurement was also used.

The multivariate calibration approach was implemented with the aid of Computational Intelligence (CI) methods from the Machine Learning field. Preliminary computational experiments, and literature methods (Spinelle et al., 2015; De Vito et al., 2018) led to use of two CI algorithms; Random Forests (RF) (Tin Kam Ho, 1995) and shallow Feed Forward Neural Networks (FFNN) (Bishop, 2006). Model results were evaluated with the aid of appropriate model performance indices as described in detail in Borrego et al., 2016.

Random Forest (RF) is an algorithm belonging to the ensemble-based classifiers that makes use of decision trees and then estimates the value of the target parameter as the average of the forecasts of each individual tree of the ensemble (Breiman, 2001). The RF algorithm uses bootstrapping; the initial feature space consisting of M observations of N input parameters (features) sampled with replacement to generate a number of M training sets. Each set is then used to train a decision tree. For each tree, a random number of features, n, is selected with replacement (the so called bagging procedure) thus formulating a random subset of the initial instances and then a decision tree is fitted (trained) on each subset and is therefore used in order to predict the parameter of interest. Here the random number of features is calculated as n=int(log2(N+1)), where int is the integer part of a real number, according to Breiman, 2001. An unlimited number of levels and nodes are used for each of the aforementioned random trees. The final prediction is calculated as a (voted or averaged) sum of individual predictions, thus making RF an ensemble-based meta-classifier. It should be noted that node split per tree, i.e., the splitting per node based on feature threshold values, was the optimum among a random subset of the features of size n. On this basis, the variance decreases due to the averaging in the ensemble, leading to an overall improvement in results.

Following the same modelling approach, shallow Feed Forward Neural Networks (FFNN) (Bishop, 2006) have been trained and tested for each of the AQ nodes. FFNN has become a reference tool for multivariate regression in the machine learning community making it a useful comparison method. Given its small operative computational complexity, it is arising as a tool of choice for implementing on board embedded intelligence when availability of computational resources is scarce e.g. when targeting mobile/wearable deployments (De Vito et al., 2018). In our implementation, we made use of 5 sigmoidal neurons in the single hidden layer for the 1-h dataset, and a three layer FFNN with 5 hidden layer neurons and a single output for the 1-min dataset. The reported performance indices are averaged along multiple implementations of the cross validation procedure, due to the inherent dependence of the performance on the random choice of the initial network weights parameter. No hyper parameter optimization procedure (Bishop, 2006) has been implemented.

Performance estimation for both RF and FFNN was done on the basis of a 10-fold cross validation procedure: the initial data set was randomly divided into 10 equal subsets. Nine of them were used for model training while the tenth was used for model evaluation. This procedure is repeated ten times, each time leaving a different subset out, to be used for model evaluation. In this way, ten models were trained and evaluated, and the overall prediction scores were calculated as averages of the individual models.

Sensor responses and IDAD reference values have been used for the cross validation based training and performance assessment. It is worth noting that for each sensor box, all available and meaningful raw sensor responses have been used to provide the multivariate input to the above mentioned CI methods. For the purposes of the present study, the WEKA computational environment was employed (Hall et al., 2009) for RF while Matlab was used as the computational environment for ANNs.

3. Results and discussion

3.1. Identification of pollutant sources

As mentioned in section 2.4 during the Aveiro campaign, principal meteorological variables were measured by the IDAD LabQAr van, using a Vaisala WXT 520 weather station. Additionally, CAM_10 and CAM_11 boxes included direct observations of wind speed, wind direction, temperature and humidity. The results are presented in Fig. 2 and Fig. 3, allowing comparison of meteorological variables and source apportionment.

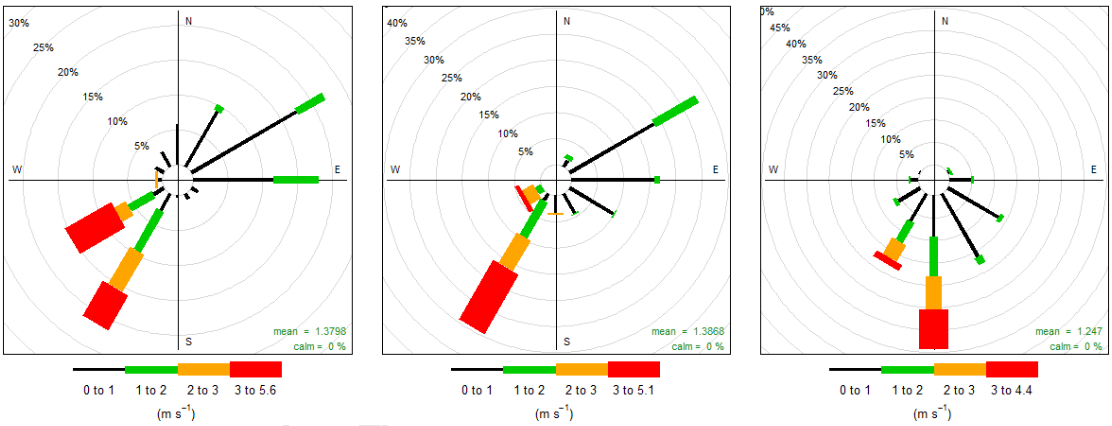


Fig. 2. Wind speed and wind direction measured (from left to right) by IDAD LabQAr and CAM_11 and CAM_10 sensors. Wind speeds are split by the coloured intervals shown in each panel, the grey circular lines show the directional frequency (%).

From the wind speed and direction data, it can be seen that both the CAM_11 and IDAD sensors capture the split between west/south-westerly (W/SW) and east/north-easterly (E/NE) winds in terms of frequency, with the strongest winds from a SW direction during the first week. The sonic anemometer on CAM_11 was partially blocked to the west by the telescopic pole, hence it has missed the frequent winds to the WSW measured by the WXT520. In contrast, CAM_10 measured predominantly southerly winds (S) throughout the study period – possibly as a result of its mounting position lower down on the van roof.

In Fig. 3 CAM sensor boxes demonstrate good agreement with the WMO certified WXT520, well capturing the diurnal trends. Nevertheless, there is a clear positive bias of 5°C in temperature readings from both CAM boxes relative to the WXT520, which is especially prevalent during the second week of measurements when the diurnal range was around 20°C. A smaller positive bias (~ +2 – 3%) is also observed in relative humidity readings compared to the WXT520.

The temperature bias appears to be systematic with the PT1000 thermocouples used in this study, and for future measurements will need to be corrected, especially under periods of high insolation as experienced during the second week of the campaign (Borrego et al., 2016). It is possible that the lack of an adequately ventilated and screened enclosure for the temperature and humidity sensors could contribute to these biases, although this requires further investigation.

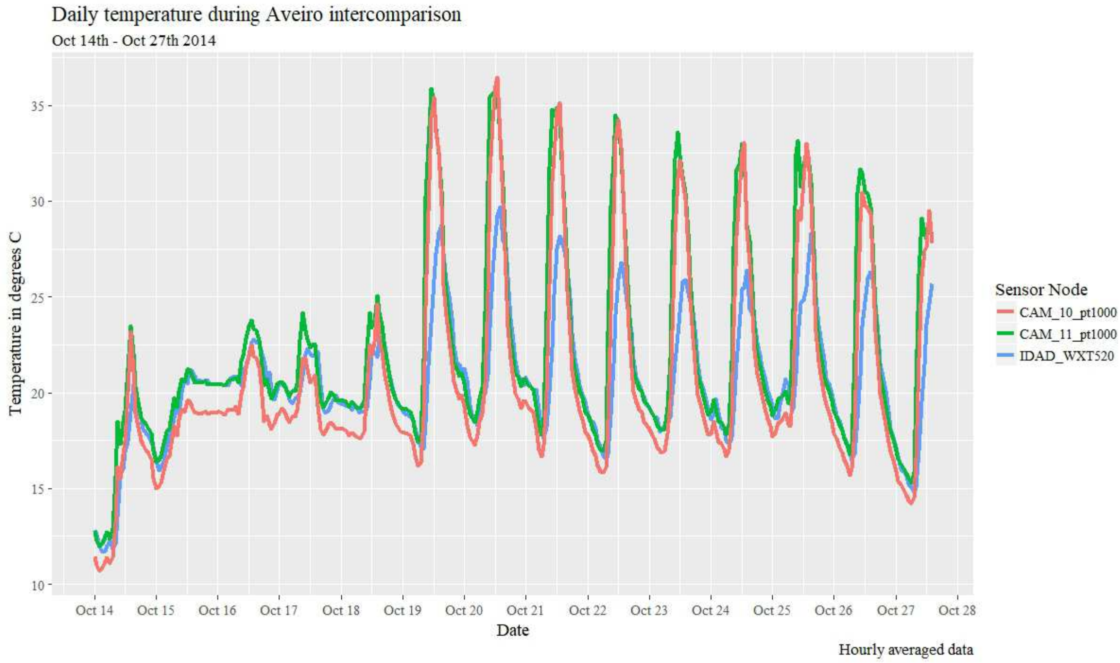


Fig. 3. Temperature °C (top) and relative humidity % (bottom) measured by IDAD LabQAr, CAM_10 and CAM_11 sensor boxes.

Nevertheless, the wind speed and direction data can be used to highlight local or regional sources of pollution. For example, Fig. 4 and Fig. 5 show pollution rose plots[1] of NO_2 and PM2.5 conditioned with CO. This process effectively plots x and y variables against a third

[1] http://www.openair-project.org

variable, in this case CO, as concentrations increased during the period of monitoring. As CO can be used as a tracer for combustion processes, this helps to pinpoint clean and polluted air.

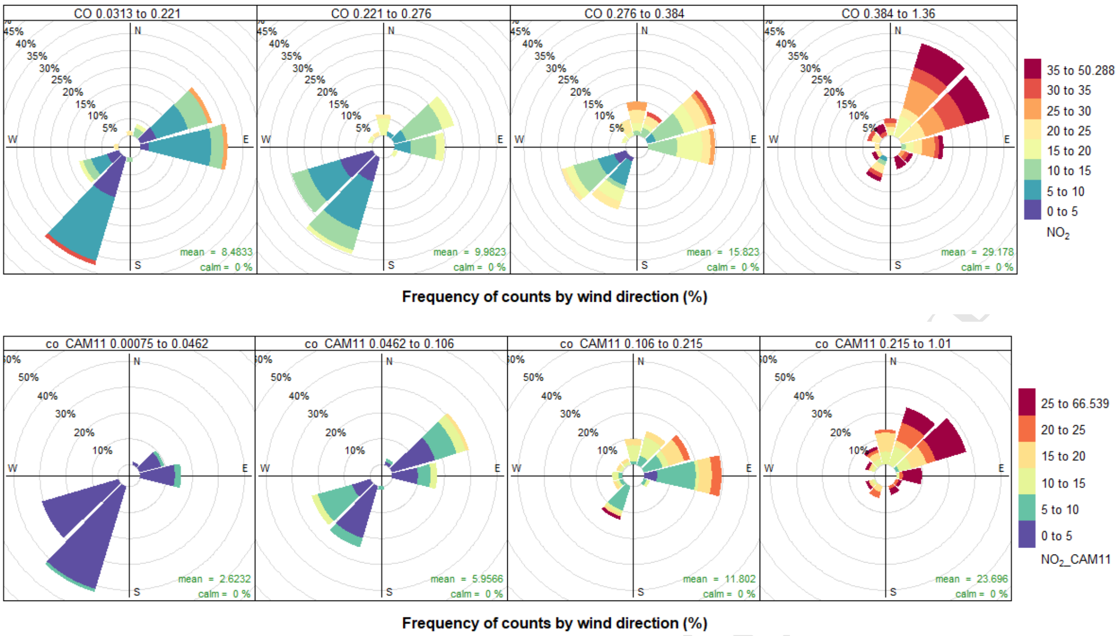


Fig. 4. Pollution rose plots of NO_2 (ppb) vs wind direction conditioned by CO. IDAD LabQAr (top) and CAM_11 (bottom). The concentration of CO (ppm) increases from left-to-right. Concentration of NO_2 is shown by the coloured intervals, and wind directional frequency by the grey contour lines. Hourly averaged data.

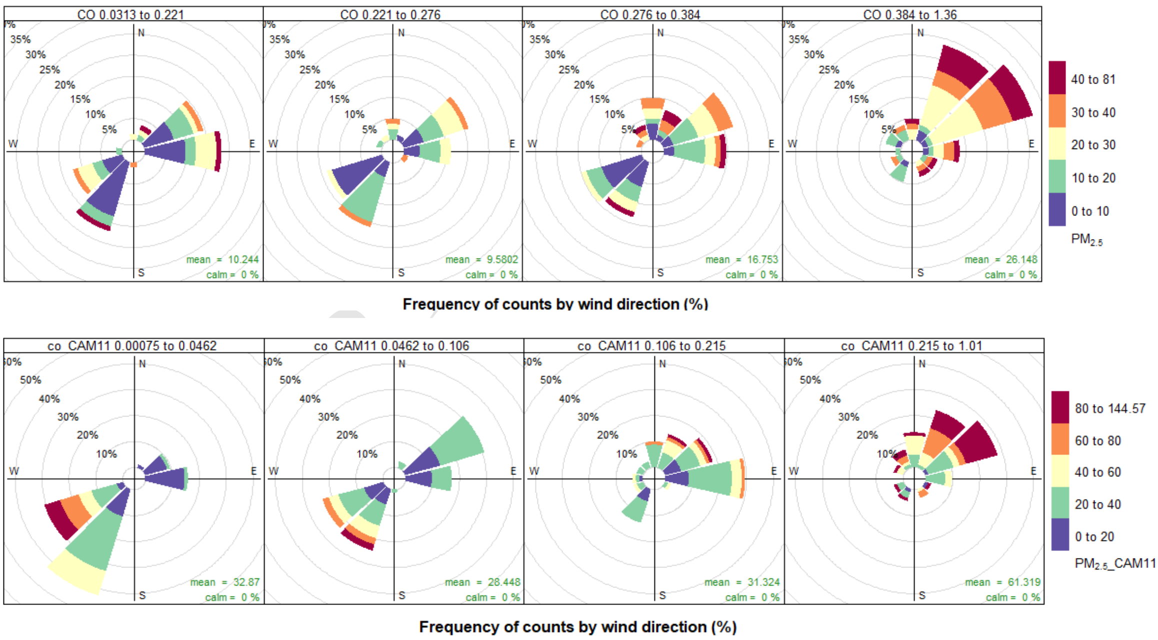


Fig. 5. Percentile rose plots of PM10 ($\mu g.m^{-3}$) *vs* wind direction derived from CAM_11 sensor data, split for day and night periods. In this case, one-minute averaged data were used. The percentile intervals are shaded and are shown by wind direction, with the concentration in ppbv indicated by the contours.

The IDAD LabQAr reference and CAM_11 sensors both indicate that as the $CO+NO_2$ burden increases, greater than 60% of the highest concentrations originate in the E/NE direction. Traffic flow was NE to SW alongside the measurement site (Fig. 1), so under low wind speeds local emissions will dominate. In contrast under S/SW conditions (Fig. 2), the $CO+NO_2$ burden is reduced dramatically (less than 15 ppb NO_2) indicating the importance of higher wind speeds for transporting cleaner background air and subsequent mixing and transport of these local emissions away from the site.

Similarly, the CO+PM2.5 concentrations are greatest in the E/NE direction for both sensor nodes, suggesting traffic and/or other local sources of combustion. During the second week of measurements, low wind speeds, high solar insolation and stagnant conditions lead to an inversion and build-up of pollutants (Borrego et al., 2016). However, there is also a source of PM2.5 to the WSW of the CAM_11 sensor node, seen to lesser extent in the IDAD data. The CAM_11 particle sensor uses an optical, not gravimetric technique and shows enhanced response under high relative humidity (experienced during the first week) and this interference cannot be ruled out.

The examples above show how local meteorological data combined with micro-sensor networks, can be of use to pinpoint pollution sources, and could be a valuable tool for policy decisions regarding urban traffic control and development. However, there are caveats. Meteorological sensors need to be properly sited with adequate protection from direct sunlight and rain, and need to be close to the pollution sensors themselves, ideally in the same enclosure. Care needs to be taken with post-processing of the data (for example, with the CAM PT1000 thermocouple temperature sensors used in this study) and where possible a WMO certified weather station should be used for quality checks.

3.2. Parameter dependencies

The analysis with the aid of the t-SNE method led to a number of visualizations like the one presented in Fig. 6.

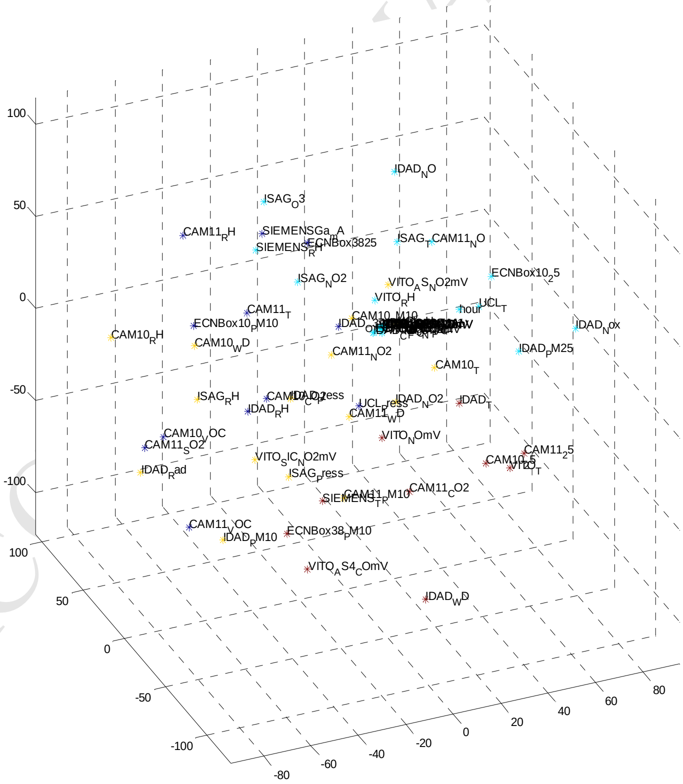


Fig. 6. A visualisation of the Aveiro dataset based on the t-SNE method. Proximity is a measure of similarity. Labels by the dots indicate the AQ node and the parameter (feature) monitored.

Although such graphs are not straightforward to analyse, it is possible to identify parameters that are visualised as being "close" to each other, via rotation and figure magnification. This "closeness" should actually be interpreted as the tendency that the data points have to organise themselves in neighbourhoods (clusters) of data, on the basis of a mathematically defined similarity. On this basis, we identified that when it comes to meteorological parameters like pressure, measurements of UCL as well as of AUTh-ISAG are close to the reference node of IDAD. For gaseous pollutants, it was evident that CAM_11 is close to the reference node, thus indicating a good overall agreement (this being suggested already by the findings of Borrego et al., 2016).

Such similarities can also be identified regarding different sensor types belonging to the same node, as it is the case with the VITO sensors for CO (MiCS-5525, MiCS-5521, Figaro 2201, Alphasense CO-BF), all of which produce mV as initial output: all four sensors were visualised very close to each other, thus suggesting that they can be considered as similar and providing “equivalent” information. Such findings are useful for selecting, in a next step, the sensor that is also more easily calibrated and possesses additional characteristics like response time for example, in order to deploy a node implementation in a future application.

3.3. Calibration of sensors

To transform the sensor responses into air pollutant concentrations, it is necessary to create a calibration function. In some cases, the calibration function is provided by the manufacturer and it is usually determined by comparing the sensor response versus reference values in a laboratory environment. In the laboratory, the conditions of temperature and relative humidity are controlled and the pollutant concentrations are precisely regulated. However, the laboratory calibration in most cases is not enough to cope with environmental variability and unpredictable interferences found in the field. In these cases, the sensor performance (in terms of concentration estimation quality) often decrease dramatically; presenting in some cases biases that can be partially corrected if a field calibration is employed as described in Spinelle et al. (2015). In this section, the results are presented from the field calibration obtained comparing the data from the sensor platforms with the data from the reference instruments in Aveiro.

The results from the linear regression univariate calibration, implemented with the aid of a simple linear regression, highlight the need for field calibration, as most of the sensors present a slope and an intercept that significantly differ from the optimal target values of 1 and 0, respectively. This confirms the findings of the target diagram regarding the bias of sensor values (Borrego et al., 2016). Actions also must be taken to correct both the zero and range of sensors in field calibration conditions. Overall results suggest that the CO sensors are generally the most effective, with intercept close to 0 and uncertainty in the slope within 35%.

To minimize the concentration estimation errors, we made use of the (1-h) IDAD dataset of reference meteorological measurements, plus microsensor data, to construct multivariate calibration models. As mentioned in section 2.4, two computational intelligence (CI) algorithms were employed, namely RF and FFNN. We compared the results received with those obtained via the basic lab-based and linear calibration methods. Table S2 defines all statistical indicators and Table S3 shows the values of all four calibration methods used (see Supplementary Material), while Fig. 7 compares all methods for all sensor nodes and pollutants in terms of correlation coefficient.

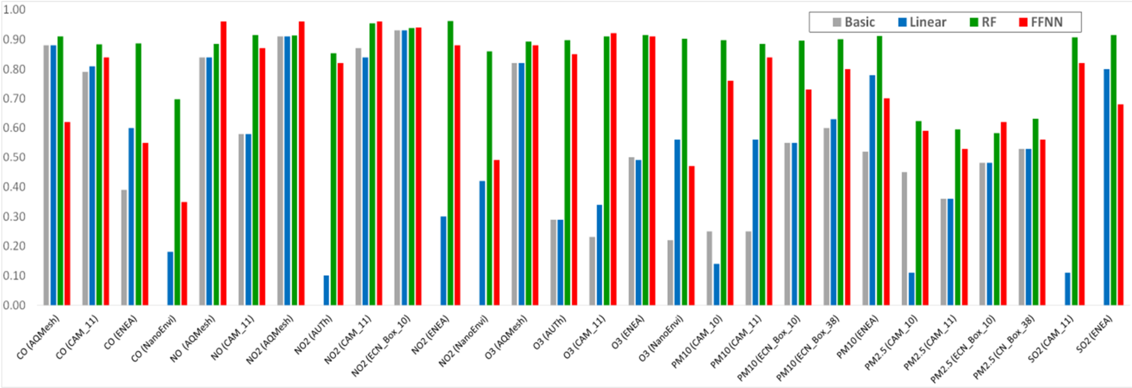


Fig. 7. Comparison of the basic, linear, RF and FFNN oriented calibration methods in terms of reference correlation coefficient achieved for all sensor nodes and pollutants based on hourly measurements. Missing values indicate insufficient data.

In the case of CO the good performance of the basic and linear calibration procedures (max(r)=0.88) is further improved via RF and FFNN (max(r)=0.91). The latter demonstrates a higher MRE and a lower FOEX (in absolute values) when compared to FFNN for AQMesh and CAM_11 thus indicating that CI-based calibration procedures are sensitive to errors and require further improvement.

Regarding NO sensors, both RF and FFNN improve the already good performance of the basic and the linear calibration method in terms of almost all statistical indices. Thus basic and linear methods reach a max(r)=0.84 that rises up to r=0.96 with FFNN.

All six NO_2 sensor nodes demonstrate good performance indicators (max(r)=0.91 with basic and linear methods, reaching up to max(r)=0.96). It is worth noting that the application of RF or FFNN algorithms leads to a strong improvement in the correlation coefficient especially for those nodes that demonstrated the worst performance in basic calibration mode, thus greatly improving their operative capabilities. This is the case for AUTh-ISAG, ENEA and NanoEnvi, where the calibrated values reach now r>0.7.

For O_3 the correlation coefficient achieved for basic calibration ranges from r=0.22 up to r=0.82, while with linear calibration ranges from r=0.29 up to 0.82. RF leads to very good correlation coefficients ranging between r=0.89 to r=0.91 while FFNNs demonstrate a range of r=0.47 to r=0.92. The MAE of both RF and FFNN calibration results is lower than basic or linear calibration results. On this basis, the two CI methods improve the calibration outcomes for O_3.

For PM10, results indicate a good overall correlation between the reference and the available measurements, A correlation coefficient of r=0.91 was the maximum value (for ENEA) and of r=0.88 was the minimum value (for CAM_11) achieved with the use of RF. On the other hand, a correlation coefficient of r=0.84 was the maximum value (for CAM_11) and r= 0.7 was the minimum value (for ENEA) achieved with the use of FFNN.. In all cases, FFNN surpasses basic and on field linear calibration methods, while RF performs even better than FFNN, for all sensor boxes. On the other hand, the CRMSE (Centred Root Mean Square Error) achieved with the use of FFNN (being in all cases below 25) is much lower than the one achieved via RF (being in all cases above 145) for all sensor nodes, the best one observed for CAM_11. Concerning the MBE, RF led to lower values in comparison to FFNN. The FOEX index of over/under estimation is closer to the ideal one (zero) for RF results with respect to FFNN results, yet both are clearly closer to zero in comparison to basic calibration, while linear calibration is similar to FFNN.

For PM2.5, results indicate that both RF and FFNN improve the correlation coefficient ranging from r=0.53 up to r=0.63 in comparison to the best one achieved with basic calibration (ranging from r=0.36 up to r=0.53). The CRMSE with FFNN is again much better compared to the one achieved with RF.

Both sensor boxes (CAM_11 and ENEA) with SO_2 measurement demonstrate high correlation coefficient (with a maximum surpassing 0.9), low MBE, and low CRMSE. The FOEX index is also improved, with RF providing with the best values.

Considering fast sampling systems suitable for mobile and pervasive deployments, results at 1-min sampling rate have been computed. Basic statistical performance indicators are reported for the sensors that provided 1-min data in Table 2, considering the same indices adopted by Borrego et al. (2016). In this case, the sensor performance analysis was carried out using the observed data from sensor platforms with 1-min sampling rate and the reference data provided by the 1-min IDAD dataset. As a further comparison baseline, univariate linear regression models have been applied to targeted sensors to estimate the target concentration using 1-min data. Results are reported in Table 3.

Table 2. Basic performance indicator for target sensors using off the shelf basic calibration (1-minute sensor data used). No off the shelf calibration was provided for Siemens node sensors. Acronyms are explained in Table S2 of the supplementary material.

Pollutant	Sensor node	MBE	r	r^2	CRMSE	NMSE	FB	FOEX	MAE	MRE
O_3	CAM_11	22.22	0.24	0.06	21.22	1.78	1.30	-2.93	24.19	1.40
O_3	VITO	-16.05	0.09	0.01	10.91	36.45	-1.92	-40.36	16.10	0.82
NO_2	CAM_11	-2.28	0.82	0.67	6.76	0.43	-0.37	-29.04	4.56	0.34
NO_2	ECN	0.93	0.87	0.75	6.56	0.26	0.14	-22.62	4.28	0.33
NO_2	VITO (NO_2MiCS2710)	-9.28	-0.30	0.09	11.44	5.27	-1.66	-28.15	9.95	0.70
NO_2	VITO (Figaro)	-10.01	-0.39	0.15	11.88	9.15	-1.77	-28.49	10.86	0.81
CO	CAM_11	-0.21	0.80	0.64	0.17	3.06	-1.38	-46.88	0.24	0.86
CO	Siemens	-	-	-	-	-	-	-	-	-
CO	VITO (Alphasense)	1.36	0.43	0.19	0.24	0.61	1.82	48.03	1.36	5.43
CO	VITO (MiCS5525)	1.81	0.30	0.09	0.25	0.69	1.88	48.03	1.82	7.11
CO	VITO (MiCS5521)	2.47	0.59	0.35	0.22	0.75	1.93	47.89	2.47	9.62
CO	VITO (Figaro)	1.03	0.77	0.59	0.18	0.54	1.73	47.81	1.03	3.95
NO	CAM_11	9.91	0.47	0.22	20.65	1.54	1.12	27.99	15.25	8.69
SO_2	CAM_11	30.14	-0.23	0.05	12.54	6.92	1.99	39.05	30.15	19.15

Table 3. Linear regression outcomes (1-minute sensor data used). Acronyms are explained in Table S2 of the supplementary material.

Pollutant	Sensor node	MBE	r	r^2	CRMSE	NMSE	FB	FOEX	MAE	MRE
O_3	CAM_11	2.86	0.34	0.12	8.87	10.47	0.28	-19.97	6.69	0.66
O_3	VITO	-0.19	0.09	0.01	10.89	108.35	-0.02	-0.26	8.80	1.85
NO_2	CAM_11	0.26	0.83	0.68	6.25	0.45	0.04	5.81	4.25	0.60
NO_2	ECN	1.85	0.88	0.78	5.92	0.31	0.26	-6.00	3.76	0.38
NO_2	VITO (NO_2MiCS2710)	-0.02	0.30	0.09	10.75	9.82	0.00	11.97	8.29	1.31
NO_2	VITO (Figaro)	0.00	0.39	0.15	10.38	5.64	0.00	13.22	7.72	1.12
CO	CAM_11	0.02	0.80	0.64	0.16	0.58	0.12	-4.39	0.11	0.40
CO	Siemens (AppSens1)	0.00	0.50	0.25	0.23	2.90	0.01	4.90	0.16	0.55
CO	Siemens (AppSens2)	0.00	0.58	0.33	0.22	2.08	-0.01	4.97	0.15	0.54
CO	Siemens (Ga2O3hp)	0.00	0.34	0.12	0.25	7.29	0.00	8.14	0.18	0.64
CO	Siemens (Ga2O3hpHeater)	0.00	0.30	0.09	0.25	10.54	0.00	10.22	0.18	0.63
CO	Siemens (MicronasPht_L)	0.00	0.26	0.07	0.26	14.69	0.00	8.67	0.18	0.65
CO	Siemens (MicronasPt_L)	0.00	0.17	0.03	0.26	31.35	0.00	8.68	0.18	0.69
CO	VITO (Alphasense)	0.00	0.43	0.18	0.24	4.43	0.01	5.65	0.19	0.71
CO	VITO (MiCS5525)	0.00	0.30	0.09	0.25	9.25	0.00	9.58	0.18	0.70
CO	VITO (MiCS5521)	0.00	0.60	0.35	0.21	1.73	-0.01	6.17	0.16	0.61
CO	VITO (Figaro)	0.00	0.78	0.59	0.17	0.65	-0.01	3.02	0.12	0.46
NO	CAM_11	0.26	0.40	0.16	20.05	5.15	0.04	16.85	11.00	4.53
SO_2	CAM_11	-0.01	0.22	0.05	0.52	21.85	-0.003	-0.75	0.44	0.27

By using the same experimental design for the 1-minute data, we also report the results obtained by:

(a) the application of the Random Forest algorithm (Table 4)

(b) a FFNN (three layers with 5 hidden layer neurons) (Table 5)

In all cases both RF and FFNN greatly improve the node performance in comparison to the basic calibration and the linear correlation based-calibration.

Obtained results show a significant improvement of the performance indices with both CI architectures. The use of RF as well as FFNN , strongly improves the performances obtained by linear regression estimators: in all cases (with the exception of CAM_11 monitored CO for FFNN), both RF and FFNN greatly outperform linear and basic calibration performance, confirming the power of multivariate non-linear regression when coupled with field data for microsensor based air quality monitor calibration. Moreover, the obtained results indicate that the RF approach outperforms the Neural one in several cases. To the best of our knowledge this is the first study to report this advantage in this particular field; providing an insight into on field calibration for further investigation.

Table 4. Random Forests (RF) based non-linear multivariate regression outcomes. Acronyms are explained in Table S2 of the supplementary material.

Pollutant	Sensor node	MBE	r	r^2	CRMSE	NMSE	FB	FOEX	MAE	MRE
O_3	CAM_11	0.00	0.98	0.97	4.20	0.03	0	0.77	1.32	0.20
O_3	VITO	0.01	0.94	0.89	14.17	0.11	0	2.68	2.05	0.29
NO_2	CAM_11	0.01	0.94	0.89	14.39	0.11	0	11.06	1.97	0.17
NO_2	ECN	-0.04	0.94	0.88	14.73	0.12	0	5.46	2.31	0.29
NO_2	VITO	-0.02	0.92	0.85	19.48	0.16	0	10.92	2.38	0.27
CO	CAM_11	0.005	0.94	0.88	0.01	0.13	0.01	4.98	0.07	2.80
CO	Siemens	0.003	0.92	0.85	0.01	0.15	0.01	0.37	0.07	4.73
CO	VITO	0.002	0.79	0.62	0.03	0.38	0	4.81	0.11	6.35
NO	CAM_11	0.07	0.86	0.74	112.22	0.26	0	21.47	4.76	1.27
SO_2	CAM_11	-0.002	0.97	0.95	0.02	0.05	0	0.39	0.09	0.05

Table 5. Feed Forward Neural network (FFNN) based non-linear multivariate regression outcomes. Acronyms are explained in Table S2 of the supplementary material.

Pollutant	Sensor node	MBE	r	r^2	CRMSE	NMSE	FB	FOEX	MAE	MRE
O_3	CAM_11	2.79	0.93	0.86	4.49	0.16	0.28	-27.72	2.60	0.18
O_3	VITO	-0.16	0.98	0.96	2.31	0.05	-0.02	0.25	1.64	0.20
NO_2	CAM_11	-3.50	0.90	0.81	4.90	0.23	-0.61	-24.77	2.00	0.25
NO_2	ECN	0.78	0.90	0.81	4.96	0.24	0.11	-31.69	3.25	0.31
NO_2	VITO	-0.03	0.91	0.84	4.50	0.19	0.00	5.15	2.84	0.30
CO	CAM_11	-0.06	0.71	0.51	0.14	0.91	-0.35	-25.80	0.09	0.34
CO	Siemens	0.11	0.76	0.58	0.17	0.71	0.04	1.47	0.12	0.44
CO	VITO	0.00	0.86	0.74	0.13	0.33	-0.01	4.28	0.10	0.37
NO	CAM_11	-6.26	0.62	0.39	7.91	1.56	-1.34	-22.37	2.39	0.84
SO_2	CAM_11	-0.06	0.93	0.86	0.22	0.16	-0.07	-27.32	0.16	0.10

3.4. Measurement of nodes expanded uncertainty

The results for the relative expanded uncertainty (see EU, 2018; EC WG, 2010) of the single sensors show that, generally, the relative uncertainty of CO targeted sensors is the lowest among the different target gases under analysis. Slightly higher values have been recorded for the uncertainty of some of the NO and NO_2 sensors. AQMesh and ECN nodes demonstrate the lower relative expanded uncertainty U_r for NO_2, with U_r reaching below 30% and 60% for concentrations above 20 ppb, for the two aforementioned nodes. However, the AQMesh NO_2 sensor seems to increase the U_r for concentrations over 30 ppb. For PM10, the results vary considerably from platform to platform; the ECN platform, for example, presents lower relative uncertainty for concentration values above 80 $\mu g.m^{-3}$ approaching DQO limits, while the CAM platforms present an U_r over 600% for all the concentration range. For O_3, AQMesh and NanoEnvi nodes show a reduction in U_r values for concentrations over 10 ppb, while with AQMesh presents the lowest relative uncertainty. SO_2 sensors are generally characterised by the highest relative uncertainty values for all the gas sensors analysed, with uncertainties values over 5000%: it should be underlined that these results are influenced by the very low concentrations recorded during the exercise. Overall, the aforementioned analysis verifies that even when some of the sensors demonstrate promising uncertainty results, sensor platforms are still not in compliance with the DQO imposed by the European Air Quality Directive (AQD) for indicative measurements when off the shelf calibration is used (EU, 2008).

Considering the results of the FFNN processed multisensors response, the results show a significant improvement with respect to off the shelf calibration. Figure 8 represents the relative expanded uncertainty of the platforms versus reference data with the FFNN processed multisensors response. As usual, figures show results obtained on test samples that have not been used in the network training phase.

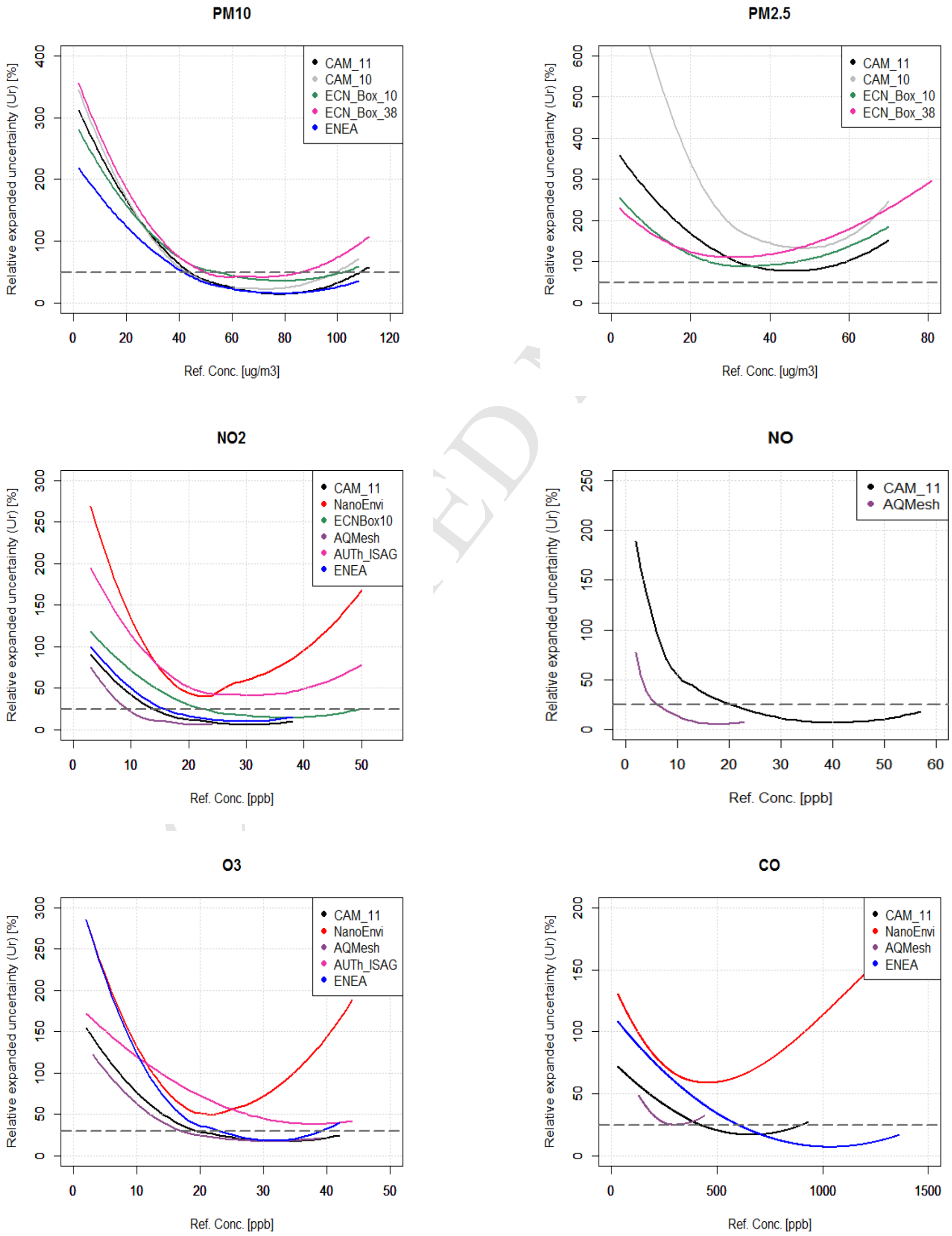
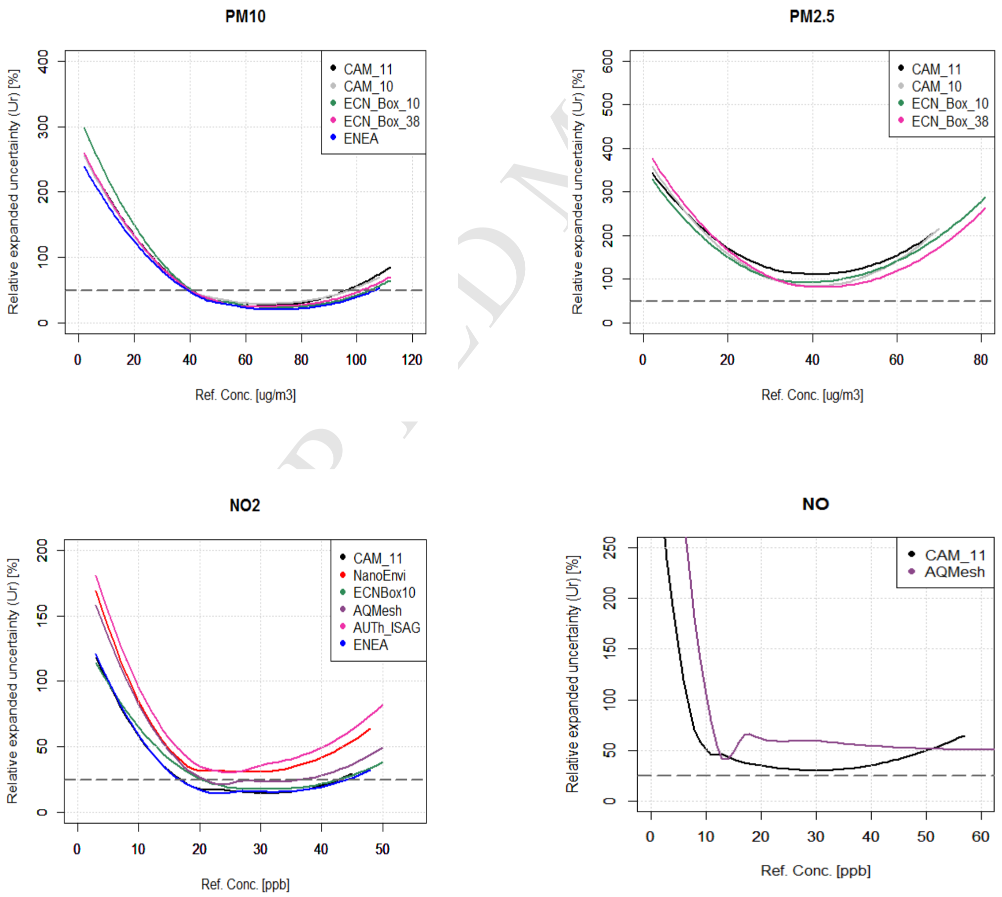


Fig. 8. Relative expanded uncertainty (%) of the FFNN processed multisensors *versus* the reference data, for PM10 (µg.m^{-3}), PM2.5 (µg.m^{-3}), NO_2 (ppb), NO (ppb), O_3 (ppb) and CO (ppb). Each colour identifies a specific platform: black (CAM_11), red (NanoEnvi), grey (CAM_10), green (ECNBox10), magenta (ECNBox38), blue (ENEA), orchid (AQMesh), pink (AUTh_ISAG).

For PM10 sensors, the estimated uncertainties generally meet the AQD data quality objective in part of the measurement range (40 - 100 µg.m^{-3}). On the contrary, for PM2.5, despite the improvement obtained with FFNN calibration, the estimated uncertainty is still higher than the AQD objective for all values in the relevant range (1 - 80 µg.m^{-3}).

With regards to NO_2, the majority of platforms show results that match with the uncertainty criteria defined in the AQD when concentration exceeds a node dependent threshold. In this case, only the results for NanoEnvi and ISAG sensors are above the AQD objective across the entire measurement range (2 - 50 ppb). For O_3 sensors, an equivalent result is obtained, with some of the platforms showing very promising results.

With the NO sensors, the limited number of data available is reflected in the uncertainty results. In this case the uncertainty is below the reference value in part of the measurement range when concentration exceeds 5 ppb and 18 ppb respectively for the AQMesh and CAM_11 node. For CO, all but the NanoEnvi and CAM11 nodes meet the AQD objective at least in a small concentration interval starting from 300 ppb for AQMesh node, 400 ppb for CAM_11 node and 600 ppb for ENEA node. Given the usual relevant concentration range for CO (0.5-10 ppm) the ENEA node gives the best results.



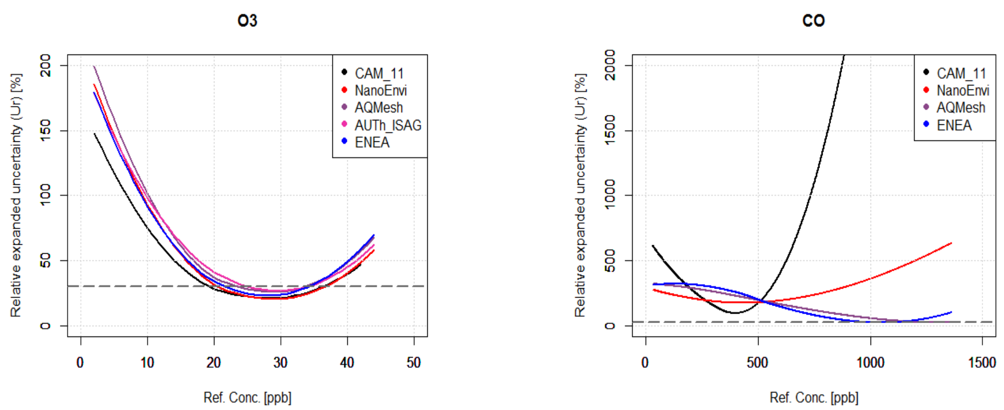


Fig. 9. Relative expanded uncertainty (%) of the RF processed multisensors *versus* the reference data, for PM10 (µg.m^{-3}), PM2.5 (µg.m^{-3}), NO_2 (ppb), NO (ppb), O_3 (ppb) and CO (ppb). Each colour identifies a specific platform: black (CAM_11), red (NanoEnvi), grey (CAM_10), green (ECNBox10), magenta (ECNBox38), blue (ENEA), orchid (AQMesh), pink (AUTh_ISAG).

Similar results are obtained by the use of the RF algorithm (Fig. 9). For PM10, the results reflect what has already been shown with the aid of FFNN, with an increased homogeneity among the node's performances. For PM2.5, again none of the nodes is capable of reaching the DQO levels. For NO_2, slightly worse results are obtained with RF, however it can be seen that most of the nodes are able to reach the DQO in relevant concentration ranges. Results for the NO sensors could not reach the quality objective achievement with the use of RF. On the other hand, results obtained for O_3 sensors suggest that by the use of RF regression, all the nodes, including NanoEnvi and AUTh-ISAG are able to meet the intended objectives. Finally, for the carbon monoxide, results with RF reflect the higher MRE already presented in Table 3.

4. Conclusions

Continuing developments in Air Quality (AQ) microsensor technologies present the need to evaluate their ability to support and complement standard monitoring procedures. This paper introduced the second part of the results of an intercomparison of AQ microsensors with reference methods during an AQ monitoring campaign in Aveiro (Borrego et al., 2016).

Microsensor nodes are low-cost devices with considerable application potential, offset by important limitations when applied to urban air quality monitoring. Heterogeneity of hardware and calibration procedures call for an additional "calibration layer" that should be implemented in any AQ information system that uses them. Sensor uncertainty and performance indicators need to be addressed in a structured way allowing for field-based sensor comparison and use.

This work focused on the analysis of the uncertainty estimation along with the development of a Computational Intelligence-based "calibration" layer focusing on the hourly data, while also addressing high temporal resolution (1-min) data. The aim was to estimate the uncertainty of the measurements according to the DQO of the European Air Quality Directive and to improve their performance with the aid of a computationally-oriented methodology. For this purpose, linear regression in addition to Feed Forward Neural Networks and Random Forests algorithms were employed, in an effort to improve sensor performance based solely on sensor data and local meteorological data from a reference station.

The results confirm that the standard in-factory calibration performances can be strongly ameliorated by CI-based algorithms with positive outcomes on several statistical indices, as already reported in the literature (De Vito et. al., 2018). Improved results can also be obtained for sensor nodes for which no in-factory calibration has been performed. For the first time, this is confirmed with tests on several multisensor nodes based on different type of solid state sensors and from different independent institutions, including academia and commercial companies, adding an unprecedented generalization value to the presented result. The latter also allows for the adoption of the proposed sensor calibration methodology in order to make sensor readings compliant with the DQO, thus also for the first time suggesting a method that may render AQ microsensors as appropriate for the support of official AQ monitoring tasks.

Nonetheless, this procedure should be regarded as an on-site calibration, as it requires ground truth data from the area of interest, thus posing a considerable challenge in terms of method scalability, required for rapid deployment.

Generalization of these results should take into account possible sensor and concept drift impacts on concentration estimation quality. It is well known, actually, that the ageing (sensor drifts) and seasonal changes in the pollutants and/or environmental variables joint probability distribution (concept drifts), may lead to suboptimal estimations negatively affecting performance estimator values (Esposito et al., 2017).

Furthermore, the performance degradation rate due to ageing effects may be different for each sensor. This behaviour could only be observed and quantified during long-term deployment campaigns and are not the subject of this study.

Further work should evaluate the robustness of the methodologies described herein. Sensor profiling with the aid of CI methods in addition to SOM can support aforementioned tasks. In addition, model performance may take benefit from an ensemble approach where different models based on different algorithms can be used in combination.

Acknowledgements

The authors would like to acknowledge the support of COST Action TD 1105 – European Network on New Sensing Technologies for Air-Pollution Control and Environmental Sustainability – EuNetAir.

The authors would like to acknowledge the support of the Municipality of Aveiro.

The UCL authors acknowledge the FP7 EU SOIHITS project for funding (grant agreement no 288481) and especially the partners UCAM (F. Udrea), CCMOSS (S. Z. Ali, F. Chowdhury) and UCL (D. Flandre).

The IDAEA-CSIC authors would like to acknowledge AQMesh for kindly providing the AQMesh pods. Support is acknowledged to Generalitat de Catalunya (AGAUR 2017 SGR41). M.C. Minguillón acknowledges the Ramón y Cajal Fellowship awarded by the Spanish Ministry of Economy, Industry and Competitiveness.

The NILU authors would like to acknowledge ENVIRA Ingenieros Asesores for participating in the campaign with the NanoEnvi platform.

References

Balzano L. and Nowak R. (2008). Blind Calibration of Networks of Sensors: Theory and Algorithms, in Networked Sensing Information and Control (Saligrama V., ed), p. 9-37, Springer, Boston, doi: 10.1007/978-0-387-68845-9_1

Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer Verlag, New Ed. 2

Borrego, C., Costa, A.M., Ginja, J., Amorim, M., Coutinho, M., Karatzas, K., Sioumis, Th., Katsifarakis, N., Konstantinidis, K., De Vito, S., Esposito, E., Smith, P., André, N., Gérard, P., Francis, L.A., Castell, N., Schneider, P., Viana, M., Minguillón, M.C., Reimringer, W., Otjes, R.P., von Sicard, O., Pohle, R., Elen, B., Suriano, D., Pfister, V., Prato, M., Dipinto, S., Penza, M. (2016). Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise, Atmospheric Environment, Volume 147, December 2016, Pages 246-263, ISSN 1352-2310, http://dx.doi.org/10.1016/j.atmosenv.2016.09.050

Breiman, L. (2001). Random Forests, Machine Learning, 45(1), 5-32. doi:10.1023/a:1010933404324

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environment International, 99, 293-302. doi:10.1016/j.envint.2016.12.007

De Vito, S., Esposito, E., Salvato, M., Popoola, O., Formisano, F., Jones, R., & Di Francia, G. (2018). Calibrating chemical multisensory devices for real world applications: An in-depth comparison of quantitative Machine Learning approaches. *Sensors and Actuators B: Chemical*, *255*, 1191-1210

De Vito, S., Massera, E., Piga, M., Martinotto, L., Di Francia, G., On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario, Sens. Actuators B: Chem., 129 (2) (2008), pp. 750–757

EC WG, 2010. Guide to the Demonstration of Equivalence of Ambient Air Monitoring Methods. Report by EC Working Group on Guidance. Available at: ec.europa.eu/environment/air/quality/legislation/pdf/equivalence.pdf

EEA, 2017. Air quality in Europe — 2017 report. EEA Report No 13/2017. ISSN 1977-8449

Esposito, E., De Vito, S., Salvato, M., Bright, V., Jones, R. L., & Popoola, O. (2016). Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. *Sensors and Actuators B: Chemical*, *231*, 701-713

Esposito, E., De Vito, S., Salvato, M., Fattoruso, G., Castéll, N., Karatzas, K., Di Francia, G., Is on field strategy robust to relocation?. Olfaction and Electronic Nose (ISOEN), 2017 ISOCS/IEEE International Symposium on, 1-3

EU, 2008. Directive 2008/50/EC of the European Parliament and the Council of 21 May 2008 on ambient air quality and cleaner air for Europe and Directive (EU) 2015/1480 of 28 August 2015 on reference methods, data validation and location of sampling points for the assessment of ambient air quality.

Fishbain, B., & Moreno-Centeno, E. (2016). Self Calibrated Wireless Distributed Environmental Sensory Networks. Scientific Reports, 6, 24382. doi:10.1038/srep24382

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann R. and Ian H. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations 11(1), pp. 10-18. Doi: 10.1145/1656274.1656278

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter,11*(1), 10. doi:10.1145/1656274.1656278

Ho, T. K. (1995). Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition. doi:10.1109/icdar.1995.598994

JCGM (2008). Evaluation of Measurement Data d Guide to the Expression of Uncertainty in Measurement.

Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., . . . Buckley, K. (2016). Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. Atmospheric Measurement Techniques,9(11), 5281-5292. doi:10.5194/amt-9-5281-2016

Kamionka, M., Breuil, P., & Pijolat, C. (2006). Calibration of a multivariate gas sensing device for atmospheric pollution measurement. Sensors and Actuators B: Chemical, 118(1-2), 323-327. doi:10.1016/j.snb.2006.04.058

Kotsev, A., Schade, S., Craglia, M., Gerboles, M., Spinelle, L., & Signorini, M. (2016). Next Generation Air Quality Platform: Openness and Interoperability for the Internet of Things. Sensors (Basel, Switzerland), 16(3), 403. http://doi.org/10.3390/s16030403

Marco, S., & Gutierrez-Galvez, A. (2012). Signal and Data Processing for Machine Olfaction and Chemical Sensing: A Review. IEEE Sensors Journal, 12(11), 3189-3214. doi:10.1109/jsen.2012.2192920

Mead, M.I., Popoola, O. a. M., Stewart, G.B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J.J., McLeod, M.W., Hodgson, T.F., Dicks, J., Lewis, a., Cohen, J., Baron, R., Saffell, J.R., Jones, R.L., 2013. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos. Environ. 70, 186–203. doi:10.1016/j.atmosenv.2012.11.060

Pascal, M., Corso, M., Chanel, O., Declercq, C., Badaloni, C., Cesaroni, G., . . . Medina, S. (2013). Assessing the public health impacts of urban air pollution in 25 European cities: Results of the Aphekom project. Science of The Total Environment, 449, 390-400. doi:10.1016/j.scitotenv.2013.01.077

Popoola, O., Mead, I., Bright, V., Baron, R., Saffell, J., Stewart, G., Kaye, P., Jones, R., 2013. A Portable Low-Cost High Density Sensor Network for Air Quality at London Heathrow Airport. EGU General Assembly 2013, held 7-12 April, 2013 in Vienna, Austria, id. EGU2013-1907 http://wwwdev.snaq.org/posters/EGU_OAMP_2013.pdf

Popoola, O. A., Stewart, G. B., Mead, M. I., & Jones, R. L. (2016). Development of a baseline-temperature correction methodology for electrochemical sensors and its implications for long-term stability. Atmospheric Environment, 147, 330-343. doi:10.1016/j.atmosenv.2016.10.024

Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz, W. A., & Bartonova, A. (2017). Mapping urban air quality in near real-time using observations from low-cost sensors and model information. Environment International, 106, 234-247. doi:10.1016/j.envint.2017.05.005

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2015). Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. Sensors and Actuators B: Chemical, 215, 249-257. doi:10.1016/j.snb.2015.03.031

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2017). Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO 2. Sensors and Actuators B: Chemical, 238, 706-715. doi:10.1016/j.snb.2016.07.036

Tsujita, W., Yoshino, A., Ishida, H., & Moriizumi, T. (2005). Gas sensor network for air-pollution monitoring. Sensors and Actuators B: Chemical, 110(2), 304-311. doi:10.1016/j.snb.2005.02.008

van der Maaten L.J.P and Hinton G.e. (2008). Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research 9 (Nov):2579-2605

Vidnerova, P., & Neruda, R. (2016). Sensor Data Air Pollution Prediction by Kernel Models. 2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid). doi:10.1109/ccgrid.2016.80

W. Tsujita, A. Yoshino, H. Ishida, T. Moriizumi, Gas sensor network for air-pollution monitoring, Sens. Actuators B: Chem., 110 (2005), pp. 304–311

Webb et al., (2005), Statistical Pattern Recognition, Wiley, 2005

WHO, 2013, Review of evidence on health aspects of air pollution — REVIHAAP Project, Technical Report, World Health Organization, Regional Office for Europe, Copenhagen

WHO, 2018 - Ambient (outdoor) air quality and health. Fact sheet. 2 May 2018.

Wu, S., Ni, Y., Li, H., Pan, L., Yang, D., Baccarelli, A. A., . . . Guo, X. (2016). Short-term exposure to high ambient air pollution increases airway inflammation and respiratory symptoms in chronic obstructive pulmonary disease patients in Beijing, China. Environment International, 94, 76-82. doi:10.1016/j.envint.2016.05.004

Highlights

- Several air quality microsensors were tested against reference methods
- Improved correlation between CO, NO_2, O_3, PM10, PM2.5, SO_2 sensors and reference methods through calibration with machine learning techniques
- Parameter dependencies and measurement uncertainty of sensors were evaluated
- Possibility of compliance with DQO of the AQD for indicative measurements
- Microsensors can improve spatiotemporal data resolution to complement current monitoring networks