Multi-sensor data fusion calibration in IoT air pollution platforms

Pau Ferrer-Cid, Jose M. Barcelo-Ordinas, Jorge Garcia-Vidal, Anna Ripoll, Mar Viana

Abstract—This paper investigates the calibration of low-cost sensors for air pollution. The sensors were deployed on three IoT (Internet of Things) platforms in Spain, Austria, and Italy during the summers of 2017, 2018, and 2019. One of the biggest challenges in the operation of an IoT platform, which has a great impact on the quality of the reported pollution values, is the calibration of the sensors in an uncontrolled environment. This calibration is performed using arrays of sensors that measure cross sensitivities and therefore compensate for both interfering contaminants and environmental conditions. The paper investigates how the fusion of data taken by sensor arrays can improve the calibration process. In particular, calibration with sensor arrays, multi-sensor data fusion calibration with weighted averages, and multi-sensor data fusion calibration with machine learning models are compared. Calibration is evaluated by combining data from various sensors with linear and nonlinear regression models.

Index Terms—IoT platform, sensor calibration, multi-sensor data fusion, low-cost sensors, uncontrolled environments, machine learning models, correlated errors.

I. INTRODUCTION

IN recent years there has been a growing interest in deploying Internet of Things (IoT) platforms in various fields such as agriculture, wildlife tracking, tracking people or air pollution monitoring [1], [2], [3], [4]. A large part of the research and published literature has focused on communication protocols and the reduction of energy consumption, as well as the transmission of data to repositories where this information is analyzed [5], [6], [7], while less attention has been paid to the quality of the data received [4], [8].

One of the biggest challenges in these platforms is to guarantee the quality of the information they collect. Depending on the field of application, the low-cost sensors that are incorporated into the wireless nodes that make up the network have better or worse quality. One of the fields in which low-cost sensors have a quality that is still being tested is the sensors that measure air pollution [4], [8]. Sensors for nitrogen dioxide (NO$_2$), nitrogen monoxide (NO), tropospheric ozone (O$_3$) or particle matter (PM2.5, PM5), among others, are generally not calibrated by the manufacturer, and if they have been, they have not been calibrated in the environmental conditions in

Pau Ferrer-Cid (pferrer@ac.upc.edu), Jose M. Barcelo-Ordinas (joseb@ac.upc.edu) and Jorge Garcia-Vidal (jorge@ac.upc.edu) are with the Universitat Politecnica de Catalunya, Barcelona, Spain.

Anna Ripoll (annaripollroca@gmail.com) is with 4sfera Innova, Girona, Spain

Mar Viana (mar.viana@idaea.csic.es) is with the Institute for Environmental Assessment and Water Research, Spanish National Research Council (IDAEA-CSIC), Barcelona, Spain

which they will provide measurement data. This is the cause of major issues for air quality managers because sensor data quality cannot be assessed and, despite this, the data are often communicated through citizen science initiatives and reported as reference values. This lack of quality control has led to the negative perception of sensor data by regulatory bodies, and driven their use with extreme caution by the scientific community. As a result, there is a major need for strategies to test and validate sensor data for air quality monitoring [9].

In the European H2020 CAPTOR project (2016-2018), three testbeds in Spain, Italy and Austria were deployed for tropospheric ozone monitoring using low-cost sensors mounted on wireless nodes [10]. Tropospheric ozone (O$_3$) is produced by the reaction of NOx and volatile components with solar radiation. Although O$_3$ is produced in large cities, it tends to spread with the wind and affect non-urban areas, having a major impact on agriculture and people. The deployed wireless nodes included sensors of O$_3$, NO$_2$, temperature and relative humidity. In addition, the sensors deployed included metal-oxide (MOX) and electro-chemical (EC) based sensor technologies. A total of sixty-one nodes were deployed, one hundred and sixty-six sensors of O$_3$, twenty-six of NO$_2$, sixty-one of temperature and relative humidity. The sensors underwent a three-week calibration process followed by a three-month deployment in volunteer homes. The data has been sent to a cloud platform, where it has been processed and made accessible to citizens through a mobile application. Thus, one of the objectives of the project was to make citizens aware of the impact of pollution, and in particular tropospheric ozone, on people. Several nodes were deployed at reference stations at all times to perform theoretical studies to better understand the calibration process, and thus improve the accuracy of contaminant concentration measurements with low-cost sensors.

In recent years, there has been a growing interest in comparing and evaluating various calibration algorithms such as multiple linear regression [11], [12], [13], [14], K-nearest neighbors [14], [15], support vector regression [13], [16] or random forest [13], [15], [17]. Most of these papers use nodes that mount an array of sensors since many air pollutants are directly or inversely related to other pollutants (e.g. ozone is inversely related to nitrogen oxide due to titration) or to environmental parameters (temperature and relative humidity) [11], [18], [19]. The results of these studies indicate that the air pollution sensor technology is still not mature enough for the sensors to give results with high accuracy. On the other hand, since the cost of the sensors is low, it is possible to mount several sensors that measure the same pollutant in the

same node to add robustness and fault tolerance [10]. This approach adds redundancy to the data captured by the sensor, which can improve the quality of the observations taken.

The aim of this paper is to study calibration mechanisms that take advantage of this redundancy to improve the quality of the final observations. These mechanisms will contribute to improving sensor data quality and facilitate their deployment for citizen communication and awareness purposes with regard to air quality. To this end, we study the use of multi-sensor data fusion techniques for calibrating air pollution sensors. In particular, in this paper, we:

- describe the data set taken by the IoT platform deployed with nodes mounting array of sensors, including O_3, NO_2, temperature and relative humidity sensors,
- determine multi-array sensor calibration for metal-oxide and electro-chemical O_3 sensors taking into account environmental conditions (temperature and relative humidity),
- define a multi-sensor data fusion framework. In particular, we will compare calibration with array of sensors, multi-sensor data fusion calibration with weighted averages, and multi-sensor data fusion calibration with machine learning models,
- compare the multi-sensor data fusion approach using several regression algorithms such as multiple linear regression, K-nearest neighbors, support vector regression and random forest,
- analyse the multi-fusion schemes for calibrating O_3 using a set of O_3 of the same or different sensor technologies: multi-sensor data fusion with only metal-oxide sensor technologies and multi-sensor data fusion with metal-oxide and electro-chemical sensor technologies.

The paper is organized as follows: in section II the related work is mentioned. Section III describes the data collection platform, including the collection nodes, the sensor technologies and the data sets obtained for calibration. Section IV discusses the calibration methods used and how the multi-sensor data fusion approach works, while section V shows the results of the multi-sensor data fusion calibration. Finally, section VI concludes the paper.

II. Related work

Sensor fusion, defined as data aggregation [20], in wireless sensor networks (WSN) has been studied in great detail in the literature. In this context, the aim is to summarize sensor measurement data to reduce the amount of messages transmitted, and the energy consumption [21]. In the context of using inertial units of measurement (IMU) sensors for orientation or activity detection, it has been proposed to fuse sensors using complementary filters or Kalman filters in order to obtain complementary information provided by different types of sensors [22]. Wu et al. [23] presents a hierarchical architecture of a system for estimating the activity of daily living of individuals from the fusion of different smartphone sensors, where the fusion of sensors by time-series is exploited in several stages of the system. In other contexts, the goal of using data fusion in multi-sensor environments [24], [25] is to obtain a lower detection error probability and a higher reliability by using data from multiple distributed sources.

In a survey paper focused on calibration in uncontrolled environments, Barcelo-Ordinas et al. [8] define the term of multi-sensor data fusion in the calibration context as the "combination of information from two or more data sources (sensors) into a single one that provides a more accurate description than that of any of the individual data sources". The concept is that in this context of calibration, a number of sensors (called *array of sensors*) participate in the calibration of the target sensor to reduce calibration errors. The sensor array principle is widely used [4], [11], [13], [17], [18], [19], [26] in the calibration of air pollution sensors, where the sensor calibration consists of measuring all cross sensitivities of the sensor array to compensate for all interfering contaminants and environmental conditions. Different calibration methods covering linear and non-linear models have been used with arrays of sensors depending on the type of contaminant and cross sensitivities. Multiple linear regression [11], [12], [13], [14] is used for calibration of O_3, NO_2, CO and PM sensors among others. In other cases, because the sensor does not behave linearly, non-linear models have been investigated for the calibration of low-cost sensors: K-nearest neighbors [14], [15], support vector regression [13], [16] or random forest [13], [15], [17].

Multi-sensor data fusion techniques are well known in many applications [27], [28], and specifically the use of multi-sensor data fusion techniques with weighted averages in ultrasonic and infrared sensors [29] with uncorrelated errors or in target tracking applications [30] with correlated errors. P. Avery [31] obtains the optimal weights for the special case of having sensors with correlated errors. Another different approach to multi-sensor data fusion in air pollution sensors is presented in [12], where the use of a multi-sensor data fusion using machine learning between the same family of sensors is introduced to improve the performance of a multiple linear regression model. Thus, it is observed that the inclusion of four O_3 sensors can improve the quality of the model that only comprises one sensor. This is the topic that is studied in more detail in this work; if the use of several sensors in a node that measure the same phenomenon can result in a significant improvement in the calibration model.

III. Testbeds and data set

The H2020 Captor project arose from the need to raise awareness about ground-level ozone pollution and its consequences for people. Three IoT testbeds were installed in Spain, Italy and Austria [10]. The data collection campaigns lasted all summer, from May to October, as O_3 is a seasonal pollutant in Europe that is formed as a reaction of NO_x and solar irradiation. The deployed wireless sensor network consisted of thirty-five nodes with a total of one-hundred-forty O_3 (4 per node) MICS 2614 metal oxide sensors, thirty-five temperature and relative humidity sensors. In addition, twenty-six other nodes were deployed with a total of twenty-six electro-chemical sensors O_3 Alphasense OX-B431 and twenty-six temperature and relative humidity sensors. As the OX-B431 sensors measure O_3 plus NO_2, it was necessary to add to each of these nodes an electro-chemical Alphasense

NO2-B43F sensor to compensate the measurement of NO_2. The data taken by the nodes deployed in the testbeds were sent to a repository using a cellular connection. The nodes transmitted the sensor data with a time stamp and a node identifier every half hour, therefore there were no bandwidth or latency restrictions.

Twenty-five nodes with MOX sensors and one with EC sensors were deployed on the Spanish platform, ten nodes with MOX and another ten nodes with EC sensors on the Italian platform and fifteen nodes with EC sensors on the Austrian platform. Areas with high ozone concentrations were chosen and which had at least one reference station managed by a government organization with which to compare and calibrate the sensors. The nodes were divided into two groups that produced two types of data sets[1]. The first group is composed of those nodes that were deployed all summer in volunteer houses. The sensors were calibrated for at least two weeks by placing them in a reference station near the volunteers' homes. Therefore this data set has corresponding samples between two or three weeks of measurements. The second group of nodes were placed in a reference station throughout the campaign. Therefore the data set produced is composed of samples corresponding to between eighteen and twenty weeks depending on the node.

To perform the multi-sensor data fusion calibration study, we need sensors that have been in the same location during the same period. Therefore, we built five data sets, Table I, extracted from the original data sets organized by having sensor samples at the same times at the same reference station. Data sets 1 and 2 correspond to nodes characterized by having a large number of sensors with MOX technology placed in the same reference station. Data sets 3, 4 and 5 correspond to reference station locations where nodes with different technologies, MOX and EC, coincided.

IV. SENSOR CALIBRATION

Calibration techniques in air pollution include linear and non-linear methods. The simpler one is multiple linear regression while the non-linear include *instance-based* methods such as K-nearest neighbors, *ensemble* methods such as random forest and *kernel* methods such as support vector regression.

Multiple linear regression (MLR) is the extension of the simple linear regression when dealing with more than one feature. It approximates the real function with a linear approximation where the regression coefficients are obtained from the normal equations. K-nearest neighbors (KNN) is based on all the training data. The output for a new observation will be the average of the outputs for the k closest training observations in feature space to the new observation. Random forest (RF) combines several decision trees which have been de-correlated by using bootstrap sampling. Then, the response for a new observation will be the average of the outputs of the different decision trees. Support vector regression (SVR) makes use of the "kernel trick" to perform a regression problem in a feature space of larger dimension than the input space.

[1]The raw data and calibration software is available at http://sans.ac.upc.edu/?q=node/231

A. Sensor array calibration

The calibration system of an array of sensors for calibrating an ozone sensor, Figure 1, consists of feeding the model with the raw data obtained by the ozone, temperature and relative humidity sensors $\{s, s_{Temp}, s_{RH}\}$ and by the reference station $y_{RefStat}$ value over a training set. This training phase produces a set of calibration coefficients or hyper-parameters. Temperature and relative humidity sensors are used in the calibration model to correct the ozone sensor signal [4]. Each calibration model (MLR, KNN, RF, SVR) produces a set of calibration coefficients or hyper-parameters that allow calibrated data to be obtained on new tuples $\{s, s_{Temp}, s_{RH}\}$ of raw data that we call uncalibrated or test data. In the case of nonlinear models, a 10-fold cross-validation procedure is applied with the training data to select the best set of hyper-parameters for each nonlinear method. Finally, a hold-out test set is used to see the performance of the different models. The outcome of the process, applied to new tuples of raw data, is an ozone value, y_{cal}, that represents the ozone sensor involved in the calibration process.
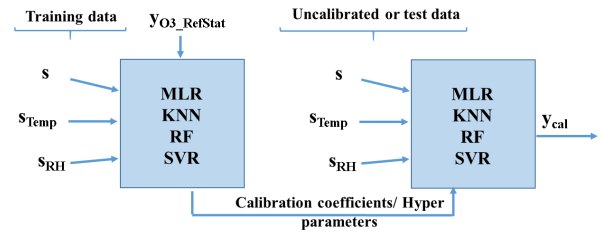

Fig. 1. Sensor array calibration model for ozone sensors.

B. Multi-sensor data fusion calibration with weighted averages

Fusion with weighted averages consists of finding a linear estimator that aggregates the measurements of all the sensors. To do this, a weight w_i such that $\sum_i w_i = 1$, is assigned to each calibrated individual y_{cal_i} sensor and the fusion is the result of the sum of each calibrated individual value y_{cal_i} weighted by its corresponding weight w_i:

$$y_{cal} = \sum_i^n w_i \, y_{cal\_i} \tag{1}$$

In the case that the measurements are independent, and therefore the errors ϵ_i between the sensors are independent, it is easy to check that the optimal weights are [31]:

$$w_i = \frac{(\sigma_i^2)^{-1}}{\sum_j^n (\sigma_j^2)^{-1}} \tag{2}$$

Where σ_i^2 is the variance of the sensor error i, with $i=1 \ldots n$, and n the number of air pollution sensors involved. This linear estimator is the best in the sense that it is unbiased and efficient [31]. However, in the general case that the sensor measurements are correlated, the errors will be correlated and we have to include these correlations in the model. Define V as the covariance matrix between the ϵ_i errors in each

TABLE I
DATA SETS USED IN THE MULTI-SENSOR DATA FUSION CALIBRATION.

Data Set Label	Reference St. Name	Year	Node Labels	Technologies	# Sensors	# of Samples
1	Manlleu	2017	C17001, C17002, C17003, C17005, C17010, C17011, C17013	MOX	28x MICS 2614	918
2	Tona	2017	C17006, C17007, C17012, C17014, C17017, C17027	MOX	24x MICS 2614	1395
3	Tona	2017	R69-17 C17017	EC MOX	1x OX-B431 + 1x NO2-B43F 4x MICS 2614	2366
4	Tona	2018	R69-18 C18017	EC MOX	1x OX-B431 + 1x NO2-B43F 4x MICS 2614	933
5	Palau Real	2019	C19000 C19027	EC MOX	1x OX-B431 + 1x NO2-B43F 4x MICS 2614	1179

sensor. If $y_{RefStat}$ is the reference value and y_{cal_i} is the calibrated value in sensor i, $\epsilon_i = y_{RefStat} - y_{cal\_i}$. We call V_{ij} the covariance between the errors at sensor i and the errors at sensor j, and $V_{ii} = \sigma_i^2$ as the error variance at sensor i. Then, it can be shown that the optimal weights are [31]:

$$w_i = \frac{\sum_j^n (V_{ij})^{-1}}{\sum_j^n \sum_k^n (V_{jk})^{-1}} \quad (3)$$

If we apply this model to our case, first, we train each sensor as if it were a calibration with sensor arrays, Figure 2. The outputs of each of these modules are (i) $y_i^{training}$, with $i=1,\ldots,n$, of the training data, and (ii) the coefficients that will allow the values of the test data set to be obtained. Second, get the covariance matrix V of the errors $\epsilon_i = y_i^{training} - y_{RefStat}$ between the data calibrated with the training set and the reference values and calculate the weights w_i. Finally, after obtaining the calibrated values over the test data set y_{cal_i}, weight them with weights w_i to derive the ozone value y_{cal}. We have to observe that if the machine learning model used in the training produces bias, the estimation of the method with weighted averages will also have bias, so that the weights used are not optimal in this sense.
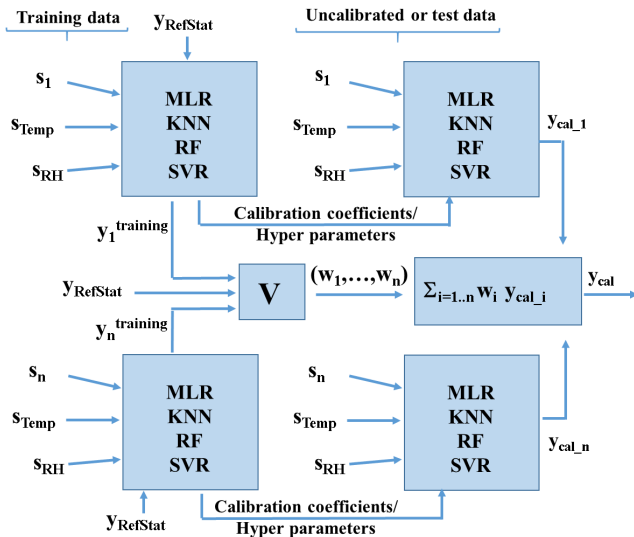

Fig. 2. Multi-sensor data fusion calibration model with weighted averages.

C. Multi-sensor data fusion calibration with machine learning

The idea of multi-sensor data fusion calibration with machine learning is to employ more than one sensor ($s_1, \ldots, s_n$) of the same pollutant at the machine learning module. The outcome of the process, applied to new tuples of data captured by the set of sensors, is an ozone value, y_{cal}, that represents all ozone sensors involved in the calibration process. The goal is for sensors that are highly correlated with the reference values to potentially improve model's performance. However, if the different sensors used in fusion are highly correlated with each other, the multi-collinearity phenomenon, that causes problems in the learning algorithms, may appear.

TABLE II
CORRELATION VALUES FOR SOME SENSORS OF DATA SET 1.

	ref	s1	s2	s3	s4	s5	s6	s7	s8	Temp	RH
ref											
s1	0.92										
s2	0.89	0.98									
s3	0.92	0.99	0.97								
s4	0.92	0.98	0.95	0.98							
s5	0.68	0.87	0.84	0.85	0.87						
s6	0.48	0.63	0.6	0.63	0.58	0.59					
s7	0.87	0.96	0.96	0.97	0.93	0.82	0.73				
s8	0.88	0.98	0.98	0.97	0.95	0.87	0.64	0.97			
Temp	0.77	0.59	0.52	0.6	0.64	0.37	0.17	0.52	0.5		
RH	-0.66	-0.54	-0.49	-0.57	-0.57	-0.39	-0.32	-0.52	-0.49	-0.82	

Table II shows the correlation values between eight of the twenty-eight sensors in data set 1. The first column indicates the correlation of each sensor with the reference values. It can be seen that all ozone sensors are correlated to the reference value in varying degrees, indicating that some sensors will produce better calibrated results than others. In addition, it can be observed that the sensors are also correlated to different degrees to each other. This can lead to a problem of multi-collinearity between sensors if we use them in the multi-sensor fusion process, i.e. the measurements of one sensor are a linear combination of the measurements of the other sensors. The *Variance Inflation Factor* (VIF) indicates whether a variable has multi-collinearity, and is defined as follows:

$$VIF(x_j) = \frac{1}{1 - R^2(x_j)} \quad (4)$$

Where the coefficient of determination R^2 is obtained by regressing x_j over the rest of features. In Figure 3, you can see the percentage of sensors with multi-collinearity given the fusion of different numbers of sensors and several combinations of the same number of sensors, with a confidence interval
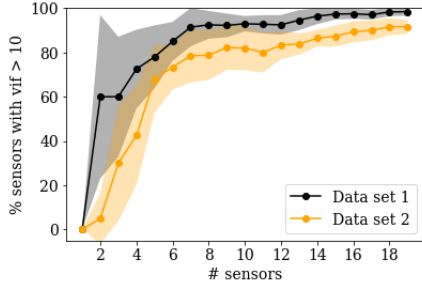
Fig. 3. Mean number of sensors with VIF larger than 10 with different number of sensors fusion. Confidence interval (shaded area) calculated as a t-student with 95% confidence level.

(a) Data set 1 with 4 MOX fusion. (b) Data set 1 with 8 MOX fusion.

Fig. 4. Variables' correlation with the first PLS components.

(shaded area) calculated as a t-student with 95% confidence level. It is observed that when dealing with a fusion of six sensors or more, almost all sensors participating in fusion have multi-collinearity (VIF greater than 10). To overcome multi-collinearity, we use the partial least squares (PLS) method to perform feature extraction. PLS is a supervised learning task, so it makes use of response values ($Y \in \mathbb{R}^N$) to find orthogonal directions that maximize the variance and the correlation between the response values and the directions. It uses an iterative algorithm to find the different directions of the partial least squares. In fact, the m-th PLS component $\hat{\phi}_m$ solves:

$$\underset{\alpha}{\text{maximize}} \quad Corr^2(y, X\alpha)Var(X\alpha) \tag{5a}$$

$$\text{subject to} \quad \|\alpha\| = 1 \tag{5b}$$

$$\alpha^T S \hat{\phi}_l = 0, \quad l = 1, .., m-1 \tag{5c}$$

Where the S matrix is the sample covariance matrix of variable x_j. Thus, at each step, we are maximizing the correlation with the response and the component's variance. In this case, we take advantage of the fact that the components found are orthogonal, so that, multi-collinearity is eliminated. The data projected into the components are used as features instead of employing the raw sensors' measures as features.

For illustrative purposes, Figure 4.a) shows the correlation of the sensors with the two first PLS components for data set 1 and a fusion of four MOX sensors. It is observed that the different ozone sensors are highly correlated with the first component, i.e. the first component represents mainly ozone sensors. In contrast, the second component has more correlation with the temperature and relative humidity sensors. Therefore, this second component will encode information about temperature and relative humidity. Also, Figure 4.b) shows the same results when fusion is performed with eight MOX sensors. The same pattern is observed, the ozone sensors are grouped and highly correlated with the first component.

The model in Figure 5 works as follows: the data from the $\{s_1, \ldots, s_n, s_{Temp}, s_{RH}\}$ sensors and the reference $y_{RefStat}$ values belonging to the training set are used to find the PLS components (PLS box in the training phase). Afterwards, the machine learning model is fed with the reference values and all training data projected to the PLS components found to train the model (MLR, KNN, RF or SVR box in the training phase). Finally, once the new uncalibrated data (test data) is acquired, these sensor values are transformed into the step
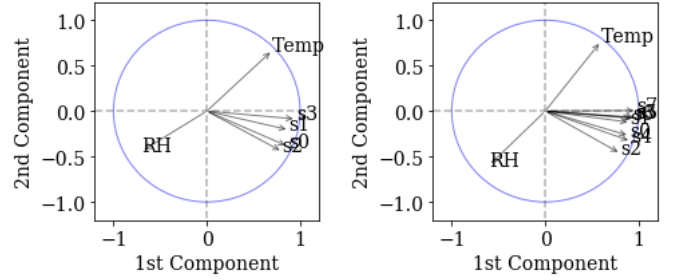
PLS-Transform in the directions found with the training data, and the prediction is made with the learning model of the trained machine model (MLR, KNN, RF or SVR box in the test phase).
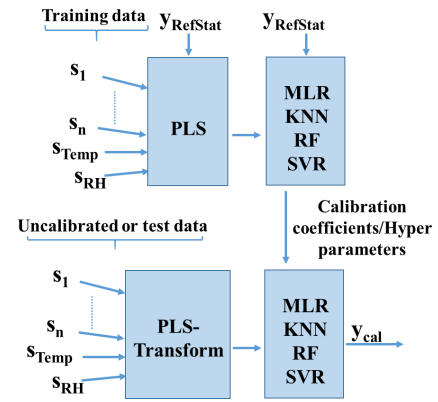

Fig. 5. Multi-sensor data fusion calibration model with machine learning.

V. RESULTS

To evaluate multi-sensor data fusion we perform two types of experiments: (i) fusion with only sensors of the same MOX family, and (ii) fusion of metal-oxide (MOX) sensors with one electro-chemical (EC) sensor. Each MOX sensor costs no more than twenty Euros and therefore up to four sensors per node were placed. On the other hand, the cost of a pair of EC sensors with their electronics is close to three hundred Euros, which makes putting more than one EC sensor per node excessively expensive in a low-cost deployment. When only one sensor has been calibrated, MOX or EC, the model in Figure 1 has been used, while when one sensor has been calibrated using the fusion of several MOX or EC, the model in Figure 2 or Figure 5 has been used. In both cases, 75% of a data set is randomly sampled for training and model selection purposes and the remaining 25% of the same data set is used to validate the performance of the different machine learning models, where the data sets are those described in Table I.

### A. Multi-sensor data fusion calibration with MOX sensors

Ferrer-Cid et. al. [26] show that identical ozone MOX sensors behave with large variability given the same calibration

model. Thus, we will first investigate the fusion of multi-sensor data between sensors of the same family. For this purpose, we will use data sets 1 and 2, each containing twenty-eight and twenty-four MOX sensors placed in the same location. Manlleu reference station for data set 1 and Tona reference station for data set 2 are used.

Figures 6.a) and 6.c) show the RMSE as we add a sensor to the fusion model. To do this, we choose the sensor that has the best RMSE, let's call it s_1, obtained using Figure 1. To calculate the fusion with two ozone sensors, we randomly choose ten sensors and form ten groups of two sensors in which each group always contains the sensor s_1. We use the calibration model in Figure 5 for each of the ten groups, and calculate the average RMSE over the ten groups of sensors and a confidence interval (shaded area) calculated as a t-student with 95% confidence level. To calculate the fusion with three sensors, we proceed in the same way. We keep sensor s_1, and form ten groups with three sensors, where the other two sensors are chosen randomly. We apply the calibration model in Figure 5 for each group, and calculate the average RMSE with its confidence interval over the ten groups of sensors. We repeat the process for the fusion of n-sensors.
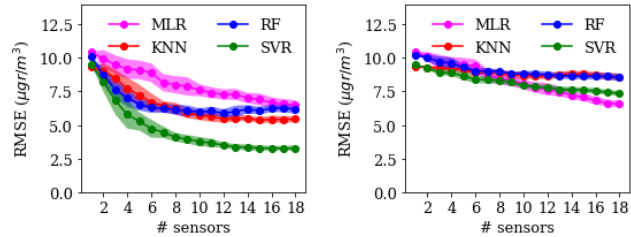
(a) Data set 1, fusion via machine learning.

(b) Data set 1, fusion via weighted averages.
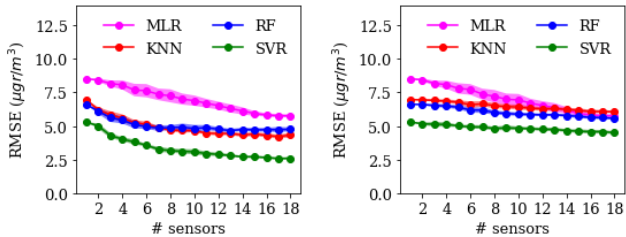
(c) Data set 2, fusion via machine learning.

(d) Data set 2, fusion via weighted averages.

Fig. 6. Mean validation RMSE with confidence intervals (shaded area) of 95% for data sets 1 and 2.

The first thing we can observe in Figures 6.a) (data set 1) and 6.c) (data set 2) is that linear methods (MLR) have worse RMSE than non-linear methods (KNN, RF and SVR). In data set 2, choosing the best of the sensors s_1, we obtain an RMSE of 8.5 μgr/m^3 for MLR, of 7.0 μgr/m^3 and 6.6 μgr/m^3 for KNN and RF and of 5.3 μgr/m^3 for SVR. In the case of data set 1, the best sensor has a very similar RMSE for the four calibration methods, but we have observed the same trend that in data set 2 if we compare the different methods: MLR the worst, KNN and RF similar, and SVR the best. When it comes to fusion by weighted averages, Figures 6.b) and 6d),

the improvement is smaller. This is because we are far from optimal conditions for fusion with weighted averages, since MOX sensors are highly correlated with each other, Table II. It is also observed that this fusion has the same behavior as fusion with machine learning for the MLR case.

In multi-sensor data fusion with machine learning, adding between four and six sensors considerably reduces the RMSE on all models. The one that less reduces the RMSE is the MLR and needs more sensors in the fusion to decrease the RMSE. On the other hand, non-linear models reduce the RMSE between 20% and 40% with four sensors. In the data set 2, whose best sensor has an RMSE between 5 and 8 μgr/m^3 depending on the model used, the reduction is smaller. However, by using three or four sensors in the fusion, the result of having only one sensor is improved. On the other hand, the use of more than five or six sensors in fusion does not improve too much the RMSE in average.

B. Multi-sensor data fusion calibration with MOX and EC sensors

We have seen in the previous section that the fusion of multi-sensor data with the same technology with low-cost sensors improves the quality of the data calibrated in terms of RMSE when we use fusion with machine learning. In this section, we investigate what happens if we merge sensor data from different technologies. To do this, we will use data sets 3, 4 and 5, which include a node with four ozone MOX sensors and a node with an ozone EC sensor. To obtain the mean validation RMSE and the confidence interval, we will use the same procedure followed in the previous subsection. The amount of shaded area due to the confidence interval depends on the data collected.

We note, Figure 7.a), c) and e), that MOX sensors in data set 3 perform much better, lower RMSE, than MOX sensors in data sets 4 and 5. For example, the best of the C17017 node sensors in the 2017 summer campaign, Figure 7.a), has an RMSE of of 12.5 μgr/m^3 for MLR, of 10.3 μgr/m^3 and 10.0 μgr/m^3 for KNN and RF and of 9.7 μgr/m^3 for SVR, while the best of the C18017 node sensors in the 2018 summer campaign, Figure 7.c), has an RMSE of 31.8 μgr/m^3 for MLR, of 27.8 μgr/m^3 and 27.5 μgr/m^3 for KNN and RF and of 27.2 μgr/m^3 for SVR. One of the reasons for the very different behaviour between the two campaigns is that the environmental conditions, with higher heat episodes in 2018 compared to 2017, led to higher ozone concentrations in 2018, and therefore the behaviour of the ozone sensors, especially in their capacity to measure very high ozone values, was lower than with lower ozone conditions. This effect makes us have a data set with sensors that behave well (data set 3), and another with sensors that behave badly (data set 4), allowing us to investigate two extreme cases of each other.

The EC sensor gives better RMSE values than any of the MOX sensors in data sets 3 and 4, Figures 7.a) and c), while in data set 5, Figure 7.e), the MOX sensors outperform the EC sensor. We also observe that the fusion of only MOX sensors does not improve the result of having an EC sensor on average, Figures 7.a) and c). In the case of the fusion with

machine learning of an EC sensor with MOX sensors with good performance (data set 3), we can observe in the four calibration models, Figure 7.a), that the RMSE obtained is better than having only the EC sensor. Adding more than one MOX sensor to the fusion with the EC sensor improves slightly in the case of MLR, does not improve or worsen in the case of KNN and RF, and improves when we use SVR.

In the case of the fusion with machine learning of an EC sensor with MOX sensors with worse performance (data set 4), we can observe in the four calibration models, Figure 7.c), that the RMSE obtained does not improve or worsen having only one EC sensor. It is observed that sometimes, e.g., MLR case, if more MOX sensors are added, the fusion can lead to a slight increase in errors. This is because in this case of fusion, one technology is clearly performing worse than the other, thus fusion does not improve the calibration model with sensor array by choosing the best sensor. In other words, the worst performing sensors are worse correlated with the concentration value of the reference station. These results show that when there is a large performance difference between sensor technologies, sensor fusion does not produce a large gain, as is also the case for data set 5, Figure 7.e). In the latter case, where MOX sensors behave better than the EC sensor, the improvement of merging both technologies is very small. We can conclude that when both technologies measure in the same range, improvements are observed, while when one technology behaves better than the other, fusing slightly improves results, especially with non-linear methods.

In the case of using fusion with weighted averages, we can observe that the results deviate from the optimal theoretical improvement. In the practical case, we are merging biased estimators obtained from the machine learning methods (MLR, KNN, RF or SVR) and highly correlated sensors between sensors of the same technology. In the three data sets, Figures 7.b), d), f), we can see that fusion between only MOX sensors produces the same trends as with the fusion with machine learning, but with fewer RMSE improvement. In one case, Figure 7.f) (data set 5) RF and KNN worsen using weighted averages. These results are not as good as those obtained by merging the sensors into the machine's learning model, but the results show a similar trend, even when conditions differ from optimal. The reason for not having as good a performance as the machine learning approach is due to the high level of correlation between the MOX sensor errors, while the optimal conditions for fusion would be to have independent errors. On the other hand, in the case of fusion between technologies, a MOX sensor plus an EC sensor, these fusion methods can introduce an improvement when the technologies do not differ much in performance.

VI. Conclusions

We have investigated the fusion of multi-sensor data in an IoT deployment with metal-oxide (MOX) and electrochemical (EC) ozone sensors. Multi-sensor data fusion consists of giving a calibrated value of an ozone sensor using a set of ozone sensors. We have shown the basic architecture, sensor array calibration, to calibrate an air pollution sensor with environmental sensors such as temperature and relative humidity. In addition, we have described two other calibration systems used for the fusion of data from multiple sensors. The first is based on obtaining the mean weighted value, where the optimum weights are obtained from the covariance matrix of the errors, since it assumes that the errors between the pollution sensors used are correlated. The second method is based on fusing raw data into the machine learning model. In this system, because there can be a multi-collinearity effect, we perform a PLS on the data before using the machine learning models. This reduces the effect of highly correlated sensors.

First we have shown how multi-sensor data fusion improves RMSE using ozone sensors with MOX technology. In fact, the average RMSE is improved with four or six MOX sensors. This technique is feasible since MOX sensors have a reduced cost and their electronics, a voltage divider, is simple, so the cost of having a small set of sensors, for example, four sensors on a board is not large. We have also seen that using two complementary technologies can reduce the RMSE. In this case, we have tested data sets with different behaviors. In two data sets, the individual EC sensor behaved better than the best of the array's MOX sensors, in another vice versa. In general, it is when both technologies have similar behaviors that multi-sensor fusion behaves best. In this case, the complementarity of technologies improves the final value obtained. On the other hand, when one technology behaves much better than the other, there is no substantial improvement in the fusion result. This is due to the fact that the fusion result is basically the result of the sensor with very good behaviour, while the sensors that behave much worse are poorly correlated with the reference ozone concentrations.

Another important conclusion is that from the methodological point of view, we have seen that when it comes to real data sets, the machine learning approach of sensor fusion works better than the weighted average approach. This is because the optimal conditions (e.g. uncorrelated errors) for sensor fusion are difficult to meet in a real data set case. As future work, we think that the fusion of multi-sensor data can also be interesting in case there is aging in the sensors. EC sensors are known to age before MOX. This means that the fusion between EC and MOX could get better results than only EC when the EC sensor starts to age and MOX does not. Moreover, we believe that the estimation of concentration values can be improved if we take into account the information from sensors located in the vicinity of the sensor to be calibrated exploiting the spatial correlation of ozone concentrations.

References

[1] J. M. Barcelo-Ordinas, J.-P. Chanet, K.-M. Hou, and J. García-Vidal, "A survey of wireless sensor technologies applied to precision agriculture," in *Precision agriculture13.* Springer, 2013, pp. 801–808.

[2] Y. Kim, H. Park, and M. B. Srivastava, "A longitudinal study of vibration-based water flow sensing," *ACM Transactions on Sensor Networks (TOSN)*, vol. 9, no. 1, p. 8, 2012.

[3] R. Tan, G. Xing, Z. Yuan, X. Liu, and J. Yao, "System-level calibration for data fusion in wireless sensor networks," *ACM Trans. Sen. Netw.*, vol. 9, no. 3, pp. 28:1–28:27, Jun. 2013.

[4] B. Maag, Z. Zhou, and L. Thiele, "A survey on sensor calibration in air pollution monitoring deployments," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4857–4870, Dec 2018.
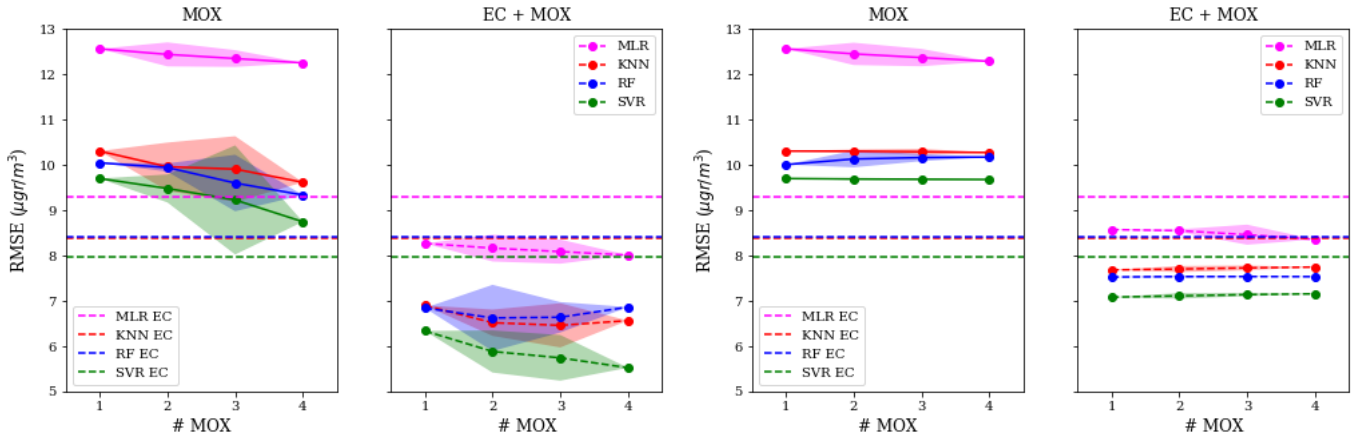
(a) Mean validation RMSE for MOX fusion and EC + MOX fusion, using fusion with machine learning, for data set 3.

(b) Mean validation RMSE for MOX fusion and EC + MOX fusion, using fusion with weighted averages, for data set 3.
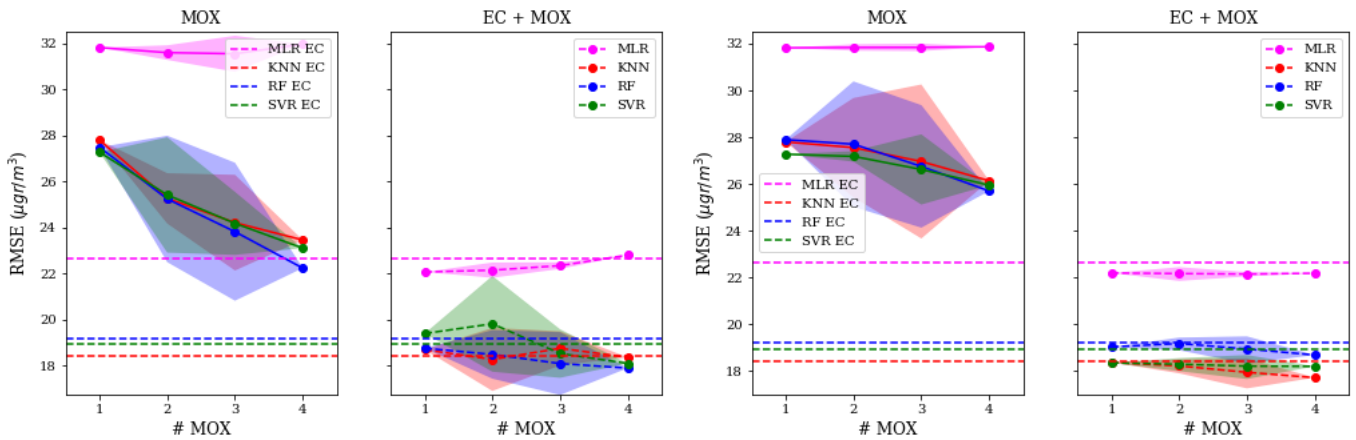
(c) Mean validation RMSE for MOX fusion and EC + MOX fusion, using fusion with machine learning, for data set 4.

(d) Mean validation RMSE for MOX fusion and EC + MOX fusion, using fusion with weighted averages, for data set 4.
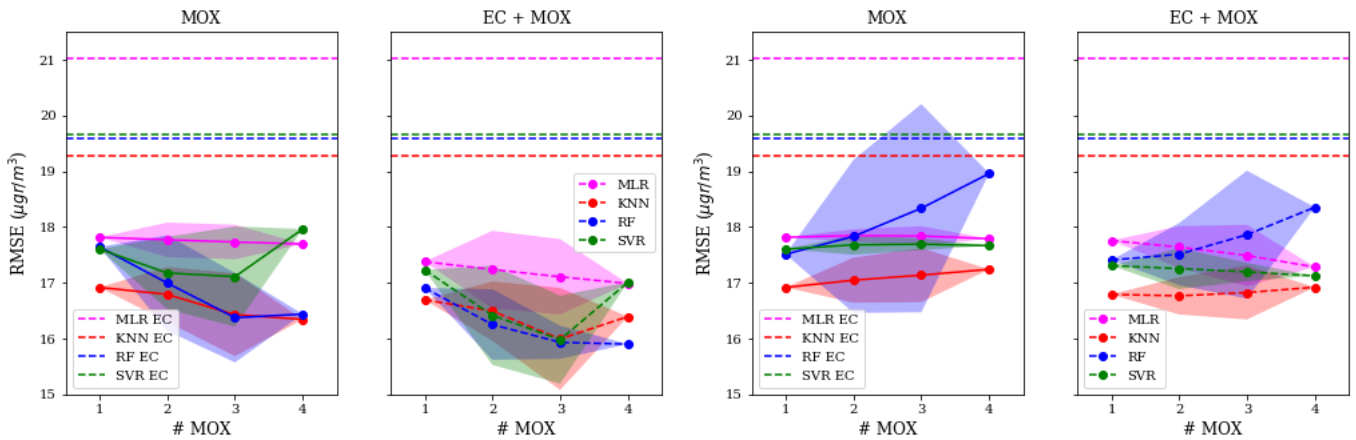
(e) Mean validation RMSE for MOX fusion and EC + MOX fusion, using fusion with machine learning, for data set 5.

(f) Mean validation RMSE for MOX fusion and EC + MOX fusion, using fusion with weighted averages, for data set 5.

Fig. 7. Mean validation RMSE for data sets 3, 4 and 5, using just one EC sensor, fusion with weighted averages and fusion with machine learning of only MOX sensors and one EC sensor with MOX sensors. Confidence interval (shaded area) calculated as a t-student with 95% confidence level.

[5] P. Huang, L. Xiao, S. Soltani, M. W. Mutka, and N. Xi, "The evolution of mac protocols in wireless sensor networks: A survey," *IEEE communications surveys & tutorials*, vol. 15, no. 1, pp. 101–120, 2013.

[6] N. A. Pantazis, S. A. Nikolidakis, and D. D. Vergados, "Energy-efficient routing protocols in wireless sensor networks: A survey," *IEEE Communications surveys & tutorials*, vol. 15, no. 2, pp. 551–591, 2013.

[7] M. Doudou, J. M. Barcelo-Ordinas, D. Djenouri, J. Garcia-Vidal, A. Bouabdallah, and N. Badache, "Game theory framework for mac parameter optimization in energy-delay constrained sensor networks," *ACM Transactions on Sensor Networks*, vol. 12, no. 2, p. 10, 2016.

[8] J. M. Barcelo-Ordinas, M. Doudou, J. Garcia-Vidal, and N. Badache, "Self-calibration methods for uncontrolled environments in sensor networks: A reference survey," *Ad Hoc Networks*, vol. 88, pp. 142–159, 2019.

[9] A. Lewis, W. R. Peltier, and E. von Schneidemesser, "Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications," *World Meteorological Organization*, 2018.

[10] A. Ripoll, M. Viana, M. Padrosa, X. Querol, A. Minutolo, K. M. Hou, J. M. Barcelo-Ordinas, and J. García-Vidal, "Testing the performance of sensors for ozone pollution monitoring in a citizen science approach," *Science of the Total Environment*, vol. 651, pp. 1166–1179, 2019.

[11] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola, "Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. part b: NO, CO and CO2," *Sensors and Actuators B: Chemical*, vol. 238, pp. 706–715, 2017.

[12] J. M. Barcelo-Ordinas, J. Garcia-Vidal, M. Doudou, S. Rodrigo-Muñoz, and A. Cerezo-Llavero, "Calibrating low-cost air quality sensors using multiple arrays of sensors," in *Wireless Communications and Networking Conference (WCNC)*. IEEE, 2018, pp. 1–6.

[13] A. Bigi, M. Mueller, S. K. Grange, G. Ghermandi, and C. Hueglin, "Performance of no, no$_2$ low cost sensors and three calibration approaches within a real world application," *Atmospheric Measurement Techniques*, vol. 11, no. 6, pp. 3717–3735, 2018.

[14] D. Hagan, G. Isaacman-VanWertz, J. Franklin, L. Wallace, B. Kocar, C. Heald, and J. Kroll, "Calibration and assessment of electrochemical air quality sensors by colocation with regulatory-grade instruments," *Atmosph. Measurement Tech.*, vol. 11, no. 1, pp. 315–328, 2018.

[15] C. Malings, R. Tanzer, A. Hauryliuk, S. P. N. Kumar, N. Zimmerman, L. B. Kara, A. A. Presto, and R. Subramanian, "Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring," *Atmospheric Measurement Techniques*, vol. 12, no. 2, pp. 903–920, 2019.

[16] S. De Vito, E. Esposito, M. Salvato, O. Popoola, F. Formisano, R. Jones, and G. Di Francia, "Calibrating chemical multisensory devices for real world applications: An in-depth comparison of quantitative machine learning approaches," *Sensors and Actuators B: Chemical*, vol. 255, pp. 1191–1210, 2018.

[17] N. Zimmerman, A. A. Presto, S. P. Kumar, J. Gu, A. Hauryliuk, E. S. Robinson, A. l. L. Robinson, and R. Subramanian, "A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring." *Atmospheric Measurement Techniques*, vol. 11, no. 1, 2018.

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

[19] J. M. Barcelo-Ordinas, P. Ferrer-Cid, J. Garcia-Vidal, A. Ripoll, and M. Viana, "Distributed multi-scale calibration of low-cost ozone sensors in wireless sensor networks," *Sensors*, vol. 19, no. 11, 2019.

[20] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In-network aggregation techniques for wireless sensor networks: a survey," *IEEE Wireless Communications*, vol. 14, no. 2, pp. 70–87, 2007.

[21] H. Wu, Z. Zhang, C. Jiao, C. Li, and T. Q. Quek, "Learn to sense: a meta-learning based sensing and fusion framework for wireless sensor networks," *IEEE Internet of Things Journal*, 2019.

[22] S. Madgwick, "An efficient orientation filter for inertial and inertial/magnetic sensor arrays," *Report x-io and University of Bristol (UK)*, vol. 25, pp. 113–118, 2010.

[23] J. Wu, Y. Feng, and P. Sun, "Sensor fusion for recognition of activities of daily living," *Sensors*, vol. 18, no. 11, p. 4029, 2018.

[24] F. Castanedo, "A review of data fusion techniques," *The Scientific World Journal*, vol. 2013, 2013.

[25] R. Tan, G. Xing, Z. Yuan, X. Liu, and J. Yao, "System-level calibration for data fusion in wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 9, no. 3, p. 28, 2013.

[26] P. Ferrer-Cid, J. M. Barcelo-Ordinas, J. Garcia-Vidal, A. Ripoll, and M. Viana, "A comparative study of calibration methods for low-cost ozone sensors in iot platforms," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9563–9571, Dec 2019.

[27] J. K. Hackett and M. Shah, "Multi-sensor fusion: a perspective," in *Proceedings., IEEE International Conference on Robotics and Automation*. IEEE, 1990, pp. 1324–1330.

[28] R. C. Luo, C.-C. Yih, and K. L. Su, "Multisensor fusion and integration: approaches, applications, and future research directions," *IEEE Sensors journal*, vol. 2, no. 2, pp. 107–119, 2002.

[29] W. Elmenreich and R. Leidenfrost, "Fusion of heterogeneous sensors data," in *2008 International Workshop on Intelligent Solutions in Embedded Systems*. IEEE, 2008, pp. 1–10.

[30] C.-Y. Chong, S. Mori, W. H. Barker, and K.-C. Chang, "Architectures and algorithms for track association and fusion," *IEEE Aerospace and Electronic Systems Magazine*, vol. 15, no. 1, pp. 5–13, 2000.

[31] P. Avery, "Combining measurements with correlated errors," *CLEO Note CBX*, pp. 95–55, 1996.

Pau Ferrer-Cid is a PhD student at the Statistical Analysis of Networks and Systems (SANS) research group, Universitat Politecnica de Catalunya (UPC). He holds a B.Sc in Computer Science and a M.Sc in Data Science by the UPC. His main research interests are the applications of novel data analysis methods to sensor data coming from IoT platforms and the analysis of other kinds of data from fields like biology and computer vision.

Jose M. Barcelo-Ordinas is an Associate Professor at Universitat Politecnica de Catalunya (UPC) from 1999. He holds a PhD and B.Sc+M.Sc in Telecommunication Engineering and a B.Sc+M.Sc in Mathematics. He has participated in many European projects such as WIDENS, EuroNGI, EuroNFI, EuroNF NoE and H2020 CAPTOR. His currently research areas are wireless sensor networks, mobility patterns, and the statistical analysis of sensor data.

Jorge Garcia-Vidal is since 2003, full professor at the Computer Architecture Department of UPC, and since 2012 responsible of the Smart Cities Initiative at Barcelona Supercomputing Center (BSC-CNS), coordinating the H2020 CAPTOR project or being the BSC-CNS responsible of the H2020 project ASGARD. His main current research interest is in problems related with the capture, processing and statistical analysis of sensor data.

Anna Ripoll is an environmental consultant at 4sfera Innova. She holds a PhD degree in environmental sciences, and has participated in several European projects including H2020 CAPTOR and ACTRIS. One of her main current research is the analysis of air quality sensor time series and calibration methods, and tools for sensor technologies.

Mar Viana is a staff researcher at the Institute of Environmental Assessment and Water Research (IDAEA-CSIC) in Barcelona, Spain, since 2011. Her research focuses on atmospheric aerosols, with main interests in outdoor and indoor air pollution, source apportionment and links with health. She is coordinator of ERANET project CERASAFE and IP of H2020 project CAPTOR, and participates in several other European and national projects.