

Calibration of Low-cost Gas Sensors for Air Quality Monitoring

Dimitris Margaritis^{1,2}, Christos Keramydas³, Ioannis Papachristos³,
Dimitra Lambropoulou^{1,4*}

¹Aristotle University of Thessaloniki / School of Chemistry, University Campus, 54124, Thessaloniki, Greece

²Centre for Research and Technology Hellas (CERTH) / Hellenic Institute of Transport (HIT), 57001 Thessaloniki, Greece

³Department of Supply Chain Management, International Hellenic University, 60100, Katerini, Greece

⁴Center for Interdisciplinary Research and Innovation (CIRI-AUTH), Balkan Center, Thessaloniki, GR-57001, Greece

ABSTRACT

Mobile monitoring devices equipped with low-cost gas sensors in fixed stations are an emerging solution to enhance the spatial coverage of air quality monitoring networks. We estimated the measurement accuracy of two AQMesh devices, evaluated their agreement, and examined the related calibration characteristics. Three widely used calibration approaches were investigated, namely uni- and multi-variate linear regression analysis and the random forest algorithm. Two identical commercial AQMesh platforms (monitoring NO, NO₂, O₃, and SO₂) were installed on a fixed municipal station for 4 consecutive weeks. Widely used statistical indexes were employed to evaluate device performance and calibration outcomes. The devices exhibited favorable performance in following the pattern of the station's reference time series in a 10-min average resolution. Nevertheless, their performance was lower, with respect to the reference values, in terms of the average error and overall bias. The calibration improved the agreement between the device and reference measurements. The emission time series of each device was consistent with the other (pre- and post-calibration) in terms of measurement patterns and point-by-point deviations. The three alternative methodologies had similar calibration performance overall. The random forest algorithm appeared to have an advantage in several cases, mostly in terms of following the pattern in the O₃ and SO₂ time series, but also in terms of the average error and bias for all pollutants.

Keywords: Electrochemical sensors, Gas emissions, Statistical analysis, Calibration models

OPEN ACCESS

Received: April 16, 2021

Revised: July 31, 2021

Accepted: September 10, 2021

* **Corresponding Author:**

dlambro@chem.auth.gr

Publisher:

Taiwan Association for Aerosol
Research

ISSN: 1680-8584 print

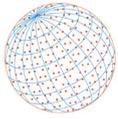
ISSN: 2071-1409 online

 **Copyright:** The Author(s).

This is an open access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

1 INTRODUCTION

Local authorities of large urban areas are obliged by the national governments to monitor the air quality of their region and to report the indicators that have been established by either the European Commission or relevant health organizations. For this reason, fixed stations are established according to an air sampling plan that is representative of all activities occurring in that area (e.g., transport, industrial, residential, and commercial activities). The costs of realizing and maintaining such a network of fixed monitoring stations is high, especially because a network consists of at least four monitoring and baseline stations. The relocation of a fixed station to another location is another disadvantage of this configuration. Emission sources might vary over the years or even within a short time period (such as seasonal points of interest), and the spatial monitoring of emission concentrations could therefore be highly important, especially when local authorities must decide on countermeasures to reduce excessive pollution.



Mobile monitoring devices equipped with low-cost gas sensors have been applied the last decade and could be a means of deploying a satisfactory network to enhance the spatial coverage that support rather than replace fixed station networks. Such mobile devices are compact in size and usually based on electrochemical sensors with satisfactory accuracy and favorable detection sensitivity to concentration variations (Borrego *et al.*, 2016; Spinelle *et al.*, 2015; Kumar *et al.*, 2015). However, electrochemical sensors are sensitive to gas interference that might produce lower or higher values compared with the reference (higher accuracy) gas analyzers (Marco and Gutiérrez-Gálvez, 2012). Frequent calibration of the measured values is an important process due to the degradation of the sensors over time and their sensitivity to environmental parameters, such as temperature and relative humidity. Calibration techniques based on statistical methodologies have been applied to improve the accuracy of passive gas analyzers, as described in the following paragraphs.

Simple linear regression (SLR) is perhaps the most common statistical methodology used in the field for device calibration. Because the relationships between the measured values of the instruments and those of the reference instruments are usually linear, as demonstrated in practice, linear regression is a reliable and widely used calibration approach (Castell *et al.*, 2017; Karagulian *et al.*, 2019; Schneider *et al.*, 2017). In SLR, the measurement (reading) of the instrument to be calibrated is an independent variable, whereas the measurement of the reference instrument is a dependent variable.

An alternative approach within the broader field of regression analysis is multiple linear regression (MLR), which allows the use of additional input variables in calibration models (Karagulian *et al.*, 2019). The goal in this case is to improve calibration quality through a more accurate interpretation, based on these additional variables, of the instrument's behavior. In international literature, the effect of temperature and relative humidity—being among the most important environmental variables—on the instrument measurement process as well as the role of these two variables in the calibration process has been documented and reported several researchers (Borrego *et al.*, 2016; Karagulian *et al.*, 2019).

In addition to the aforementioned linear analyses, several nonlinear approaches have been applied in research and measurement practices (Karagulian *et al.*, 2019). These include nonlinear regression (NLR), which generally follows the principles of regression analysis, with the main difference being the use of a nonlinear function as an adjustment function in measurement data for exponential, logarithmic, or quadratic functions.

Several algorithms related more broadly to machine learning methodologies have also been used for instrument calibration. For example, the random forest (RF) algorithm is a supervised learning technique that is used to solve classification and regression problems (Borrego *et al.*, 2016; Karagulian *et al.*, 2019; Zauli-Sajani *et al.*, 2021). The algorithm creates decision-trees through bootstrapping during training.

Artificial neural networks are commonly used for calibrating measurement devices (Borrego *et al.*, 2016). Finally, reports on the use of hybrid models are based on a combination of two techniques (i.e., RF and linear regression) (Zauli-Sajani *et al.*, 2021).

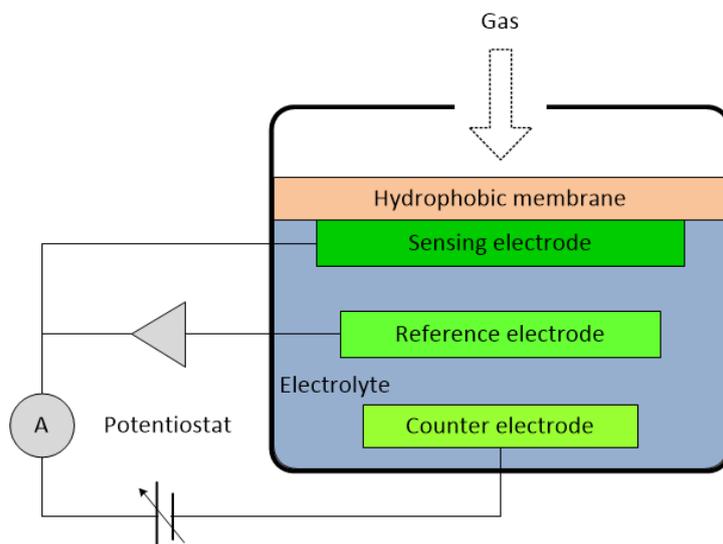
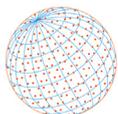
In the present paper, the performance of a commercial AQMesh platform (monitoring NO, NO₂, O₃, and SO₂) located roadside in a city center affected by urban traffic is evaluated. Two identical devices were installed for 4 consecutive weeks on the roof of a fixed monitoring station for a municipality, providing synchronized data for characterizing the temporal devices' performance and determining variations between identical devices. The objectives of this study were the estimation of the measurement accuracy of the two widely used devices, the evaluation of their agreement, and the examination of the related calibration characteristics. For this purpose, three calibration approaches were investigated.

2 METHODS

2.1 Measurement Equipment

2.1.1 AQMesh pods

The AQMesh devices (Environmental Instruments Ltd., UK, <https://www.aqmesh.com/>) are portable, small (22 × 16 × 20 cm), and lightweight (< 2 kg). They consist of a metal box with



electrochemical sensors for the detection and measurement of NO (Alphasense NO-B4), NO₂ (Alphasense NO₂-B43F), O₃ (Alphasense OX-B431), and SO₂ (Alphasense SO₂- B4) as well as three sensors that measure temperature, relative humidity, and atmospheric pressure (AQMesh, 2020).

Electrochemical sensors measure the concentration of gaseous pollutants through the positive or negative currents created by reactions between gases and electrodes. The pollutant penetrates through a membrane into the sensor housing, and an oxidation or reduction reaction begins. Oxidation (or a reduction with an opposite current direction) causes electrons to flow from one electrode to another through an external circuit (Fig. 1). The produced current is proportional to the concentration of the pollutants, and an external circuit detects and amplifies this signal (Zauli-Sajani *et al.*, 2021).

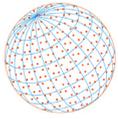
The data are collected through a built-in unit, which forwards the data to a server through a Global System for Mobile Communications (GSM) connection. To correct the effect of meteorological parameters and other gases' cross-interference, the company processes the raw measurements of the sensors using its own algorithm. A dedicated website hosts the postprocessed data that can be downloaded or analyzed using tools available on the web platform. The maximum possible data recording frequency is 1 Hz; however, due to the instability of the electrochemical sensors, the company provides a maximum frequency of 0.017 Hz (one record per minute), and that frequency was used in the measurements of the present study.

The two aforementioned devices were acquired in the summer of 2016 and in 2017, they were sent to Environmental Instruments Ltd. in UK for the replacement of their NO₂ and O₃ sensors with a new generation of sensors with improved accuracy.

2.1.2 Municipal monitoring station

The station is located at the intersection of the main road axis of Thessaloniki City in Greece (40°38'15.4"N, 22°56'27.9"E), which is called the Egnatia–Monastiriou axis, in the road link that has the largest traffic volume in all of the city (from Dimokratias Square to Agia Sofia), and Venizelou Street, which connects the Port of Thessaloniki with Ano Poli and is considered the most important commercial axis of Thessaloniki. The Egnatia station contributes to the air pollution control network by recording the maximum concentrations of all pollutants (other than O₃) in the city and simultaneously warning of an impending atmospheric episode. It started its operation in 1989 and is 12 m above sea level and 650 m from the sea.

The concentrations of gases, which were used as reference values, were measured using Environnement S.A. analyzers (an AC32M - Chemiluminescent for NO_x, an O₃42M - UV photometric for O₃, and an AF22M - UV Fluorescent for SO₂). The municipality of Thessaloniki has two Environnement programmable calibrators with VE 3M and MGC 101 dilution systems; they are suitable for the calibration of NO_x, O₃, and SO₂ analyzers.



2.2 Sampling Process

2.2.1 Thessaloniki City

Thessaloniki (40°37'45.3684"N, 22°56'50.6832"E), with at least 1 million inhabitants, is located on the northeast coast of Thermaikos Bay. The city has a hill range with an altitude of 300–1200 m. Thessaloniki's climate is typically Mediterranean, with cold and wet winters and hot and dry summers. The city is characterized by dense construction, including tall buildings, narrow streets, and virtually no uncovered spaces, such as parks. High levels of air pollution have been observed only in the commercial center of the city due to its overconcentration of activities and heavy traffic. High concentrations for certain pollutants, such as nitrogen dioxide and particulate matters, are often notified by the city authorities (Kassomenos *et al.*, 2011; Vlachokostas *et al.*, 2009). For example, according to the 2014 air quality city center data, the average NO_x value was 58 μg m⁻³, with the permissible yearly limit set at 40 μg m⁻³ (Pasxalia, 2019).

2.2.2 Data collection

The two pods were mounted 1 m away from the inlet probe of the reference station and at approximately the same height (with a difference of a few centimeters). Their relative distance was 40 cm. The data collection took place from September 5, 2017 (24:00), to September 30, 2017 (24:00), and the final data corresponded to 3,744 10-min average values per pollutant.

2.3 Statistical Analysis

2.3.1 Data processing

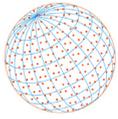
A first processing of the data was conducted by AQMesh at the company's data collection site in the United Kingdom. After their acquisition, the measured time series data (emissions and environmental values) had to be synchronized to the respective data of the reference municipal station due to the time differences between the two countries. Negative and missing values were excluded from the analysis. The data were finally aggregated to a 10-min average resolution.

2.3.2 Statistical indices

Metrics that are widely used in the relevant literature were employed to evaluate the relationship between the time series of the two devices (pod and municipal reference) both before and after calibration (Borrego *et al.*, 2016; Karagulian *et al.*, 2019; Schneider *et al.*, 2017). The metrics, which are presented in Table 1, were designed to capture the main aspects of time series behaviors. To that end, Pearson's coefficient of correlation (*r*) was used to capture the level

Table 1. Statistical indicators for data evaluation before and after calibration.

Metric	Formula	Range	Target value
Pearson's coefficient of correlation	$r_p = \frac{\sum_{i=1}^n (ref_i - \overline{ref})(pod_i - \overline{pod})}{\sqrt{\sum_{i=1}^n (ref_i - \overline{ref})^2} \sqrt{\sum_{i=1}^n (pod_i - \overline{pod})^2}}$	-1 to +1	+1
Relative Bias	$RBS = \frac{\overline{pod} - \overline{ref}}{\overline{ref}}$	$-\infty$ to $+\infty$	0%
Normalized Mean Absolute Error	$nMAE = \frac{\frac{1}{n} \sum_{i=1}^n pod_i - ref_i }{\overline{ref}}$	0% to $+\infty$	0%
Normalized Mean Squared Error	$nRMAE = \frac{\frac{1}{n} \sum_{i=1}^n (pod_i - ref_i)^2}{\overline{ref}}$	0% to $+\infty$	0%
Mean Absolute Percentage Error	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{pod_i}{ref_i} \right $	0% to $+\infty$	0%



of agreement in the trend of the two time series; relative bias (rBS) was employed as a measure of the difference between the two time series; the normalized mean absolute error (nMAE) and the normalized root mean squared error (nRMSE) were used as indices of the respective average absolute error, and the nRMSE penalizes larger errors. The mean absolute percentage error (MAPE) was employed to measure the average relative error. The overall analysis was conducted in R (R Core Team, 2021).

2.3.3 Alternative approaches and the calibration process

The following three widely used alternative calibration approaches were selected for comparison: (i) SLR analysis, where the regressor variable is the measurement (reading) of the device to be calibrated; (ii) MLR analysis, where three regressor variables (the measurement of the device for calibration, the ambient temperature, and the relative humidity of the environment) are employed; and (iii) the RF algorithm, where the regressor variables are the same as those employed in the MLR case. The selected approaches represent simple classical and advanced nonclassical methodologies available in the field, and they also cover the potential uni- and multi-variate aspects of the problem and its potential linear or nonlinear behavior.

The Multiple Linear Regression methodology is the generalization of the linear regression concept to multiple explanatory variables (features). It is based on the linearity of the relationships between the regressors and the response variable under study, i.e., in our case, the concentration of gas pollutants. The outcome of an MLR model is an algebraic multivariable function. In a mathematical point of view, linear regression is a parametric approach that also has some assumptions in order to provide reliable results, as for example linearity, normality, homoscedasticity, etc.

On the other hand, the Random Forest algorithm is an ensemble learning method that is based on decision trees. Therefore, there is some kind of classification that takes place. In contrast to MLR, the RF algorithm is a non-parametric methodology, i.e., practically, there are no assumptions to be made, and no parameters of the respective distributions are used during the implementation of the model. The RF algorithm just creates decision trees, i.e., practically, the algorithm creates bins in a tree-oriented approach and classifies each observation within a specific bin according to its features' values. The main advantage of RF, in comparison to the simple version of regression trees, is that it reduces the variability of the results due to its main idea of sampling both the subjects and the features in the training dataset.

The calibration process comprised the following steps:

Step 1, data processing: This involved synchronizing the involved time series, handling negative and missing values in the measurement readings, and temporally aggregating the associated time series (emissions and environmental values).

Step 2, calibration modeling: Each of the two devices was calibrated independently in relation to the corresponding reference device (station) of the municipality of Thessaloniki City by using all three calibration methods. The outcome of this step was the creation of three calibrated time series per measuring device and pollutant.

Step 3, evaluation of outcomes: The effectiveness and the accuracy of the overall calibration process were evaluated based on statistical indicators. The outcomes of this process were examined to select the most appropriate calibration method.

3 RESULTS AND DISCUSSION

3.1 Device Accuracy Evaluation

The time series of the measurements for all the three devices (i.e., Pod₁, Pod₂, and the municipal device) and for NO, NO₂, O₃, and SO₂ are presented in Fig. 2. The Pod₁ and Pod₂ devices appear to be in agreement regarding the pattern of the respective emission time series for all pollutants. This seems to also be the case for NO, O₃, and SO₂ with respect to the measured emissions levels of Pod₁ and Pod₂. However, for NO₂, an apparent offset was observed between the two measurements; the level of the respective NO₂ measurements (time-series) of the Pod₁ device was higher than the respective level of the Pod₂ device.

The two devices seemed to follow the trend of the reference time series, exhibiting a generally

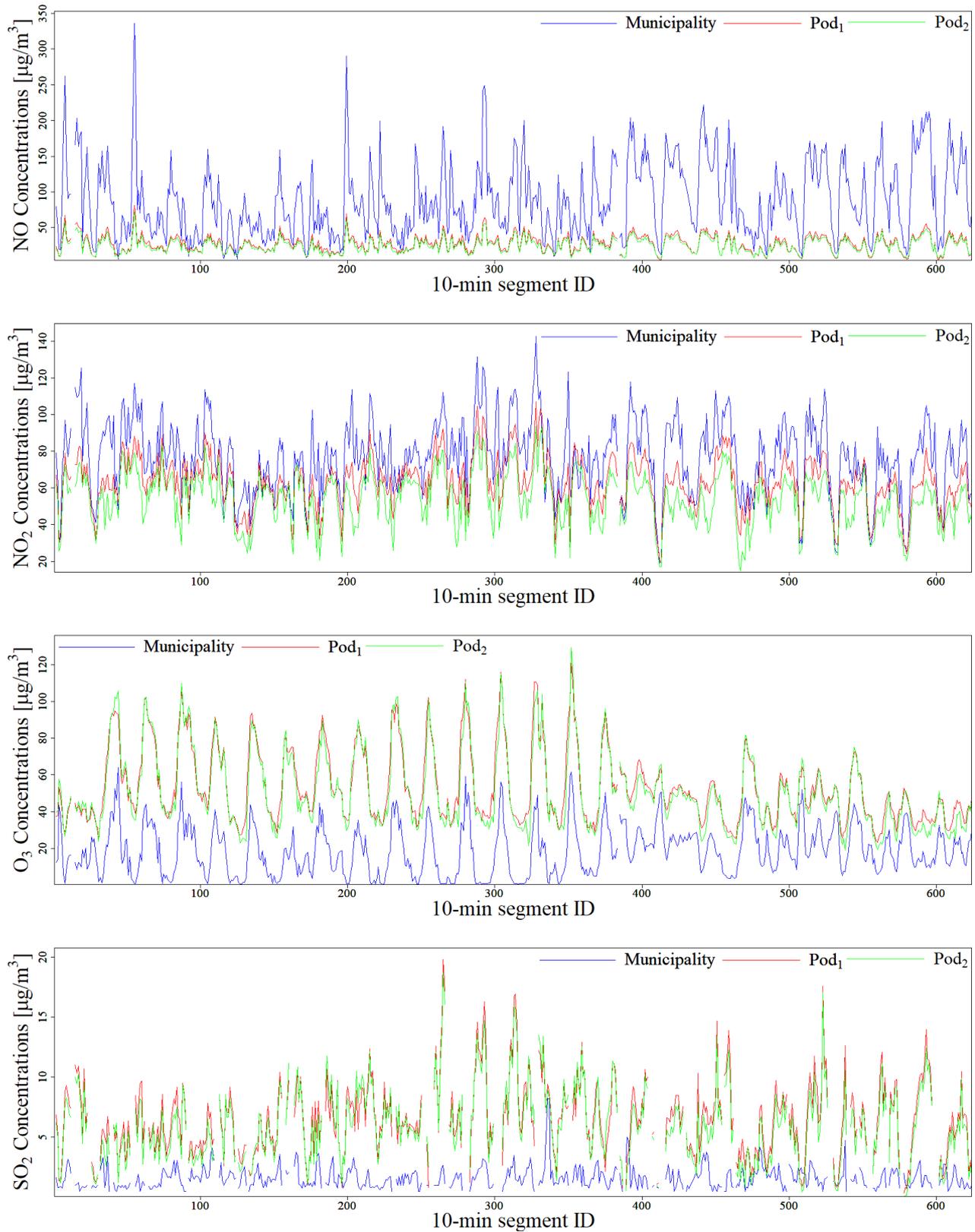
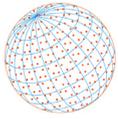
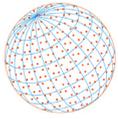


Fig. 2. Time series of the Pod₁, Pod₂, and municipal devices regarding NO, NO₂, O₃, and SO₂ pollutants.

similar pattern as that reported by the municipality. However, regarding the comparison of the respective levels of the time-series, i.e., comparing the average values of the municipality time-series against the average values of the time-series of the two pods, the two devices did not have



similarly favorable performance (Athiyarath *et al.*, 2020). The two devices either overestimated (positive bias) emissions of NO and NO₂ pollutants or underestimated (negative bias) emissions of O₃ and SO₂ pollutants.

Fig. 3 presents the scatterplots of the measurements reported by the municipal station and the Pod₁ and Pod₂ measurements along with the corresponding regression and diagonal lines. The relatively moderate spread of the points on either side of the regression line indirectly confirms the agreement of the patterns for the involved time series. By contrast, the majority of the NO and NO₂ points are above the diagonal, and the O₃ and SO₂ points are below the diagonal; this further confirms the findings of overestimation and underestimation.

The accuracy of the two measurement devices, Pod₁ and Pod₂, with reference to the respective values reported by the municipality of Thessaloniki, was quantitatively evaluated based on trend-, level- and error-related indicators (see subsection 2.3.2), and the calculations are presented in Tables 2 and 3. Both devices appeared to obtain measurements that followed the trend of the reference time series for NO, NO₂, and O₃ emissions, as indicated by the related values of the coefficient of correlation, but the device measurements did not follow the trend of the reference time series for SO₂. The average error and the level-related bias was generally low for NO₂ measurements and high for the NO, O₃, and SO₂ measurements; also, O₃ and SO₂ had remarkably high deviations from the corresponding reference series.

The measurement consistency between the two devices was also quantitatively evaluated on the basis of this set of statistical indicators, as presented in Table 4. Table 4 indicates that the consistency of the measurements between the two devices is higher than the consistency between each device and the municipal reference time series. The two devices appear to similarly measure both the trend of level of the associated time series.

3.2 Benchmarking Alternative Calibration Approaches

The comparison of the three alternative approaches' calibration performance was also based on the aforementioned statistical indicators. The indicators were calculated to assess the level of agreement between the calibrated Pod₁ reference time series, the calibrated Pod₂ and reference time series, and the calibrated Pod₁ and calibrated Pod₂ time series; these comparisons are presented in Tables 5, 6, and 7, respectively.

The three methodologies seemed to achieve the stated goal; they led to high correlation values between two time series (i.e., the calibrated and the reference ones) and to a zero-mean deviation (Tables 5 and 6). The RF approach better calibrated the individual devices (though marginally in NO and NO₂, in terms of the respective coefficient of correlation values) in terms of both the coefficient of correlation index and the average error indexes at the expense of a marginally higher but still negligible increase in the overall bias for O₃ and SO₂.

The outcomes of the calibration process seem to be more beneficial for SO₂ and O₃, since these gases appear to have a greater improvement in terms of the respective statistical indices, for both Pod₁ and Pod₂. For SO₂ and both devices, there is an average increase of the *r* index by 0.15, and an average decrease of the MAPE, nMAE, and nRMSE indices by 651%, 150%, and 157%, respectively. For O₃ and both devices, there is an average increase of the *r* index by 0.31, and an average decrease of the MAPE, nMAE, and nRMSE indices by 348%, 235%, and 288%, respectively.

The agreement between the calibrated time-series and the respective municipality series appears also to be improved for NO and NO₂ compared to the agreement between the original time-series and the municipality, however to a lower extent compared to SO₂ and O₃. In the case of NO and NO₂, though the coefficient of correlation index is marginally improved, the rest of the statistical metrics that assess the deviations between the calibrated time-series and the municipality's reference ones, are considerably improved. For NO and both devices, there is an average increase of the *r* index by 0.03, and an average decrease of the MAPE, nMAE, and nRMSE indices by 18%, 47%, and 59%, respectively. For NO₂ and both devices, there is an average increase of the *r* index by 0.05, and an average decrease of the MAPE, nMAE, and nRMSE indices by 7%, 11%, and 13%, respectively.

After calibration, as regards NO and NO₂, our results coincide with the results of previous studies, which are shown in Table 8. Before the calibration, looking into the results of Borrego *et al.* (2016), we also see a high correlation of the AQMesh sensor data with the municipal station

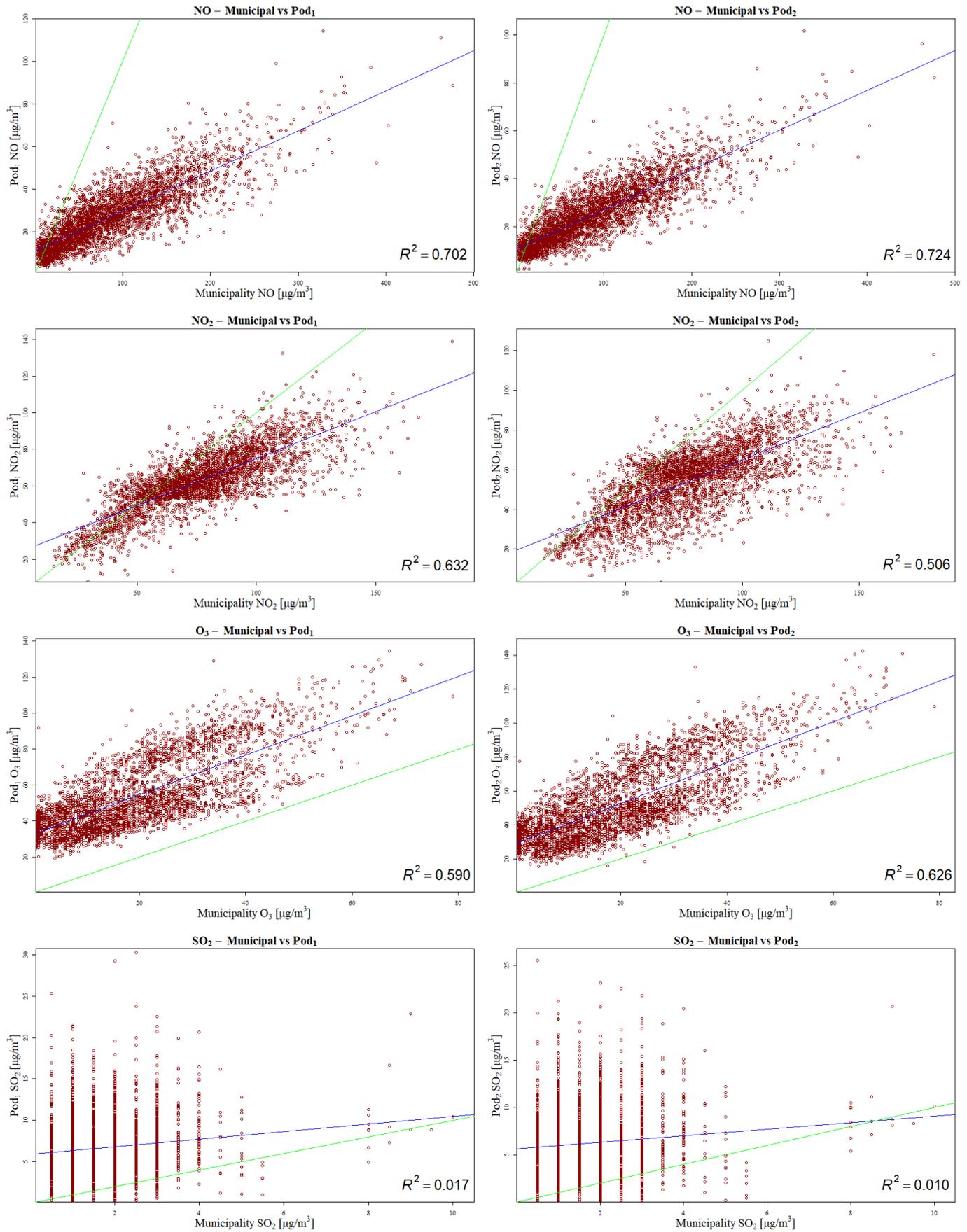
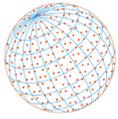


Fig. 3. Scatterplots of the municipal reference versus the Pod₁ and Pod₂ time series for NO, NO₂, O₃, and SO₂.

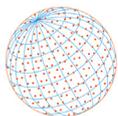


Table 2. Statistical indicators for comparing Pod₁ and municipal station values.

	Pod ₁ vs. Municipality				
	r	MAPE [%]	nMAE [%]	nRMSE [%]	rBS[%]
NO	0.84	65	70	91	69
NO ₂	0.79	19	21	27	18
O ₃	0.77	756	183	195	-183
SO ₂	0.13	421	288	354	-283

Table 3. Statistical indicators for comparing Pod₂ and municipal station values.

	Pod ₂ vs. Municipality				
	r	MAPE [%]	nMAE [%]	nRMSE [%]	rBS[%]
NO	0.85	67	72	94	72
NO ₂	0.71	29	31	38	29
O ₃	0.79	691	171	185	-171
SO ₂	0.10	390	264	331	-255

Table 4. Statistical indicators for comparing Pod₁ and Pod₂ values.

	Pod ₁ vs. Pod ₂				
	r	MAPE [%]	nMAE [%]	nRMSE [%]	rBS[%]
NO	0.992	9.6	9.7	13	8.5
NO ₂	0.926	16	14.9	17	14
O ₃	0.971	8.7	7.9	11	4.1
SO ₂	0.973	16	12	15	7.2

Table 5. Comparison of the three alternative calibration methods between the calibrated Pod₁ and the reference time series.

Indic.	r			rBS			MAPE			nMAE			nRMSE		
	RF	SLR	MLR	RF	SLR	MLR	RF	SLR	MLR	RF	SLR	MLR	RF	SLR	MLR
NO	0.87	0.84	0.87	-0.1	0.0	0.0	50	57	51	26	29	26	35	39	36
NO ₂	0.82	0.80	0.81	-0.2	0.0	0.0	16	17	17	14	15	15	19	20	20
O ₃	0.92	0.77	0.88	0.0	0.0	0.0	72	175	108	21	39	28	30	48	36
SO ₂	0.43	0.13	0.14	-0.4	0.0	0.0	57	64	53	41	45	45	55	60	60

Table 6. Comparison of the three alternative calibration methods between the calibrated Pod₂ and the reference time series.

Indic.	r			rBS			MAPE			nMAE			nRMSE		
	RF	SLR	MLR	RF	SLR	MLR	RF	SLR	MLR	RF	SLR	MLR	RF	SLR	MLR
NO	0.88	0.85	0.88	0.0	0.0	0.0	49	55	52	25	28	26	35	38	35
NO ₂	0.78	0.71	0.74	-0.1	0.0	0.0	18	21	20	16	18	17	21	23	22
O ₃	0.93	0.79	0.87	0.1	0.0	0.0	72	172	126	21	38	29	28	46	36
SO ₂	0.41	0.10	0.13	-0.4	0.0	0.0	58	65	64	41	45	45	55	60	60

($r = 0.89$ for NO and $r = 0.94$ for NO₂ in their study). This fact results to a minimal margin of gain by the application of any calibration model. Moreover, the higher frequency data (higher resolution) that we used in our study do not contribute to the aggregation of the outcomes as much as the lower resolution hourly average values that were employed in other studies. Instant picks of the gases' concentration in the ambient air cannot be measured accurately by the sensors, because of the time needed for their stabilizations.

Notably, the calibrated time series of the two devices (Pod₁ and Pod₂ in Table 7) had a favorable level of agreement in all three approaches. The calculated values for the coefficient of correlation revealed a high level of similarity between the corresponding patterns, whereas

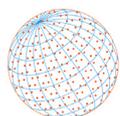


Table 7. Comparison of the three alternative calibration methods between the calibrated Pod₁ and the calibrated Pod₂ time series.

Indic. Type	r			rBS			MAPE			nMAE			nRMSE		
	RF	SLR	MLR	RF	SLR	MLR	RF	SLR	MLR	RF	SLR	MLR	RF	SLR	MLR
NO	0.98	0.99	0.99	0.1	0.0	0.0	10	12	10	8.4	5.8	4.5	11	7.8	6.2
NO ₂	0.92	0.93	0.94	-0.1	0.0	0.0	8.5	9.0	8.3	8.0	7.8	7.4	11	9.9	9.4
O ₃	0.96	0.97	0.95	0.1	0.0	0.0	20	14	64	13	11	16	18	14	20
SO ₂	0.71	0.97	0.69	0.0	0.0	0.0	14	1.8	2.2	14	1.8	2.2	19	2.4	2.9

almost zero bias was observed between the two measurements as well as between the relatively low absolute errors. This observation provides evidence of the accuracy and robustness of the devices in evaluating relative differences between different environmental scenarios and the associated measurements. This observation seems to be valid in general and irrespective of calibration method because the advantage of the RF algorithm, which related to the calibration of individual devices versus the municipal-reported values, seemed to disappear.

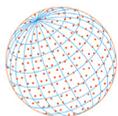
Ambient temperature and relative humidity are key factors that appear to influence the low-cost sensor measurements as indicated by the relevant literature (Borrego *et al.*, 2016; Karagulian *et al.*, 2019). Our findings are in agreement with the respective literature, according to the results of the respective analysis of variance or ANOVA-based procedures, where both ambient temperature and relative humidity also seem to have a statistically significant role in explaining the variance of the response variable, i.e., the gas pollutant's concentration. The only exceptions were the ambient temperature in the calibration of NO₂ measurements of Pod₁, as well as in the calibration of SO₂ measurements of Pod₁ and Pod₂.

The contribution of these variables in the final calibration outcomes could be assessed using the change in the respective statistical indices, when shifting from the SLR model, where the input variable is the reading of the device, to the corresponding MLR model, where the temperature and humidity were added as two extra regressor variables. In this context, and in terms of the respective statistical indices, we observe that the inclusion of these variables in the calibration model are associated to some marginal improvement for the NO, NO₂, and the SO₂ gases, and some greater improvement in the case of O₃. For the first three gases, in terms of the average change in the statistical indices (all three gases and two pods included), we observe a +0.02 increase in the r index, and a -3.7%, -1.0%, and -1.2% reduction in the MAPE, nMAE, and nRMSE, respectively. As regards O₃, on average, we observe a +0.10 increase in the r index, and a -56.5%, -10.0%, and -11.0% reduction in the MAPE, nMAE, and nRMSE, respectively. These results confirm the beneficial impact of ambient temperature and relative humidity on the calibration processes of the NO, NO₂, SO₂, and O₃ gaseous pollutants.

3.3 Comparison of the Study Outcomes to Those in the Relevant Literature

In this subsection, we compare the findings of the current study with those in the literature. To that end, we examined the outcomes of four relevant studies in which the authors tested similar AQMesh pods to determine their measurement performance and associated calibration characteristics; we also focused on the common calibration approaches between the current study and those in the literature (i.e., SLR, MLR, and RF). In the reviewed literature, a 1-h average resolution was employed in calculating the statistical metrics of the corresponding emission time series, whereas the current study used the 10-min average. The comparison was based on the Pearson's coefficient of correlation (r) because this metric is typically used by authors to evaluate calibration approaches (Table 8).

Zauli-Sajani *et al.* (2021) tested three AQMesh pods and used a hybrid RF and SLR method to calibrate the corresponding measurements to those of a reference device. The results in Table 8 refer to the respective summer and site-specific calibration outcomes of the study. Castell *et al.* (2017) examined 24 identical AQMesh devices and calibrated them to the reference device using SLR. In the study of Borrego *et al.* (2018), various low-cost sensors, including an AQMesh pod, were tested to determine measurement uncertainty and improve related performance through calibration. Jiao *et al.* (2016), studied the accuracy of several low-cost sensors, including two AQMesh pods.

**Table 8.** Correlation values for the SLR-, MLR-, and RF-based calibration approaches of the current study and four relevant studies.

Study	Pollutant	Resolution	Calibration Method	r value
Current study	NO	10 min	SLR	0.84–0.85
Borrego <i>et al.</i> (2018)	NO	1 h	SLR	0.84
Castel <i>et al.</i> (2017)	NO	1 h	SLR	0.60–0.98 (0.86)
Jiao <i>et al.</i> (2016)	NO	1 h	SLR	0.77
Current study	NO	10 min	MLR	0.87–0.88
Jiao <i>et al.</i> (2016)	NO	1 h	MLR	0.75
Current study	NO	10 min	RF	0.87–0.88
Borrego <i>et al.</i> (2018)	NO	1 h	RF	0.89
Current study	NO₂	10 min	SLR	0.71–0.80
Borrego <i>et al.</i> (2018)	NO ₂	1 h	SLR	0.91
Castel <i>et al.</i> (2017)	NO ₂	1 h	SLR	0.21–0.72 (0.49)
Current study	NO₂	10 min	MLR	0.74–0.81
Zauli-Sajani <i>et al.</i> (2021)	NO ₂	1 h	Hybrid RF+LR	0.60–0.82 (0.71)
Current study	NO₂	10 min	RF	0.78–0.82
Borrego <i>et al.</i> (2018)	NO ₂	1 h	RF	0.91
Current study	O₃	10 min	SLR	0.77–0.79
Borrego <i>et al.</i> (2018)	O ₃	1 h	SLR	0.82
Castel <i>et al.</i> (2017)	O ₃	1 h	SLR	0.09–0.81 (0.54)
Current study	O₃	10 min	MLR	0.87–0.88
Current study	O₃	10 min	RF	0.92–0.93
Zauli-Sajani <i>et al.</i> (2021)	O ₃	1 h	Hybrid RF+LR	0.97–0.98 (0.97)
Borrego <i>et al.</i> (2018)	O ₃	1 h	RF	0.89

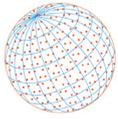
Regarding the calibration of devices for NO measurements, all five study groups (i.e., the four previous studies and the current one), had similar calibration performance irrespective of the applied calibration approach. The corresponding *r* values for the AQMesh pods that were examined fell within the range of 0.60–0.98. With respect to the corresponding NO₂ readings of the devices, a wide range of calibration outcomes was observed, as indicated by the associated *r* value range (0.21–0.91). A narrower range (0.60–0.91) was reported for the three studies in which the RF algorithm had been employed. Finally, regarding O₃ measurements, the three studies with O₃-related results had high correlation values under the RF calibration approach, and the related outcomes were better than the those for MLR and SLR (for the two studies in which SLR was also applied as an alternative calibration methodology). Overall, the O₃ *r* values ranged from 0.09 to 0.98, but this wide range was mainly due to the measurements of Castel *et al.* (2017), who examined 24 pods, because the corresponding range of the other studies was 0.77–0.98, indicating much less variance in calibration outcomes among the examined devices.

The performance of the alternative calibration methodologies was marginally better for NO and O₃ emissions than for NO₂ emissions. The RF algorithm appeared to achieve a marginally better performance than such methodologies in some cases. Nevertheless, the simpler linear regression approach appeared to have similar performance to the more complex RF algorithm.

In general, a part of this agreement in terms of the coefficient of correlation is due to the fact that all devices that were calibrated in the different studies originate from the same manufacturer, i.e., AQ Mesh, therefore, some kind of agreement was expected. Additionally, the measurements, in several cases, took place under similar environmental conditions.

4 CONCLUSIONS

Air quality monitoring platforms are complementary systems for measuring concentrations of emission gases as well as environmental parameters. Each year, their reliability and accuracy increase, and they have become a valuable component of monitoring station networks and offer portability, thus supplying detailed spatial environmental information. Mobile platforms that use low-cost sensors still lack accuracy, and thus, data validation and calibration are essential.



In the present study, we tested two identical AQMesh devices that are widely used in several countries for 4 weeks in a two-fold manner, that is, we estimated their measurement accuracy based on data gathered from field measurements at an urban location and examined their potential calibrations. To that end, device performance was evaluated based on a set of statistical indexes, and three widely used alternative calibration approaches were employed. The reference for all evaluations was the respective measurement data reported by the associated municipal station.

The two devices, Pod₁ and Pod₂, seemed to have favorable performance in following the patterns (trend) of the reference time series (i.e., the emission concentrations reported by the municipal station [10-min average resolution]). Nevertheless, their performance was less favorable in terms of respective average error and overall bias. The NO₂ measurements appeared to be the most accurate among the four pollutants, whereas the SO₂ measurements appeared to be the least accurate. The calibration improved the agreement between the device and reference measurements in terms of time series patterns, average error, and overall bias.

A notable and practically useful finding is that the measurements of the two devices were consistent with each other both before (i.e., in their “raw” form) and after calibration stage (i.e., in their “calibrated” version); the corresponding bias was almost zero. The consistency between the calibrated time series of the two measurement devices was significantly better, in terms of the respective patterns, errors, and biases, than that between an individual device and the reference municipal station measurement. Therefore, similar devices could be used in studies wherein the relative differences in emissions, rather than the respective absolute scale values, are the main point of interest.

The calibration performance of the three alternative approaches was generally similar. The RF algorithm seemed to have an advantage in several cases, mostly in terms of following the pattern in the O₃ and SO₂ time series, but also in terms of the associated error and bias metrics.

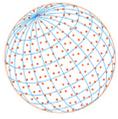
The calibration results appeared to agree with the findings of similar studies, and some high coefficients of correlation were noted. A set of indexes used in the overall evaluation, including the coefficient of correlation, was calculated at a higher frequency (10-min average) than that of the 1-h average in the literature. When both the initial raw measurements of the devices and the outcomes of calibration processes in the current study and similar studies were taken into account, low-cost sensors appear to be an important alternative solution for air quality monitoring.

ACKNOWLEDGMENTS

We would like to thank Dr. A. Kelesis from the Directorate of Urban Environment Management of the Municipality of Thessaloniki for sharing with the authors the reference data of the city centre air quality monitoring station.

REFERENCES

- AQMesh (2020). <https://www.aqmesh.com/> (accessed 15 December 15 2020).
- Athiyarath, S., Paul, M., Krishnaswamy, S. (2020). A comparative study and analysis of time series forecasting techniques. *SN Comput. Sci.* 1, 175. <https://doi.org/10.1007/s42979-020-00180-5>
- 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., *et al.* (2016). Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise. *Atmos. Environ.* 147, 246–263. <https://doi.org/10.1016/j.atmosenv.2016.09.050>
- 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., *et al.* (2018). Assessment of air quality microsensors versus reference methods: The EuNetAir Joint Exercise – Part II. *Atmos. Environ.* 193, 127–142. <https://doi.org/10.1016/j.atmosenv.2018.08.028>
- 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? *Environ. Int.* 99, 293–302. <https://doi.org/10.1016/j.envint.2016.12.007>
- Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., Judge, R., Caudill, M., Rickard, J., Davis, M., Weinstock, L., Zimmer-Dauphinee, S., Buckley, K. (2016). Community Air Sensor Network (CAIRSENSE) project: Evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmos. Meas. Tech.* 9, 5281–5292. <https://doi.org/10.5194/amt-9-5281-2016>
- Karagulian, F., Barbiere, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., Redon, N., Crunaire, S., Borowiak, A. (2019). Review of the performance of low-cost sensors for air quality monitoring. *Atmosphere* 10, 506. <https://doi.org/10.3390/atmos10090506>
- Kassomenos, P.A., Kelessis, A., Paschalidou, A.K., Petrakakis, M. (2011). Identification of sources and processes affecting particulate pollution in Thessaloniki, Greece. *Atmos. Environ.* 45, 7293–7300. <https://doi.org/10.1016/j.atmosenv.2011.08.034>
- Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., Britter, R.R. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environ. Int.* 75, 199–205. <https://doi.org/10.1016/j.envint.2014.11.019>
- Marco, S., Gutiérrez-Gálvez, A. (2012). Signal and data processing for machine olfaction and chemical sensing: A review. *IEEE Sens. J.* 12, 3189–3214. <https://doi.org/10.1109/JSEN.2012.2192920>
- Pasxalia, A. (2019). Study the air quality in the area of Thessaloniki, Greece. Master thesis, School of Physics, AUTH, Thessaloniki, Greece.
- R Core Team (2021). R: A language and environment for statistical computing. Vienna, Austria.
- 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. *Environ. Int.* 106, 234–247. <https://doi.org/10.1016/j.envint.2017.05.005>
- Spinelle, L., Gerboles, M., Villani, M.G., Aleixandre, M., Bonavitacola, F.F. (2015). Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sens. Actuators, B* 215, 249–257. <http://dx.doi.org/10.1016/j.snb.2015.03.031>
- Vlachokostas, Ch., Achillas, Ch., Moussiopoulos, N., Hourdakis, E., Tsilingiridis, G., Ntziachristos, L., Baniyas, G., Stavrakakis, N., Sidiropoulos, C. (2009). Decision support system for the evaluation of urban air pollution control options: Application for particulate pollution in Thessaloniki, Greece. *Sci. Total Environ.* 407, 5937–5948. <https://doi.org/10.1016/j.scitotenv.2009.07.040>
- Zauli-Sajani, S., Marchesi, S., Pironi, C., Barbieri, C., Poluzzi, V., Colacci, A. (2021). Assessment of air quality sensor system performance after relocation. *Atmos. Pollut. Res.* 12, 282–291. <https://doi.org/10.1016/j.apr.2020.11.010>