

Optimal Field Calibration of Multiple IoT Low Cost Air Quality Monitors: Setup and Results

E. Esposito[1(✉)], G. D'Elia[1,2], S. Ferlito[1], A. Del Giudice[1], G. Fattoruso[1], P. D'Auria[3], S. De Vito[1], and G. Di Francia[1]

[1] ENEA, DTE-FSD-SAFS, P. le E. Fermi, 1, 80055 Portici, NA, Italy
elena.esposito@enea.it
[2] Department of Industrial Engineering (DIIn), University of Salerno, via Giovanni Paolo II, 132, Fisciano, SA, Italy
[3] ARPAC, Campania Regional Agency for Environmental Protection, Torre 1, CDN, 80143 Naples, Italy

Abstract. The assessment of Low Cost Air Quality Multisensor Systems (LCAQMS) performance is a crucial issue in the Air Quality (AQ) monitoring framework. The devices calibration model is one of the most important drivers of the overall performances. As we know, on field calibration is increasingly considered as the best performing approach for air quality monitor devices. Field recorded sensor data together with co-located reference data allow to build suitable datasets that are more representative of the complexity of real world conditions. In this work a co-location experiment is presented, in which four multisensor devices are co-located with a mobile ARPAC (Campania Regional Agency for Environmental Protection) reference analyzer station. Two types of calibration models, linear and nonlinear have been tested on the recorded datasets, in order to determine the best calibration strategy to use to optimize the calibration procedure time in the real world operative phase. The results show that for pervasive AQ scenario, a reasonable choice is provided by a multilinear approach during one-week short co-location period.

Keywords: Air Quality Multisensor Systems · Sensor calibration · Backup reference data

1 Introduction

The next generation of air quality monitoring network will rely on the integration among regulatory grade analyzers and IoT based smart chemical/particulate multisensor devices [1]. The former will provide a backbone of sparse but high reliability, high quality, measurements at a significant procurement and operational costs. Smart multisensor devices will provide high resolution and possibly redundant measurements with affordable costs but with reduced precision and accuracy. Data assimilation models will be then capable to integrate the stream of incoming data with emissions or regression factors and weather data to produce high resolution air quality maps capable to produce informed remediation decisions. PM and NO_2 are foreseen to be the most problematic pollutants since

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12253, pp. 700–708, 2020.
https://doi.org/10.1007/978-3-030-58814-4_57

most of urban chemical pollution is constantly decreasing, in high GDP countries, over the past 20 yrs. Smart air quality monitors, either developed for fixed or mobile deployment, suffer from intrinsic issues linked with their sensors transduction mechanism. Low cost gas sensors, whichever technology they rely on, are ultimately hampered by lack of sensitivity, selectivity and stability. Most of them have they response be significantly influenced by so called interferents being other pollutants or just environmental variables, specifically Temperature and Relative Humidity [2]. In the long term their response is influenced by slow poisoning or just ageing of their components, including the active surface. Nonlinear behaviour can be observed in some of the relevant technology family. Field calibration relying on statistical or, more generally, machine learning models seems the only viable and feasible method to guarantee the short term accuracy and precision of these systems [3, 4]. Although its robustness to long term deployment and so different environmental and pollution composition is criticized [5–7] it allows to rapidly and so cheaply expose the sensors grabbing their response to a variety of (uncontrollable) condition that are similar to the ones that will be encountered during operational life in opposition with lab based calibration that would need significant time and human efforts to achieve similar variability in controlled settings. As shown by several recent works, long term operational deployment may keep the accuracy level if enjoying a stream of high accuracy data by either *rendez-vous* with fellow nodes, periodic visit to regulatory grade analyzers in the mobile case or projected in their position through an adequate model by nearby stations [8–10]. One of the major challenges in field calibration is the optimal choice of the calibration model and the length for the colocation period [11]. Additionally, we still have no clue of the variability of the performance indices among different calibrated mobile analyzers that are known to exhibit significant fabrication variability. In this contribution, we analyse this issue by a field calibration experiment lasting two months using a mobile regulatory grade analyser deployed in the city of Portici located nearby Naples in the south of Italy during the first months of 2020. The experiment is part of the efforts of the AIR-Heritage project, an Urban Innovative Action EU funded project which foresee the deployment of tenths of mobile analyzers to be used by citizens [12] during their daily mobility routines so to contribute AQ ground data to high resolution mapping facilities.

2 Experimental Framework

2.1 IoT AQMS Station Architecture

In this study, we rely on the use of the ENEA MONICA AQMS, designed for cooperative mobile air quality quantitative sensing operations. MONICA device is based on electrochemical sensors array using Alphasense™ A4 class sensor units, respectively targeted to Carbon Monoxide (CO), Nitrogen Dioxide (NO_2) and Ozone (O_3). Relative Humidity (RH) and Temperature (T) sensors complete the sensing array. The sensors analog front-end is provided by the same company and allows to connect sensors to an ARM microcontroller based ST Microelectronic Nucleo board. The latter captures and digitalize the two relevant sensors terminal voltages of each sensor, namely Working Electrode (WE) and Auxiliary Electrode (AE), along with the temperature and RH readings from a SHT11 sensor device. In particular, AE readings may be used to partially

correct for temperature interference affecting these sensors WE. The effect of temperature on the different electrodes readings are different and temperature still affect their difference (WE-AE), due to their particular geometry and manufacturing difficulties (see Sensors Datasheet in reference [13]). Raw sensors data are captured by a 10-bit ADC and transmitted via a Bluetooth serial interface to a Raspberry Pi Mod. 3 + based datasink with Raspbian OS providing for local storage and WAN connectivity services through a mobile router wi-fi tplink M7650. Data is captured at 15 samples/minute rate. At remote side, an ad-hoc IoT backend architecture relying on a contained NodeJs REST APIs server and MongoDB provides data inception, device management, storage, preprocessing and map based visualization functionalities (see Fig. 1).

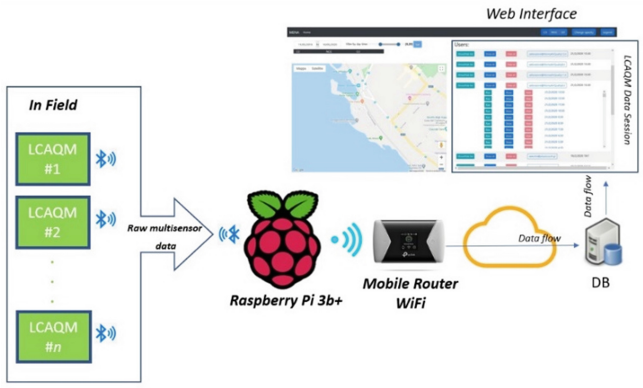

Fig. 1. Scheme of IoT AQMS architecture.

2.2 Recorded Datasets Description

Four MONICA devices (since now on identified as AQ6, AQ8, AQ11, AQ12) have been located on a mobile ARPAC station in a Portici city area, for two months (January 2 to March 2, 2020). In Fig. 2 the co-location architecture is shown.

The recorded datasets consist of 1440 h captured in a continuous sampling fashion. Specifically, for each node, two datasets, with samples averaged at minute and hourly rate, have been built. These datasets contain averaged data from each of sensors embedded into the device, i.e. WE and AE raw sensors readings (mV) for NO_2, CO, O_3 targeted sensors plus T (°C) and RH (%), joined to same time scale averaged data from a mobile ARPAC reference analyzer for NO_2 ($\mu g/m^3$), CO (mg/m^3), O_3 ($\mu g/m^3$). Table 1 resumes the acquired data from the 4 co-located nodes and data losses. Recorded data have been preprocessed, analyzing the missing values, detecting the possible outlier's carrying out a correlation analysis.

The analysis of the ARPAC validated data, collected during the co-location period shows a cyclical and long-term trend, according to the emission characteristics of the site. The significant decrease in average hourly NO_2 concentrations in the time series is due to the wind influence, with different intensity and direction (N-NW and N-NE). This allows



Fig. 2. MONICA AQMS network co-located with ARPAC mobile reference station in Portici city area.

Table 1. Recorded data and data losses.

Monica node	Acquired data (hours/minutes)
AQ6	1432 h/83990 min
AQ8	1392 h/81854 min
AQ11	1422 h/81460 min
AQ12	1268 h/73832 min

the dispersion of pollutants of emission origin. In Fig. 3, four similar meteorological conditions occurring during the co-location period are highlighted by ellipses (Fig. 4).

2.3 Sensor Calibration Models

It is now common knowledge that chemical sensors array raw data needs to be processed by a calibration function to accurately and precisely estimate target gases concentration taking care of nonlinearities and interferents. Several and extensive on-field experiments along with theoretical results led to select two main calibration approaches, each of which has shown to be suitable in specific conditions: Multiple Linear Regression model (MLR) and Shallow Neural Network model (SNN).

Assuming that X is the input features and y the predicted value, the MLR model is the classic linear regression with multiple input features, mathematically expressed by

$$y = X\beta + c \tag{1}$$

where c is the intercept. The selected nonlinear model is SNN, that has already proven very efficient for AQMS on field calibration. The analyzed SNN model is a three layers' architecture, empirically equipped with [3, 5, 7] standard sigmoidal tangent neurons units in the hidden layer and a linear output layer. Automatic Bayesian Regularization (ABS) was used as training algorithm. In this study, we focused on NO_2 hourly averaged concentration estimation problem using hourly averaged WE and AE sensors data of

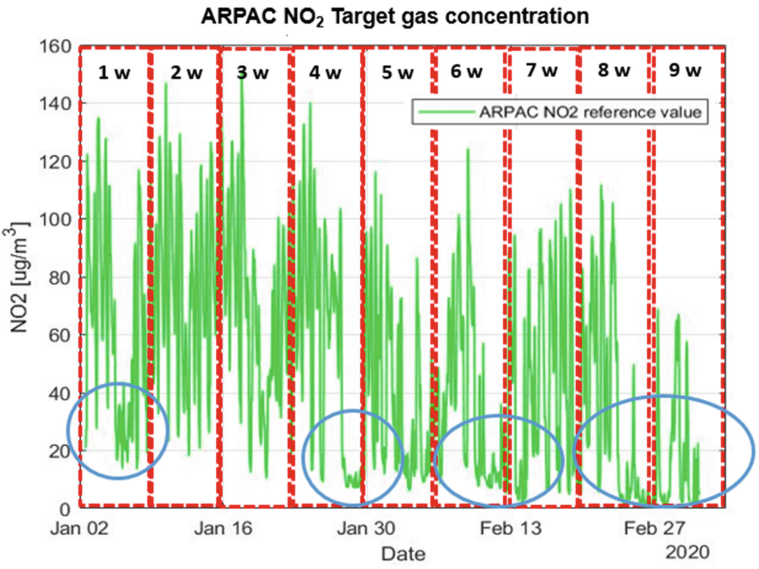

Fig. 3. NO_2 gas concentration provided by mobile ARPAC reference analyzer. It can be observed the gas behavior during the entire co-location period and also the daily trend, depending on the meteorological conditions.

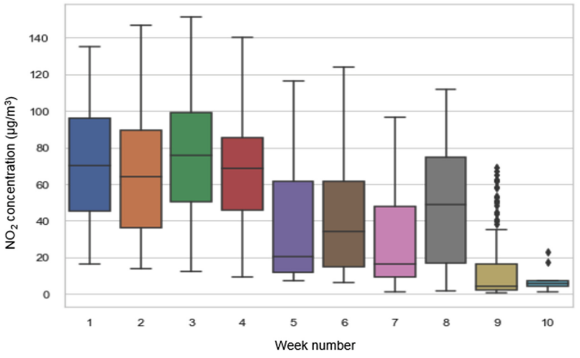

Fig. 4. Box-plot representation of NO_2 gas concentration distribution along the entire co-location period.

NO_2, O_3 sensors plus T and RH data as inputs for the two calibration algorithms. The input matrix X, thus, consists of 6 features vectors as columns (WE_NO_2, AE_NO_2, WE_O_3, AE_O_3, T, RH) and the rows number depends on the training set length. The two calibration algorithms have been tested using different choices of training/validation/test sets. In particular, calibration is performed in an ex-post fashion by selecting for testing purposes only those samples that are temporally located after all the data used for training and validation purposes. This setting is the most adequate to simulate real conditions

where nodes will be operated after the calibration took place. Details and results are reported in Sect. 3.

3 Results

As previously mentioned, performance assessment experiments have been carried out using different training/validation/test sets combinations, aiming to the optimization of the involved parameters. Table 2 captures the preliminary results of our experimentation. At a glance, it is possible to spot the different performance obtained by the different analyzers. Irrespective of the calibration model and at each training/validation/test combination AQ8 node appear as the worst performing node. AQ12 node seems instead to express the best performance with respect to all the other nodes particularly when calibrated using an MLR approach. NN model performances are hampered when using data from second week; actually, without resorting to a validation set the learning process result in overtraining conditions that prevent the network to obtain good generalization capabilities. Generally, results obtained by MLR and NN models appear similar with MLR keeping a limited edge on the performance obtained by NN.

Table 2. Mean absolute errors obtained using two calibration procedures with different choices of operational and test sets for each AQMS node.

Training + Validation setting	Mean Absolute Error (MAE) [$\mu g/m^3$]							
	AQ6		AQ8		AQ11		AQ12	
	NN	MLR	NN	MLR	NN	MLR	NN	MLR
1st w + 0 w	11.70	7.94	17.42	23.36	8.20	7.78	11.45	**6.56**
1st w + 2nd w	7.15	7.70	24.04	16.79	10.69	9.51	8.27	**6.92**
2nd w + 0 w	***54.46***	***9.55***	***92.41***	***21.92***	***64.26***	***10.37***	***97.68***	***8.21***
1–2nd w + 3rd w	*8.65*	*7.73*	*17.63*	*13.30*	*11.17*	*8.86*	*8.29*	***6.50***
1–3rd w + 4th w	8.13	7.56	15.98	12.63	9.50	9.89	6.63	**6.31**
1–4th w + 5th w	7.44	7.63	15.57	11.37	10.38	9.66	6.53	**5.15**

The first two experiments have been performed using the first week for the training phase (1st and 2nd row in Table 2), while in the second experiment we use only the second week for the training (3rd row in Table 2). We can observe that the NN performances decrease when the meteorological conditions change in the operative phase wrt the training phase (result highlighted in blue in Table 2). Then, we compared these results with those obtained incrementally enlarging the training set up to one month (rows 4–6 of Table 2). The obtained results show that no significant improvements have been achieved extending the co-location period. Figure 5 depicts the behavior of NN model based estimations along with true concentrations for all the four nodes under calibration when using the first two weeks of data for training purposes and the third for validation (results highlighted in Italic in Table 2). It is easy to spot the sudden

performance worsening attained at the end of the colocation period when concentration of the target pollutant is significantly lower than those encountered during training and validation period. Performance worsening are more evident for node AQ8 confirming its low capability to accurately estimate the target gas concentrations.

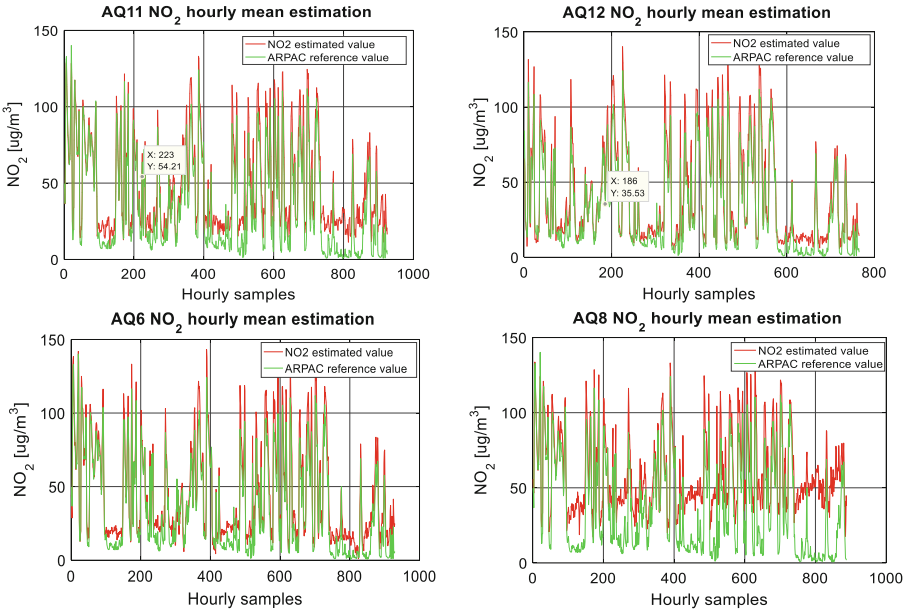

Fig. 5. NO_2 hourly concentration estimations versus target gas concentration starting from the 4th week of the co-location period.

By focusing on Fig. 6 the underlying drivers of the poor performance of AQ8 node become evident with a significant bias error found along all the target concentration range and dramatic lack of precision when dealing with low target gas concentrations. A careful and continuous estimation of performance of the nodes can be helpful to rapidly detect poorly performing node and freeing its calibration spot to accommodate a new node. When benefitting from a continuous access to ground truth data significant differences in the linear fit parameters between node estimations may establish a fast and robust way to determine anomalies in sensors operative performances (see Fig. 7).

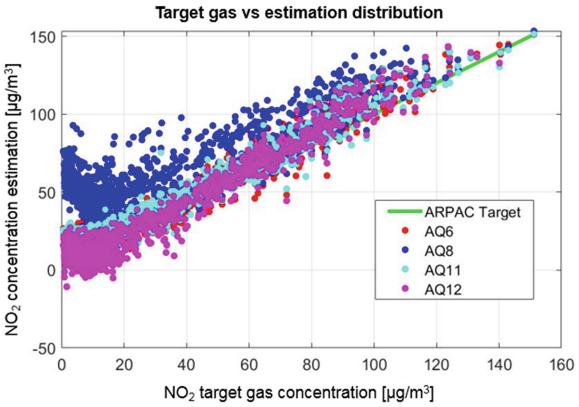

Fig. 6. NO_2 gas concentration estimation computed for each node, along the entire co-location period versus target gas concentration line. The differences among the four sensors performances become apparent when considering low true concentration of the target pollutant.

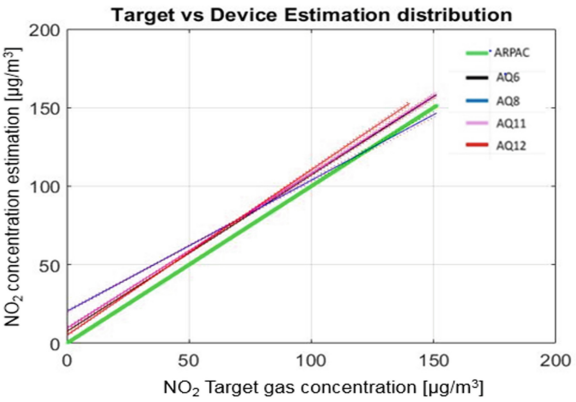

Fig. 7. Differences in linear fit parameters among the difference nodes highlight the malfunctioning node.

4 Conclusions

A mid-term colocation experiment has been carried out for the purpose of analyzing and comparing the performance of calibrated smart air quality monitoring nodes. The analysis of the dependence of performances from calibration extension did not highlighted a significant correlation among training duration and performances with good performances obtained by considered model starting as early as just one week of calibration data. Performance seems consistent among the two considered model while sustained performance variance has been recorded among a node and the majority of the others. In fact, while three of the node undergoing the calibration process expressed similar performances while the remaining one suffered both from high amount of bias and lack of precision and accuracy at lowest end of the target gas observed concentration range.

Further works will include the analysis of performances along an extended duration as well as the increase of the robustness of the recorded performance indexes by averaging on different combination of calibration set start, i.e. by adequate resampling process.

Acknowledgments. This work has received funding by EU through UIA 3rdcall Project "AirHeritage".

References

1. Schneider, P., Castell, N., Vogt, M., Dauge, F.R., Lahoz, W.A., Bartonova, A.: Mapping urban air quality in near real-time using observations from low-cost sensors and model information. Environ. Int. **106**, 234–247 (2017). ISSN 0160-4120
2. Cross, E.S., et al.: Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements Atmos. Meas. Technol. **10**, 3575–3588 (2017)
3. Zimmerman, N., et al.: A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. Atmos. Meas. Technol. **11**, 291–313 (2018)
4. Hagan, D.H., et al.: Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments. Atmos. Meas. Technol. **11**, 315–328 (2018)
5. Masey, N., et al.: Temporal changes in field calibration relationships for Aeroqual S500 O3 and NO2 sensor-based monitors. Sens. Actuators B Chem. **273**, 1800–1806 (2018)
6. De Vito, S., Esposito, E., Castell, N., Schneider, P., Bartonova, A.: On the robustness of field calibration for smart air quality monitors. Sens. Actuators B Chem. **310**, 127869 (2020). https://doi.org/10.1016/j.snb.2020.127869. ISSN 0925-4005
7. Lewis, A.C., Edwards, P.: Validate personal air-pollution sensors. Nature **535**, 29–31 (2016)
8. Kizel, F., et al.: Node-to-node field calibration of wireless distributed air pollution sensor network. Environ. Pollut. **233**, 900–909 (2018)
9. De Vito, S., Di Francia, G., Esposito, E., Ferlito, S., Formisano, F., Massera, E.: Adaptive machine learning strategies for network calibration of IoT smart air quality monitoring devices. arXiv preprint arXiv:2003.12011 (2020)
10. Miskell, G., et al.: Reliable data from low cost ozone sensors in a hierarchical network. Atmos. Environ. **214**, 116870 (2019). https://doi.org/10.1016/j.atmosenv.2019.116870. ISSN 1352-2310
11. Esposito, E., De Vito, S., Salvato, M., Bright, V., Jones, R.L., Popoola, O.: Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. Sens. Actuators B Chem. **231**, 701–713 (2016)
12. Castell, N., Liu, H.-Y., Schneider, P., Cole-Hunter, T., Lahoz, W., Bartonova, A.: Towards a personalized environmental health information service using low-cost sensors and crowdsourcing EGU general assembly (2015)
13. NO2-A4 datasheet downloaded in February 2020 from Alphasense ltd website. www.alphasense.com