COMPAIR

DELIVERABLE

# D3.5 Data Processing and Data Calibration of Air Quality Sensors

| | | |
|---|---|---|
| Project Acronym: | **COMPAIR** | |
| Project title: | Community Observation Measurement & Participation in AIR Science | |
| Grant Agreement No. | 101036563 | |
| Website: | www.wecompair.eu | |
| Version: | 1.1 | |
| Date: | 25 October 2024 | |
| Responsible Partner: | IMEC | |
| Contributing Partners: | - | |
| Reviewers: | Christophe Stroobants (VMM)<br>Carolina Doran (ECSA)<br>Andrew Scott<br>Karel Jedlicka<br>Martine Van Poppel (VITO)<br>Karen van Campenhout (NC) | |
| Dissemination Level: | Public | X |
| | Confidential, only for members of the consortium (including the Commission Services) | |

This project has received financial support from the European Union's Horizon 2020 Programme under grant agreement no. 101036563

# Revision History

| Version | Date | Author | Organization | Description |
|---|---|---|---|---|
| 0.1 | 03.01.24 | Burcu Celikkol | IMEC | Initial structure |
| 0.2 | 14.02.24 | Shaojie Zhuang | IMEC | Input added |
| 0.3 | 16.02.24 | Jasper Fabius | IMEC | Input added |
| 0.4 | 23.02.24 | Burcu Celikkol | IMEC | Ready for review |
| 0.5 | 03.03.24 | Carolina Doran | ECSA | Internal review |
| | 06.03.24 | Charalampos Alexopoulos | UAEG | Internal review |
| | 11.03.24 | Christophe Stroobants | VMM | Internal review |
| 0.6 | 14.03.24 | Burcu Celikkol | IMEC | Ready for external review |
| 0.7 | 21.03.24 | Karel Jedlicka | UWB (CZ) | External review |
| | 25.03.24 | Andrew Stott | Independent (UK) | External review |
| | 27.03.24 | Martine Van Poppel | VITO | External review |
| 1.0 | 27.03.24 | Burcu Celikkol | IMEC | Final version |
| 1.1 | 25.10.24 | Burcu Celikkol | IMEC | Update with results until October 2024 |

# Table of Contents

Executive Summary 7
1. Introduction 8
2. Sensor Types 9
2.1. PM Sensing Device (SODAQ Air) 9
2.2. $NO_2$ Sensing Device (OnePlanet NitroSense) 11
3. Calibration Approaches and Description 13
3.1 Factory Calibration 13
3.2 Field Calibration 13
3.3 Distant Calibration 14
3.3.1 Applying Distant Calibration Approach in COMPAIR 15
3.3.1.1 Technical Limitations 15
3.3.3.2 Practical Limitations 17
3.3.3.3 Additional Functionalities 18
3.3.2 Optimization of Distant Calibration (October 2024) 18
3.4 Auto-Calibration (October 2024) 19
4. Technical Implementation 19
4.1 Pipeline PM 20
5.2 Pipeline $NO_2$ 21
5. Evaluation of Distant Calibration 22
5.1 PM 22
5.2 $NO_2$ 25
5.2.1 $NO_2$ Results (March 2024) 25
5.2.1 $NO_2$ Results (October 2024) 27
6. Evaluation of Auto-Calibration (October 2024) 28
6.1 $NO_2$ 28
6.1.1 Preliminary Evaluation 28
6.1.2 Evaluation using COMPAIR data 29
6.1.3 Comparison of distant calibrated and auto-calibrated data 31
7. Conclusions and Discussion 33
8. References 34
Books/Articles 34
Websites 34
9. Supplementary Information 35

# List of Tables

Table 3.1 Distant calibration parameters that were optimised and the optimization range. 19
Table 5.1 Summary of the SODAQ AIR devices used for the performance evaluation. 22
Table 5.2 The performance evaluation period and location of the NitroSense sensor boxes deployed at station BETR702* in Ghent (Belgium). 25
Table 6.1 Correlative statistics of the calibrated data using different calibration approaches: Pearson correlation coefficient (r), coefficient of determination (R2) and mean absolute error (MAE). 29
Table 6.2 Descriptive statistics of the device data: mean, median, SD (standard deviation), min (minimum), max(maximum) during the period. Unit is $\mu$g/m3. 30
Table 9.1 Comparisons between the original and field calibrated SODAQ AIR measurements on PM10 and PM2.5 throughout the complete colocation period at the reference station. 35
Table 9.2 Comparisons among the original, distant calibrated and field calibrated SODAQ AIR measurements on PM10 and PM2.5 over the period during which distant calibration was available. 36
Table 9.3 Comparisons among the distant, factory and field calibrated NitroSense measurements on NO2 over the colocation period in Ghent, Belgium. 37

# List of Figures

Figure 2.1 Red box: Location of the optical sensor (Sensirion SPS30) inside the a disassembled SODAQ AIR device (Source: COMPAIR D3.2) 9
Figure 2.2 Working principle of an optical PM sensor. 9
Figure 2.3 Location of electrochemical NO2 sensor (Alphasense A43F) on the OnePlanet NitroSense device (Source: OnePlanet Research Center). 11
Figure 2.4 Working principle of an electrochemical gas sensor. Gas molecules react at the sensing electron surface, producing or consuming electrons that are supplied by the counter electrode. 11
Figure 3.1 Daily variability of pollutant concentrations from five high-end monitoring stations in Antwerp, Belgium. Left: PM2.5, Middle: PM10 and NO2 (right). Source: Hofman et al. (2022) 14
Figure 3.2 Calculated hourly variance between the five monitoring stations in Antwerp, Belgium. Night hours are highlighted in light blue when the variance is considered low. Source: Hofman et al. (2022) 14
Figure 3.3 Plots showing the data availability of NO2 reference station data from DiscoMap. The y axis represents the reference station codes and scale bar shows NO2 sensor concentration readings (µg·m-3). White gaps show the periods where data was not found. Each heatmap shows data availability from a pilot region. 17
Figure 4.1 Main architecture of the calibration pipeline set up in the cloud environment. In the Calibration part, a calibration model is built per sensor per device. In the Inference part, the latest calibration model is applied to sensor measurements of the last day. 20
Figure 5.1 Hourly averaged uncalibrated (raw), distant calibrated (dist) and field calibrated SODAQ AIR measurements on PM10. 24

Figure 5.2 Hourly averaged uncalibrated (raw), distant calibrated (dist) and field calibrated SODAQ AIR measurements on PM2.5. 24

Figure 5.3 Time series of NO2 measurements taken by NitroSense devices during the colocation period with the reference station BETR702. Each NitroSense device consisted of two NO2 sensors, which were denoted by the suffix. 26

Figure 5.4 Reference data (black) and NitroSense data calibrated using optimised distant calibration algorithm (orange) of one device deployed for project LIFE Critical at Rotterdam Statenweg. 27

Figure 6.1 Reference data (black) and NitroSense data calibrated using optimised distant calibration algorithm (green) of two devices deployed for project LIFE Critical at Rotterdam Statenweg. 28

Figure 6.2 The box plots show that the 5 NitroSense sensors co-located at Ghent reference station, represented by the different colors are measuring in a similar range. 30

Figure 6.3 Time series comparison of the auto-calibrated data from 5 devices and the data from the reference monitor. 31

Figure 6.4 Scatter plots where NitroSense data (y-axis) is plotted against the reference monitor data (x-axis). 31

Figure 6.5 R2 values comparing sensor to monitor. 31

Figure 6.6 Distribution of distant calibrated NO2 concentrations calibrated using the approach described in section 3.3 (source: VMM, D5.6 Public Round Report). 32

Figure 6.7 Distribution of auto-calibrated NO2 concentrations calibrated using the auto-calibration approach described in section 2.2. 32

Figure 9.1 Scatter plots of SODAQ AIR device deployed at Antwerp (ID ***7833) and Berlin (ID ***2725) against the colocated reference station on PM10 and PM2.5 measurements. Solid lines denote the regression line of total least squares, and dashed lines the perfect match. 38

# List of Abbreviations

| Abbreviation | Definition |
|---|---|
| PM | Particulate matter |
| $NO_2$ | Nitrogen dioxide |
| O3 | Ozone |
| OEM | Original equipment manufacturer |
| RIVM | National Institute for Public Health and the Environment |
| API | Application Programming Interface |
| EEA | European Environmental Agency |
| DiscoMap | Discovery map (data collection of environmental data from EEA) |
| IRCEL - CELINE | API of the Belgian Interregional Environment Agency |
| OGC-STA | Open Geospatial Consortium SensorThings API |
| FROST | The Fraunhofer Open source SensorThings Server |
| MAE | Mean absolute error |
| RMSE | Root mean square error |
| MSE | Mean squared error |
| MAD | Median absolute deviation |
| Pearson's r | Pearson correlation coefficient |
| R2 | Coefficient of determination |
| kU | Expanded uncertainty |
| LCS | Low-cost sensor |

# Executive Summary

COMPAIR aims to evaluate the effects of urban mobility experiments performed by citizens in the five pilot regions on traffic and air quality. Citizen science air quality sensors typically involve simple components, and their measurement output can be sensitive to the surrounding environment. This can lead to collecting data that deviates from the “true value”; especially in the case of outdoor measurements where the sensor is exposed to a wide range of environmental conditions. Improving the sensor output through validated calibration methods allows a better evaluation of the urban experiments to optimally steer potential behavioural and policy changes.

One calibration approach for outdoor air quality measurements is to deploy all sensors next to a high-end outdoor reference station of known high accuracy (“field calibration”) before deploying the sensor at the location of interest. However, this approach is not always feasible, lacks scalability, is resource intensive and does not account for effects such as ageing of the sensor. Here, we describe an alternative calibration approach that is applied in COMPAIR which aims to improve the sensor measurements without the need for deployment of all sensors at a reference site: distant calibration. We describe the algorithm which uses reference station data at-a-distance to train a multilinear regression model which is used to calibrate low-cost sensor data with a rolling time window. The calibration is implemented via a cloud pipeline in COMPAIR. In the pipeline, the raw data from 250 particulate matter sensors (PM: SODAQ AIR) and 15 nitrogen dioxide sensors ($NO_2$: OnePlanet NitroSense) are ingested, calibrated and exposed to COMPAIR data manager and made available to the applications such as the Policy Monitoring Dashboard (PMD). We also describe the modifications to the algorithm that were introduced to accommodate the use cases in COMPAIR and deviations from the original implementation of the distant calibration approach.

We evaluate the performance of the distant calibrated PM and $NO_2$ output in comparison with raw output of the low-cost sensors and typically well-performing field calibration. We evaluate the data in terms of measurement error, biases, correlation and measurement uncertainty. Our main findings show that applying distant calibration results in some improvements in PM output in terms of absolute error compared to the raw data, while it improves measurement uncertainty of only some of the tested devices. In the case of $NO_2$, the distant calibration approach leaves room for improvement. In the v1.1 version of this deliverable published in October 2024, we describe further efforts in improving $NO_2$ calibration, including applying a different calibration approach called auto-calibration which shows significant improvements over distant calibration. The re-calibrated $NO_2$ data is made available to the pilots in the public round, enabling air quality analyses with improved accuracy.

# 1. Introduction

Sensor data calibration is crucial to ensure accurate and reliable measurements, which helps end users make well-informed analyses and interpretations. In COMPAIR, low-cost citizen science sensors (LCS) are deployed mainly outdoors to monitor the impact of urban mobility implementations on traffic and air quality. Low-cost air quality sensors make use of simple measurement principles (see section 2) that differ significantly from the reference methods and as such they are prone to drift, ageing and their output is sensitive to environmental conditions such as temperature, humidity and interfering pollutants.

In COMPAIR, a novel sensor calibration approach called “distant calibration” is applied and evaluated to calibrate PM and $NO_2$ sensor data. The approach combines high-end reference station data with low-cost sensor data to account for sensitivity of sensors to environmental conditions and the changes they undergo during their deployment outdoors. The approach enables automated near-real-time calibration of sensor data without the need for lengthy and labour-intensive laboratory or field testing.

This deliverable first describes the air quality sensing devices that are calibrated (section 2) to describe the underlying physicochemical processes that influence raw sensor output. In section 3, we describe the calibration approach as well as the challenges faced in the implementation of the approach during the project. Section 4 explains the technical implementation of the approach which involves training and applying a calibration model for each sensor with dynamically updated parameters in the cloud environment. Section 5 and 6 evaluate the performance of different calibration approaches.

# 2. Sensor Types

Data from two types of sensors are calibrated in COMPAIR: SODAQ AIR PM sensing device and OnePlanet NitroSense $NO_2$ sensing device. This section covers the working principles of each sensor and why a calibration strategy is needed for continuous reliable measurements when the sensor of interest is deployed outdoors in varying environmental conditions.

## 2.1. PM Sensing Device (SODAQ Air)

SODAQ AIR device that measures airborne particulate PM2.5 and PM10 uses a Sensirion SPS30 sensor which is based on light scattering. The following figures show the location of the sensor within the device and its working principle.

***Figure 2.1*** *Red box: Location of the optical sensor (Sensirion SPS30) inside the a disassembled SODAQ AIR device (Source: COMPAIR D3.2)*

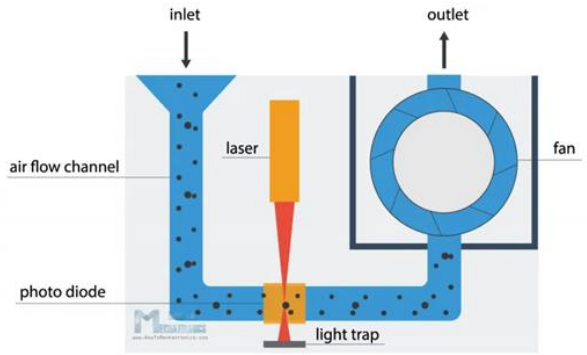


***Figure 2.2*** *Working principle of an optical PM sensor.*

As seen in the figure above, the sensor makes use of a fan which draws in air through an inlet. The air flows through a laser diode channel; where the particles dispersed in air cause the laser light to scatter and hit the light sensor. The non-scattered laser light is absorbed into a special surface to avoid detection of scattered light that does not result from the particle-laser interaction. A microcontroller unit processes the data measured by the photodiode to determine the size and concentration of particles from the collected data. Then, the air is exhausted out of the unit.

Such sensors can benefit from sensor data calibration, which can minimise the fluctuations, drift or deviations in the raw sensor response due to the following points:

- Exposure to environmental conditions: Particle size can deviate from “true size” in different environmental conditions such as high humidity due to water absorption of the particle. Whereas a low-cost PM sensor would register this change as an enlarged particle size, a high end PM sensor makes use of a conditioned air inlet which minimises such effects.
- Ageing: Parts of the sensor such as the fan or laser can age over time or dust accumulation in the unit can occur which can change the baseline response of the sensor.
- Particle composition: the various sources of particulate matter lead also to varying particle composition, which in turn will change the surface, shape and reflective properties of individual particles. Hence a measurement method relying on light scattering by particles will be affected by changes in the composition. Assuming particle composition is fairly homogeneous in a certain area, calibration can help correct for these effects (i.e. more salts near coasts, more tyre abrasion in cities, more wood burning in rural areas).
- Difference in measurement principles: LCS estimate PM size based on the interaction of light with particles, whereas high-end sensors are based on different measurement principles such as separating particle size via a cyclone and detecting particle size and concentration based on attenuation of radiation.

Therefore we apply a calibration algorithm to minimise these influences. The calibration approach is described in section 3.

## 2.2. $NO_2$ Sensing Device (OnePlanet NitroSense)

OnePlanet NitroSense device that measures gaseous nitrogen dioxide uses an electrochemical Alphasense A43F $NO_2$ sensor. The following figures show the location of the sensor on the device and its working principle.


***Figure 2.3*** *Location of electrochemical NO2 sensor (Alphasense A43F) on the OnePlanet NitroSense device (Source: OnePlanet Research Center).*

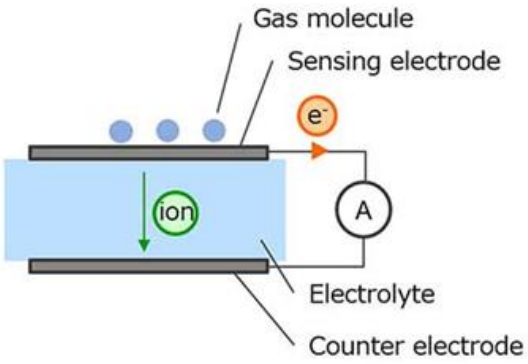


***Figure 2.4*** *Working principle of an electrochemical gas sensor. Gas molecules react at the sensing electron surface, producing or consuming electrons that are supplied by the counter electrode.*

In an electrochemical sensor, gas molecules undergo a chemical reaction at the sensing electrode surface. The reaction produces electrons which flow between the sensing and counter electrodes, and the generated current can be either positive or negative depending if

oxidation or reduction occurs at the electrodes. The magnitude of the current is proportional to the gas concentration, and with the help of a reference electrode the sensor electronics process and amplify the electronic signal that is detected as the output. Gas sensors may have a chemical filter to capture potentially interfering gases that the sensor may be sensitive to - called cross-sensitivity - before the gas makes contact with the sensing electrode surface. In the case of $NO_2$, Alphasense NO2-A43F is equipped with an ozone (O3) filter.

The measured current is expressed in nanoAmperes (nA) and can be converted into a pollutant concentration using calibration factors derived from measurements in the manufacturer's testing facility. Typically, these measurements include exposing each sensor to two distinct pollutant concentrations: one at 0 ppm (parts per million mole fraction) and the other at the so-called span level depending on the designed range of the sensor.

The zero-response and the sensitivity of a sensing unit can vary greatly even within the same batch of production, e.g. an $NO_2$ sensor may have a sensitivity of -150 nA per ppm of pollutant, whereas another from the same batch may respond -450 nA / ppm. Therefore even the fresh out-of-the-box sensors show different nA output when exposed to the same concentration of airborne pollutant.

Once the sensing device is deployed in outdoor conditions as intended, environment-related parameters such as outdoor temperature, humidity and cross-interfering pollutants also have an influence on the output signal. Additionally, the sensor slowly degrades over time as the sensor is deployed in the field which is also a function of the conditions that the sensor is exposed to. Due to these reasons, gas sensors can also greatly benefit from calibration. In the next section, we describe different calibration approaches, their benefits and drawbacks.

# 3. Calibration Approaches and Description

There are several calibration approaches that can be applied to a sensor that will be deployed outdoors. Some of the approaches can be listed as the following:

## 3.1 Factory Calibration

Factory calibration refers to obtaining the calibration coefficients from the test facilities of the manufacturer. The test facilities are often controlled environments where temperature, humidity and concentration of the target gas or particle types are fixed. If such testing is performed after the sensor has been shipped by the manufacturer, this process can also be called laboratory calibration. The obtained calibration coefficients are typically incorporated in ready-to-use products like the SODAQ AIR, whereas an OEM (original equipment manufacturer) component or prototypes like the Alphasense $NO_2$ sensor would require the users and developers to apply the calibration accordingly.

While factory calibration is a convenient solution, it has several downsides. Firstly, the sensor response to its environment will change over time due to factors like hardware degradation, contamination and decrease of consumable parts such as the electrolyte of an electrochemical sensor. Typically, periodic and even frequent recalibration is required to compensate for these changes and thereby maintain the performance of low cost sensors. Secondly, the calibration coefficients are valid for the set of environmental conditions in which the sensors are tested in the laboratory (e.g. 60% RH, 20°C). The factory calibration may work well in one circumstance but not another.

## 3.2 Field Calibration

Field calibration refers to obtaining calibration coefficients under a field condition which ideally could resemble the actual environment where the sensors will be measuring. Field calibration typically involves colocating the sensors with one or multiple more reliable measurement devices by placing them side by side and acquiring data for a certain amount of time. This approach allows incorporating factors that are difficult to know in advance or to recreate in a conventional laboratory environment.

A drawback, however, is the lack of control over these factors. Therefore, field calibration may require a long test period under sufficiently varying conditions in order to obtain a proper data set. Moreover, the availability and accessibility of the reference devices also constraint the practicality of the approach. It may not be possible to obtain the permit to access reference sites that are operated by local environmental agencies, especially if a large number of sensors are to be tested.

## 3.3 Distant Calibration

Given the drawbacks of the previous standard approaches, in COMPAIR we explore an alternative novel calibration approach developed within the VLAIO City of Things project called distant calibration (Hofman et al. 2022). Distant calibration aims to account for sensor sensitivities related to exposure to environmental conditions (like temperature and humidity), and long-term ageing effects by calibrating sensor measurements using publicly available high-end reference station measurements during hours of low human activity where pollutant concentrations are low and relatively homogeneously distributed. The homogeneity of pollutants during certain hours of the day is demonstrated using data from Antwerp reference stations in Figure 3.1.

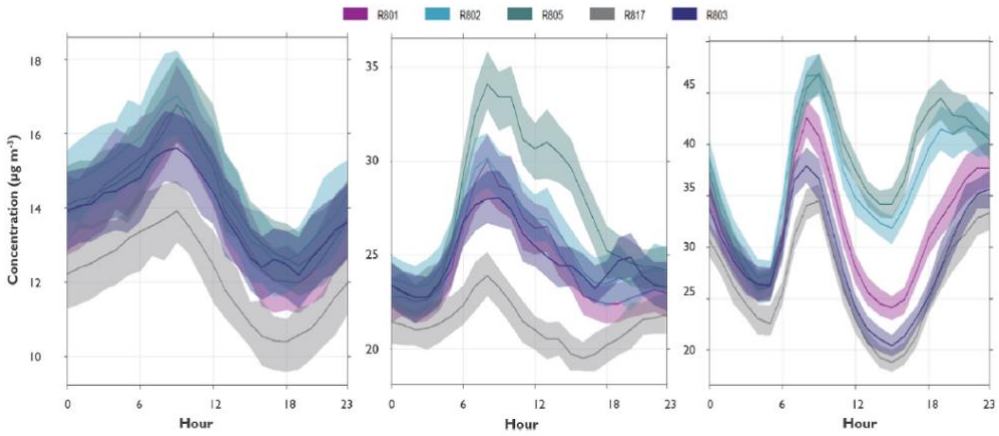


***Figure 3.1*** *Daily variability of pollutant concentrations from five high-end monitoring stations in Antwerp, Belgium. Left: PM2.5, Middle: PM10 and NO2 (right). Source: Hofman et al. (2022)*

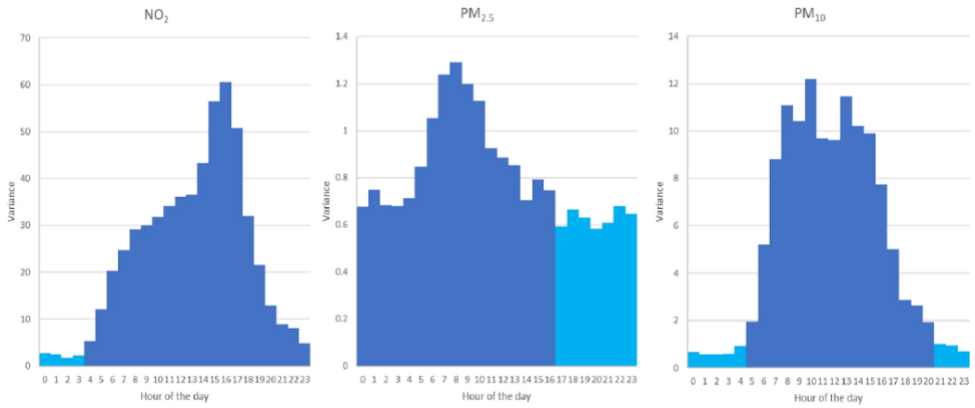


***Figure 3.2*** *Calculated hourly variance between the five monitoring stations in Antwerp, Belgium. Night hours are highlighted in light blue when the variance is considered low. Source: Hofman et al. (2022)*

The algorithm uses the following steps:

- Collect measurements from the last 34 days of all reference stations within 15 km from the sensor location.

- Select the reference station that shows the highest similarity with the raw sensor measurements, in terms of relative fluctuations over time. This selection is performed using pairwise Pearson's correlation of the timeseries of the reference stations and the sensor.
- Slice the data to only include measurements taken at night time. This step is included because during night time the variability between reference stations tends to be lower than during day time (see Figure 3.2) i.e., the assumption is that if all reference station measurements are roughly equal during these time points, so should the measurements of the sensor be.
- Perform preprocessing of the data (e.g. remove outliers, apply transformation).
- Build a multilinear regression model using sensor, local temperature and local relative humidity measurements to estimate the measurements from the selected reference station. In the case of $NO_2$, cross-interfering pollutant ozone (O3) is also used as a parameter.

$$NO_{2,\,ref} \sim a + b_1 * NO_{2,\,sensor} + b_2 * T_{sensor} + b_3 * RH_{sensor} + b_4 * O_{3,\,ref}$$

- The model and the extracted parameters a, $b_1$, $b_2$, $b_3$ and $b_4$ are used to calibrate the low-cost sensor data.
- A moving window of the previous 34 days of collected measurements is used to update the parameters on a daily basis.

The distant calibration algorithm was developed and validated using sensors deployed in 5 experiments ("testbeds") in total (3 unique testbeds NO2, 3 unique testbeds for PM) in Belgium and the Netherlands. Different types of sensors were used for the validation: SDS011, Alphasense OPC-N3 and Sensirion SPS30 for PM, Alphasense NO2-A43F for $NO_2$. Previously, the validation results showed that the calibration improves low-cost sensor data significantly where both sensor measurement accuracy and correlation relative to the reference monitor were significantly increased after the calibration.

### 3.3.1 Applying Distant Calibration Approach in COMPAIR

In COMPAIR, the main objective of using and further developing the distant calibration approach was that data with potentially higher accuracy can be provided to the pilot experiments in near real-time that helps the citizens and policy makers to evaluate effects the urban pilot experiments have on air quality. Furthermore, the large number of sensors and different pilot locations in Belgium, Germany, Greece and Bulgaria would allow evaluating the scalability and cross-site applicability of the approach; especially in the pilot locations with significantly different climate conditions to Belgium and the Netherlands.

The inclusion of mobile sensors and coverage of different European regions inspired modifications in the implementation of the distant calibration approach. Furthermore, related to the scale-up in the number of sensors we also experienced challenges in implementation and evaluation of the approach. These challenges, the applied mitigation strategies and corresponding risks are described in this section.

#### 3.3.1.1 Technical Limitations

Mobile sensors with low time coverage stored indoors for the night

In distant calibration, the calibration model is trained during nighttime hours where the pollutant concentrations are relatively homogeneously distributed as described in section 3.3. Additionally, a time coverage threshold of 75% was applied in the original approach, i.e. if the data that did not meet this time coverage, a new model was not built. If the new model is not built, then the calibration parameters from the previous model remain in use.
One challenge related to mobile SODAQ Air sensors was that the citizens typically stored the sensors indoors or switched them off during the night after removing them from backpacks and bikes, and the indoor sensor readings should not be combined with the outdoor pollutant data to build a calibration model. Furthermore, the time coverage of SODAQ Air sensors was low since they were often only used during trips. To mitigate these points, a different approach to selecting the training time window was applied where instead of using nighttime data for training, data with low variance between the surrounding reference stations were used instead (see section 5.1). Additionally, the time coverage threshold was decreased to cover a minimum of 30 hours during the training period.

### Low spatial density of reference stations

Ideally, to apply the distant calibration algorithm, a high density of reference stations around the low-cost sensors present in a similar microenvironment (e.g. urban, rural) is required. However, a low number of reference stations were present in some pilot regions, such as Plovdiv and Athens. Therefore the original reference station search perimeter was expanded from 15 km to 30km for PM and 50 km for $NO_2$. The accepted risk was that the algorithm may have worse results in all regions due to this modification; the more distance between LCS and reference sensor, the less representative the conditions which may lead to biases in the calibration parameters.

### Delayed or missing reference data

In the formerly designed calibration pipeline, reference station data was collected through the API of the Dutch National Institute for Public Health and the Environment (RIVM) which is generally reliable in terms of timeliness and completeness of data. In the COMPAIR implementation, we first performed selection of a data source that can be used as a source for reference data for all pilot locations. After evaluating a few options, the European Environmental Agency (EEA) Discovery map (DiscoMap) dataset was found to be most convenient, since it includes data from all pilot locations. However, data gaps from reference stations were found to be common and it was also often observed that data is delayed by one to several days. This could be due to the fact that DiscoMap relies on the original data sources to expose data and not all sources may have the same data availability frequency. One modification that was implemented due to this delay was that the calibration model was built weekly rather than daily to increase the likelihood that reference data is present.

Figure 3.3 visualises common data gaps in a post-hoc analysis.

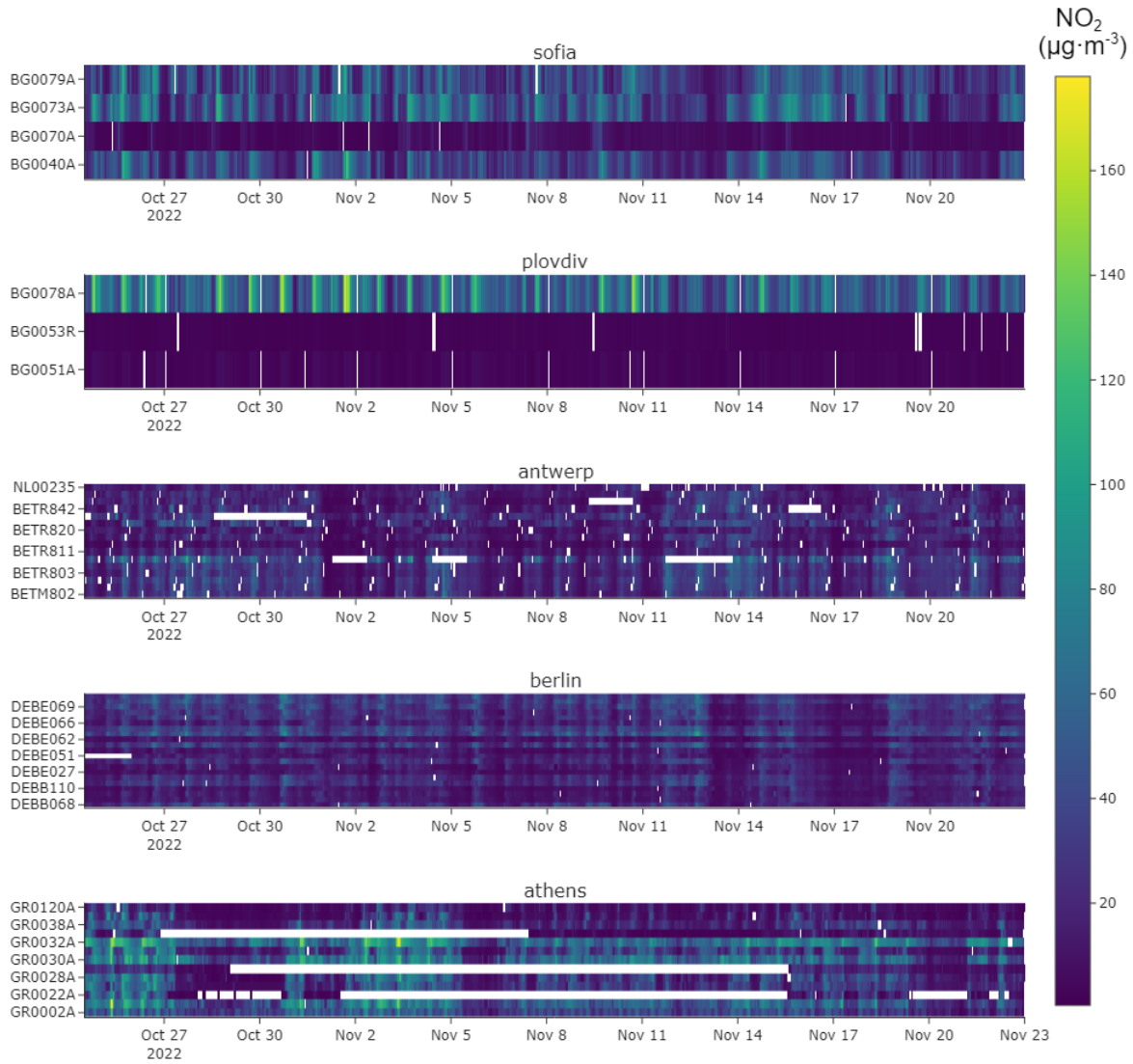


***Figure 3.3*** *Plots showing the data availability of $NO_2$ reference station data from DiscoMap. The y axis represents the reference station codes and scale bar shows $NO_2$ sensor concentration readings ($\mu g \cdot m^{-3}$). White gaps show the periods where data was not found. Each heatmap shows data availability from a pilot region.*

To mitigate the effects of this data delay, an additional script was developed which can be triggered manually to retroactively calibrate data when the reference station data has become available. The same script can also be used to retroactively calibrate the first 34 days of sensor data where the online pipeline does not output any calibrated data. Furthermore, for sensors deployed in the Flanders region the API of the Belgian Interregional Environment Agency (IRCEL–CELINE) was used instead of DiscoMap to improve calibration performance.

#### 3.3.3.2 Practical Limitations

##### Delayed evaluation period

OnePlanet NitroSense was not the original $NO_2$ sensor planned to be included in the project and was included as a fall-back option when SODAQ $NO_2$ sensing devices did not pass initial testing (See COMPAIR D5.2 Closed Round Testing Report). This delayed the deployment of

NitroSense $NO_2$ devices and therefore collection of data to evaluate the calibration performance. Time constraints related to the implementation of policies in pilot experiments, resulted in prioritising the policy experiment and performing evaluation benchmarks post-experiment.

Limited evaluation dataset

In order to evaluate and validate the calibration, “ground truth” (i.e. the concentrations reported by a validated method at the measurement location) measurements are needed. This is only possible through deploying devices next to a high-end reference station and collecting measurements, ideally a minimum of 3 devices for more than one month; longer if the effect of different seasons is to be evaluated. In COMPAIR all pilot locations were encouraged to perform such deployments to enable evaluation of the calibration especially in geographical locations of varying climate conditions. However, after deliberation of pilot partners with city authorities many of the reference sensor deployments did not turn out to be possible. While the distant calibration algorithm was applied to all sensors explained in section 2, the performance of the calibration was only evaluated using the following devices and locations:

- 4 SODAQ AIR devices measuring PM: 3 deployed in Antwerp, Belgium and 1 in Berlin, Germany
- 5 OnePlanet NitroSense devices measuring $NO_2$: All deployed in Ghent, Belgium

Results of the evaluation are found in section 4.

#### 3.3.3.3 Additional Functionalities

In the calibration pipeline, Open Geospatial Consortium SensorThings API (OGC-STA) framework was used to standardise the structuring, storage and access of data. Applying this standardisation allowed minimising differences in ingestion of sensor data into the calibration pipeline and allowed scalable deployment of the pipeline for different types of sensors. Details and benefits of applying OGC-STA are found in COMPAIR D3.2: section 5.2 Sensor device functional and technical design report and examples in D3.6: Digital Twin CS Integration report.

### 3.3.2 Optimization of Distant Calibration (October 2024)

After the release of D3.5 v1.0, there was an additional round of optimization of parameters used in distant calibration algorithm to improve its performance. As a first step, a parameter optimization was performed by iterating through 17,600 different parameter configurations to find the best performing set of parameters. Table 3.1 shows a list of the optimised parameters.

| **Parameter** | **Range** |
|---|---|

| | |
|---|---|
| Search radius for reference (km) | 5 - 30 |
| Number of training days (d) | 13 - 44 |
| Variance type | Absolute, relative |
| Low variance reference (quartile) | 0.1 - 0.9 |
| Low variance sensor (quartile) | 0.1 - 0.9 |
| Model formula combinations | Required: reference, NO2 signal +<br>Optional: temperature, (absolute / relative) humidity, ozone |
| Curve fitting method | Ordinary least squares, total least squares |

***Table 3.1*** *Distant calibration parameters that were optimised and the optimization range.*

Since the results of the evaluation had to be applied to the situation of COMPAIR, where sensors would be deployed in Ghent, St. Niklaas and Athens where there are few to none reference stations nearby, a subset of the parameter configurations were used. This subset only included configurations where a search radius of 30 km was used, i.e. all reference stations within a radius of 30 km were considered for the calibration.

## 3.4 Auto-Calibration (October 2024)

In parallel to the distant calibration approach applied in COMPAIR, IMEC has been working on calibrating sensors using the calibration coefficients supplied by the sensor manufacturer (AlphaSense) without using data from the reference stations. Although the development is still in progress, it is possible to apply auto-calibration to the $NO_2$ data. The sensors remain susceptible to relative humidity, however this effect can be partially corrected using the humidity measurements from the sensor. So, the auto-calibration consists of two steps: 1) apply the manufacturer's calibration coefficients to convert the raw signal to a concentration, 2) estimate the relative humidity effect on the raw concentration and subtract it.

An additional benefit of requiring fewer third-party data sources (e.g. reference stations) for calibration is the potential for more consistent delivery of complete calibrated data compared to distant calibration. The calibration approach relies on interacting with environmental institutes' application programming interfaces (APIs) which could experience delays in providing data, resulting in less timely calibration.

# 4. Technical Implementation

Two calibration pipelines were developed, one to calibrate the particulate matter sensors for SODAQ AIR and one to calibrate the $NO_2$ sensors from OnePlanet Research Center.

Although the pipelines are distinct, the differences are subtle and the main components are the same (Figure 5.1). A calibration model is built per sensor, per device. This calibration model is built using a modified version of the distant calibration algorithm (Hofman et al., 2022). Each model is rebuilt every 7 days, on Monday morning 2 a.m. Every day in the following week at 4 a.m., the latest calibration model build is used to calibrate the raw sensor measurements in the past 24 hours. This 24h delay was implemented due to delays in the DiscoMap reference data availability. Both parts of the pipeline are run as *Jobs* on Databricks, where each calibration model is logged with mlflow.

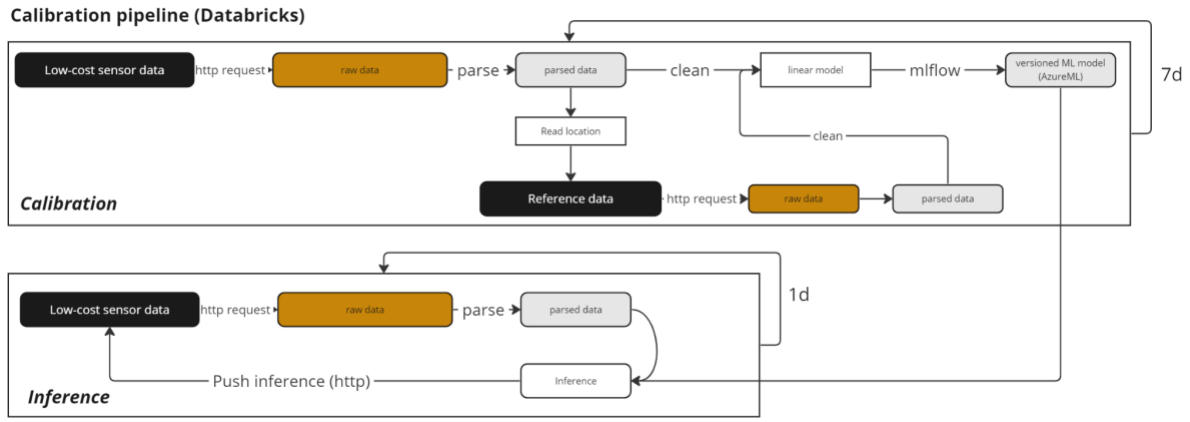


***Figure 4.1*** *Main architecture of the calibration pipeline set up in the cloud environment. In the Calibration part, a calibration model is built per sensor per device. In the Inference part, the latest calibration model is applied to sensor measurements of the last day.*

## 4.1 Pipeline PM

Raw data from the 230 $PM_{2.5}$ and $PM_{10}$ sensors are requested from the FROST server hosted by ATC. Additionally, temperature, relative humidity and the location (GPS coordinates) of the sensors are requested. Because the PM sensors can be mobile and potentially stored indoors overnight, it was decided to only include daytime hours for the calibration (from 7 a.m. to 1 a.m., local time). NB this method only provides a rough estimate of whether a device is outdoor, an estimate based on heuristics. A more direct metric for a sensor being indoors or outdoors is required to achieve a higher accuracy with the current calibration pipeline.

Raw data are resampled to hourly values and cleaned by discarding statistical outliers (observations > 3.5 standard deviation of a datastream). The GPS coordinates are cleaned differently, first all zeros are removed, and next the lowest and highest 2.5% are discarded to remove outliers.

From the GPS coordinates, the sensor location is set to the median of the coordinates. Reference stations within 50 km of this location are used for the calibration. Data from the reference stations are requested from the DiscoMap of the EEA.

For further processing, only time points are used for which both sensor and reference data are available. For every time point, the variance between reference stations is determined. The median of the variance across timepoints is computed, and only the time points lower or equal

to the median are used for the calibration, i.e. timepoints with the lowest variance between reference stations in the selected period.

From the remaining data, the time point-by-time point average of the reference data is used as the calibration target. The raw sensor output and temperature and humidity are used as predictors in a linear model:

$$PM_{ref} \sim a + b_1*PM_{sensor} + b_2*T + b_3*RH$$

$PM_{ref}$ are the average $PM_{10}$ or $PM_{2.5}$ measurements in $\mu g \cdot m^{-3}$ of nearby reference stations in the timepoints that met the inclusion criteria (described in 5.1), $a$ is the intercept, $PM_{sensor}$ are the raw PM sensor measurements in $\mu g \cdot m^{-3}$, $T$ are local temperature measurements, $RH$ are local relative humidity measurements. Thus, for every SODAQ device two calibration models are constructed, one for $PM_{10}$ and one for $PM_{2.5}$. This model is updated every week.

During the week, the most recent model is used to calibrate the data of the last day. These data are requested from the FROST server hosted by ATC, calibrated and then pushed back to the FROST server.

## 5.2 Pipeline $NO_2$

The $NO_2$ sensor calibration pipeline for OnePlanet NitroSense device is highly similar to the pipeline for PM. However, there are five differences. First, the reference $NO_2$ data are requested for reference stations within 30 km of the location of the device. Second, these reference data are requested from the IRCEL - CELINE for sensors deployed in Flanders. Third, because the $NO_2$ sensors were not in mobile but static devices, all hours were used for the calibration model, not just daytime hours. Fourth, raw data was requested from the database of OnePlanet Research Center. Fifth, an additional predictor was used: $O_3$ from nearby reference stations, akin to the distant calibration algorithm as described by Hofman et al. (2022) which is added to the pipeline that calibrates the 7 devices in Flanders and Athens:

$$NO_{2,\,ref} \sim a + b_1*NO_{2,\,sensor} + b_2*T + b_3*RH + b_4*O_{3,\,ref}$$

$NO_{2,\,ref}$ are the average $NO_2$ measurements in $\mu g \cdot m^{-3}$ of nearby reference stations in the timepoints that met the inclusion criteria (described in 5.1), $a$ is the intercept, $NO_{2,\,sensor}$ are the raw $NO_2$ sensor measurements in nA, $T$ are local temperature measurements, $RH$ are local relative humidity measurements, $O_3$ are average $O_3$ measurements in $\mu g \cdot m^{-3}$ from nearby reference stations (note that these are not necessarily the same stations as used for the $NO_2$ reference measurements).

The NitroSense devices in this study had two $NO_2$ sensors each. For every sensor a separate calibration model was constructed, updated and used. The week after the calibration models were updated, raw data was requested from the database of OnePlanet Research Center, calibrated and pushed to the FROST server of ATC.

# 5. Evaluation of Distant Calibration

The SODAQ AIR and NitroSense devices were collocated at national air quality monitoring stations to evaluate their performance. Measurements from these stations were regarded as the reference and were downloaded from the Air Quality database via the DiscoMap portal (https://discomap.eea.europa.eu/). The SODAQ AIR and NitroSense measurements were calibrated applying 1) factory calibration, 2) distant calibration and 3) field calibration. In the scope of this evaluation work, the field calibration takes the complete collocation period to estimate calibration coefficients and the performance is evaluated on the same period. The calibration outcome would indicate a performance cap of the device as if the same set of coefficients could be obtained by other calibration coefficients. On top of the calibration, these measurements were downsampled to hourly average in order to match the data frequency of the reference measurements. In this report each individual device is referred to by the last four digits of its unique ID (Table 4.1 and Table 4.2).

Five matrices are used to indicate the accuracy with respect to the reference measurements, which included: root mean square error (RMSE), mean absolute error (MAE), median absolute deviation (MAD), Pearson correlation coefficient (r) and expanded uncertainty (kU). [1] The expanded uncertainties on $PM_{10}$, $PM_{2.5}$ and $NO_2$ were determined at 50, 25 and 40 $\mu g \cdot m^{-3}$ respectively, and a coverage factor of 2 was used.

## 5.1 PM

Four SODAQ AIR devices were collocated with a reference station for the performance evaluation (Table 4.1). In this document these devices are referred to by the last four digits of their ID number. One of the devices (7420) was not operating properly (raw data availability was about 1%) and thus was excluded from the performance evaluation. The data coverage rates of the other three devices were around 93%, where two (8220 and 7833) were collocated at station BETR801 in Antwerp, Belgium and one (2725) at DEBE065 in Berlin, Germany.

**Table 5.1** Summary of the SODAQ AIR devices used for the performance evaluation.

| Device ID | Colocation station[1] | City | Data period[2] |
|---|---|---|---|
| 350457790907420 | BETR801 | Antwerp (BE) | 2023-12-04 to 2024-01-30 |
| 350457790908220 | BETR801 | Antwerp (BE) | 2023-12-04 to 2024-01-30 |
| 350457790917833 | BETR801 | Antwerp (BE) | 2023-12-04 to 2024-01-30 |
| 350916067032725 | DEBE065 | Berlin (DE) | 2023-09-07 to 2024-01-30 |

[1] Station EoI code as used in the past AirBase system.
[2] Up to the date till which the data were used for the evaluation in this document.

[1] The meaning of correlative statistics matrices (RMSE, MAE, MAD, r) are described in https://statisticsbyjim.com/regression/ , and kU is described in EU Directive EU 2008/50/EC.

The measurements from devices 8220 and 7833 were highly similar. The mean absolute differences were 0.61 and 0.51 $\mu g \cdot m^{-3}$ for $PM_{10}$ and $PM_{2.5}$, respectively, and the correlation coefficients were >0.998. In general, the raw SODAQ AIR measurements were able to roughly follow the same fluctuation patterns (Figure 4.1 and Figure 4.2). The correlation between raw $PM_{10}$ measurements and the reference was between 0.75 and 0.79, and the mean errors were at a magnitude of 3–11 $\mu g \cdot m^{-3}$. The lowest expanded uncertainty – which is calculated at 50 $\mu g \cdot m^{-3}$ – on the raw $PM_{10}$ measurements was 35.9%. The correlation between raw $PM_{2.5}$ measurements and the reference was between 0.91 and 0.94, and the mean errors were around 2–8 $\mu g \cdot m^{-3}$. The lowest expanded uncertainty – which is calculated at 25 $\mu g \cdot m^{-3}$ – on the raw $PM_{2.5}$ measurements was 58.4%. Filed calibration significantly improved the performance on both $PM_{10}$ and $PM_{2.5}$. The correlations with the reference was increased to 0.83–0.90 and 0.94–0.96 for $PM_{10}$ and $PM_{2.5}$, respectively, and the expanded uncertainties were reduced to 25–35% and 17–19%. Such improvements majorly came from the correction of the slope coefficient (Annex Figure 8.1). Detailed performance matrices are listed in Table 8.1 in the annex.

For both $PM_{10}$ and $PM_{2.5}$, the distant calibration approach introduced small reductions in RMSE and MAE in comparison to the raw measurements, but MAD increased slightly. There was no significant change in the correlation coefficient. Nevertheless, the expanded uncertainties of the $PM_{10}$ measurements after applying distant calibration become >100%, which was a substantial increase from the >35% of the raw measurements. The uncertainties of $PM_{2.5}$ increased slightly - compared to raw performance - in two of the devices with the distant calibration, but the other one had reduced uncertainty. However, across the range field calibration outperformed the COMPAIR-application of distant calibration. The reason behind this kind of performance change after applying the distant calibration was likely that the implemented calibration algorithm was biassed towards low concentration measurements. Consequently, the algorithm was effective in correcting the sensor baseline shifts, but the slope coefficients were poorly estimated due to a lack of high concentration reference points. This can be seen from the time series plots where the fluctuation at low concentrations were followed reasonably well by the distant calibrated values but not the peaks (Figure 4.1).

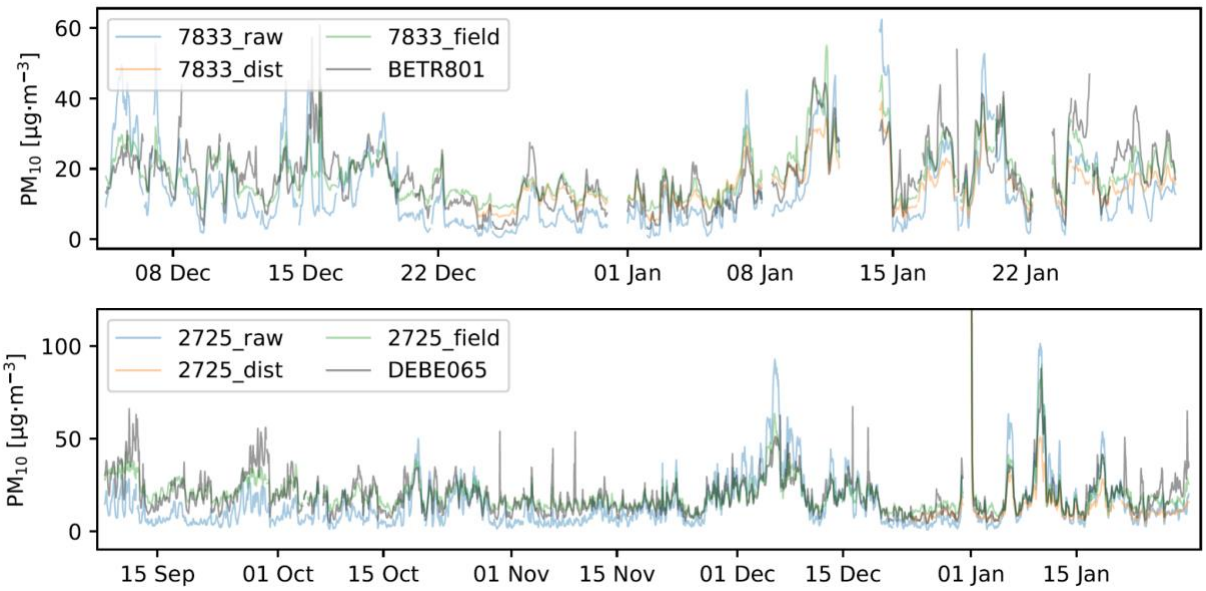


***Figure 5.1*** *Hourly averaged uncalibrated (raw), distant calibrated (dist) and field calibrated SODAQ AIR measurements on $PM_{10}$.*

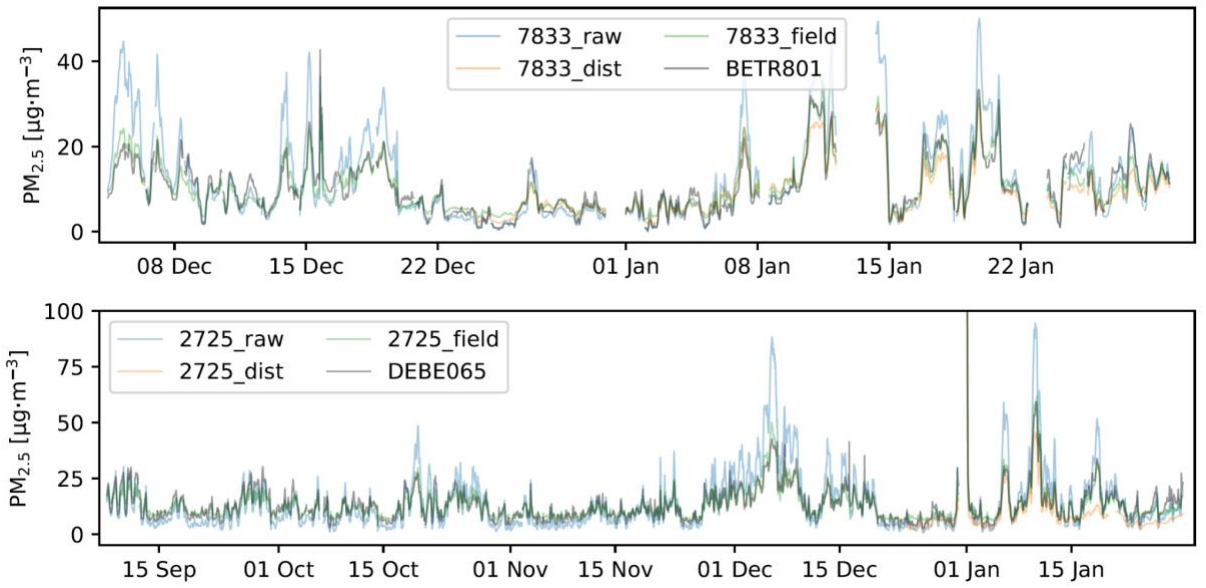


***Figure 5.2*** *Hourly averaged uncalibrated (raw), distant calibrated (dist) and field calibrated SODAQ AIR measurements on $PM_{2.5}$.*

## 5.2 $NO_2$

### 5.2.1 $NO_2$ Results (March 2024)

Five NitroSense sensor boxes were collocated at station BETR702 in Ghent (Belgium). Each sensor box consists of two $NO_2$ sensors of the same type, but these sensors were subjected to different configurations (level of analogue signal amplification, physical position in the circuitry, order of sensor readout, etc.) and thereby exhibit different sensitivities and noise levels. The test periods of these sensor boxes are given in Table 4.2.1. The raw data availability was around 97% for most devices, except the device 1319 had only 89.4% data due to unsuccessful update of an expired credential for IoT devices. Moreover, all field calibrated measurements had additional missing data between 9 - 18 October due to incomplete ozone measurements from nearby reference stations.

**Table 5.2** The performance evaluation period and location of the NitroSense sensor boxes deployed at station BETR702* in Ghent (Belgium).

| Device name | Period |
|---|---|
| op_D96A4FA650504335382E314AFF012E21 | 2023-09-12 to 2023-10-30 |
| op_461250D250504335382E314AFF061416 | 2023-09-12 to 2023-10-30 |
| op_C6D2F69450504335382E314AFF091319 | 2023-04-13 to 2023-06-28<br>2023-09-12 to 2023-10-30 |
| op_6C58DF7B50504335382E314AFF062612 | 2023-09-12 to 2023-10-30 |
| op_2E3AF1DF50504335382E314AFF061F0F | 2023-05-11 to 2023-06-28<br>2023-09-12 to 2023-10-30 |

* Station EoI code as used in the past AirBase system

We first discuss the field calibrated dataset as they yielded the best performance. The field calibrated NitroSense measurements on $NO_2$ followed the fluctuation pattern reported by the reference station at a reasonable accuracy (Figure 4.3). The correlation coefficients with respect to the reference measurements ranged between 0.71 – 0.92. The mean error of all sensors was at the level around 4 $\mu g \cdot m^{-3}$, and the between sensor disagreement (median absolute deviation from the group median) was 1.7±0.9 $\mu g \cdot m^{-3}$. The measurement in September and early October contained quite a few negative spikes which significantly affected the measurement accuracy especially at low concentrations. These spikes were likely caused by an innate behaviour of electrochemical sensors in response to temperature and humidity fluctuations (Farquhar et al., 2021). The expanded uncertainties of individual sensors

ranged between 24–87% after the field calibration. More details are provided in Table 8.3 in Annex.

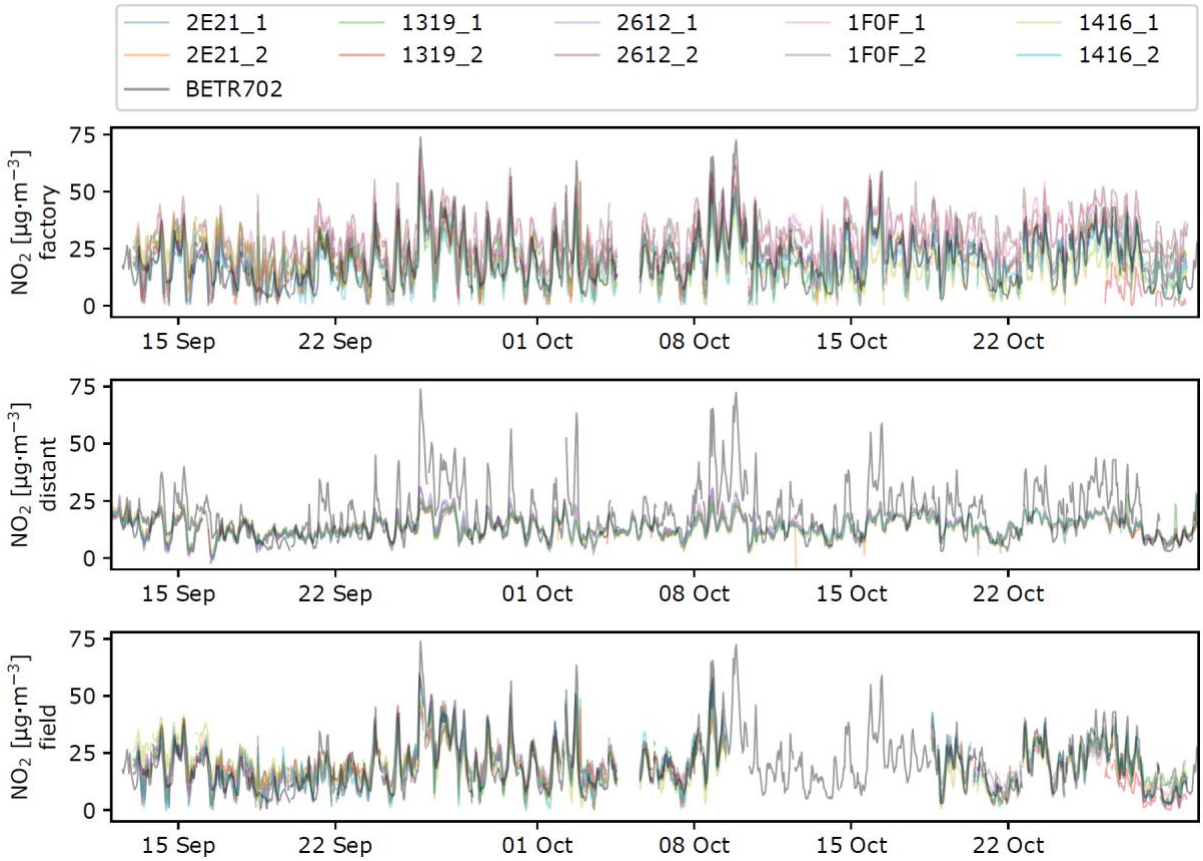


***Figure 5.3*** *Time series of $NO_2$ measurements taken by NitroSense devices during the colocation period with the reference station BETR702. Each NitroSense device consisted of two $NO_2$ sensors, which were denoted by the suffix.*

The overall performance of factory calibration was slightly lower than field calibration (Annex Table 8.3). The correlation coefficients with respect to the reference measurements ranged between 0.64–0.84. The expanded uncertainties were higher than field calibrated values by roughly 10%. The between-sensor disagreement was 4.0±5.1 µg·m$^{-3}$. This was higher with the factory calibration and could be seen from the wider spread in the measurement time series (Figure 4.3). The poorer reproducibility could be resulted from the limitation of the factory calibration coefficients, where the effects of ambient condition and sensor ageing were not taken into account.

The distant calibration yielded a between-sensor disagreement of 0.5±0.35 µg·m$^{-3}$, which was notably lower than the other calibration approaches. The accuracy of the distant calibration was nevertheless not ideal (Annex Table 8.3). While the measured $NO_2$ roughly matched the fluctuation pattern at concentrations below 25 µg·m$^{-3}$, peaks at higher levels were evidently underestimated (Figure 4.3). Consequently, the expanded uncertainty of these distant calibrated measurements – which is calculated at 40 µg·m$^{-3}$ – was >200%. This outcome was discussed in depth between COMPAIR experts at OnePlanet and VMM, key hypotheses and conclusions have been incorporated in Chapter 6.

### 5.2.1 $NO_2$ Results (October 2024)

Parameter search for the distant calibration algorithm was evaluated using 2 OnePlanet NitroSense devices that were placed alongside (“co-located”) with a reference station as a part of European project LIFE Critical[2]. These devices have been co-located at the reference station in Rotterdam Statenweg (NL01493) for over one year, so this data was selected to give better insights into the long term behaviour of the algorithm. Specifically, data between 19 December 2022 and 2 February 2024 was used. Reference data and calibrated NitroSense data are shown in Figure 3.1.

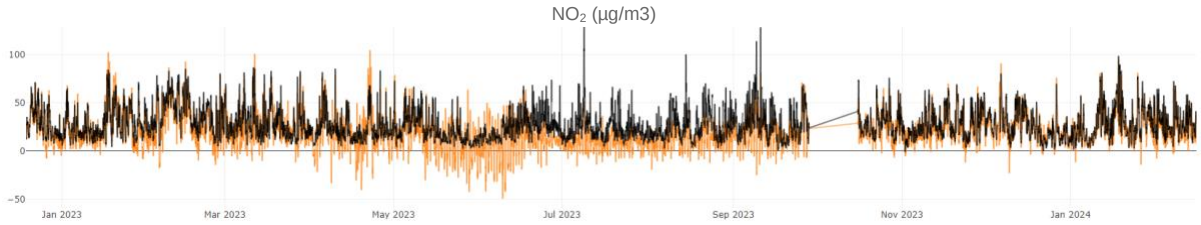


***Figure 5.4*** *Reference data (black) and NitroSense data calibrated using optimised distant calibration algorithm (orange) of one device deployed for project LIFE Critical at Rotterdam Statenweg.*

Although the optimised parameter configuration performs better than the March 2024 results, the $NO_2$ calibration performance was still not satisfactory, especially in the month of June 2023 when the temperature is high and humidity is low (see Figure 3.1).

[2] https://lifecritical.eu/en/

# 6. Evaluation of Auto-Calibration (October 2024)

## 6.1 $NO_2$

### 6.1.1 Preliminary Evaluation

Figure 6.1 compares the auto-calibrated NitroSense $NO_2$ output with the reference station data, using the same dataset and evaluation period as section 3.1. Results are in Figure 3.2.

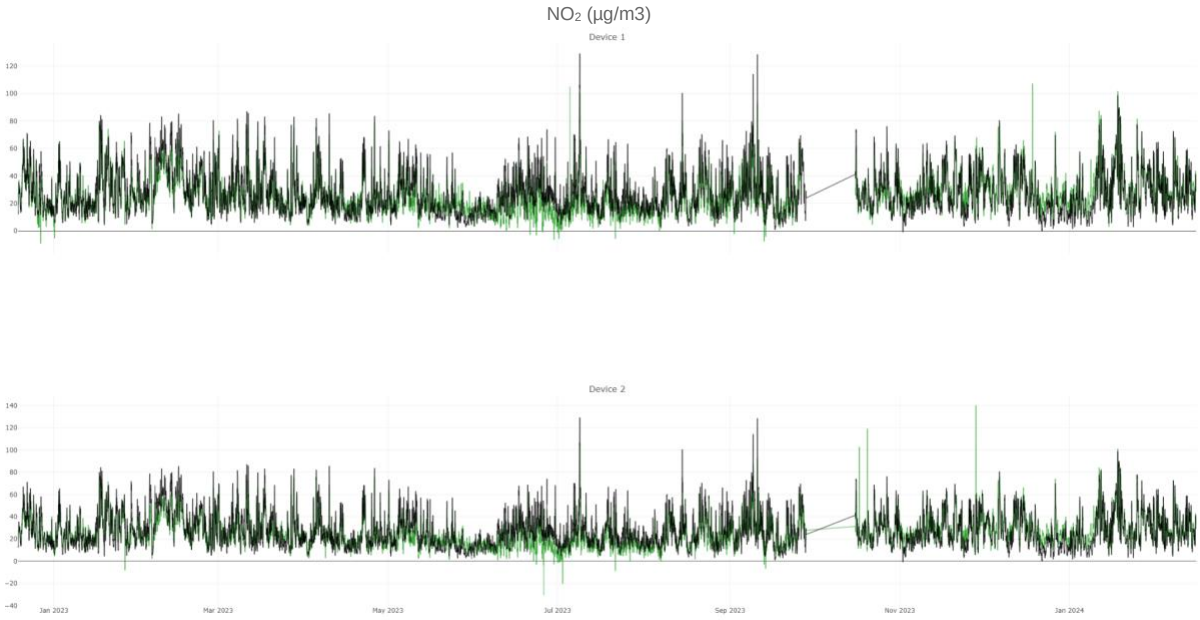


***Figure 6.1*** *Reference data (black) and NitroSense data calibrated using optimised distant calibration algorithm (green) of two devices deployed for project LIFE Critical at Rotterdam Statenweg.*

It is clear that auto-calibration outperforms parameter-optimised distant calibration. To compare the results, three metrics per method is compared to the reference $NO_2$ data. The metrics are computed over the full dataset.

| *Device 1* | Pearson's r | $R^2$ | MAE |
|---|---|---|---|
| *raw* | 0.757 | 0.382 | 8.529 |
| *distant calibration* | 0.797 | 0.366 | 9.210 |
| *auto-calibration* | 0.792 | 0.567 | 7.593 |

| *Device 2* | Pearson's r | $R^2$ | MAE |
|---|---|---|---|
| *raw* | 0.745 | 0.366 | 8.556 |
| *distant calibration* | 0.792 | 0.329 | 9.366 |
| *auto-calibration* | 0.876 | 0.753 | 5.745 |

***Table 6.1*** *Correlative statistics of the calibrated data using different calibration approaches: Pearson correlation coefficient (r), coefficient of determination ($R^2$) and mean absolute error (MAE).*

Table 6.1 shows that even though the correlation with the reference of all three methods (raw, distance cal, auto-cal) is acceptable, the error is substantially different. This is reflected in the differences in $R^2$ and MAE between the different methods. From the correlation, $R^2$ and MAE combined it can be concluded that there must be an offset or scaling issue for the raw and distant calibrated values. These results encourage a transition from distant calibration approach in COMPAIR to auto-calibration. The next section includes a similar analysis using a more expansive number of metrics to evaluate auto-calibration results using COMPAIR data.

### 6.1.2 Evaluation using COMPAIR data

The Flemish environmental agency (VMM) also performed a first evaluation of the data calibrated using the distant calibration approach which was described in section 3.3. This evaluation is found in D5.6 Public Round report section 3.3.

This evaluation was repeated using auto-calibrated NitroSense data, collected from the 5 sensors co-located with the VMM reference station Gustaaf Callierlaan in Ghent (BETR702) from mid-September 2023 until the end of October 2023. It is important to note that these sensors were shipped in February 2023 and first deployed in April 2023, and the comparison takes place in September which may lead to somewhat lower performance than performing the same test with freshly deployed sensors.

The evaluation results are shown in the following figures, courtesy of VMM. Figure 3.3 shows that the measurement range of the 5 co-located NitroSense devices are similar, as expected.

$NO_2$ (µg/m3)

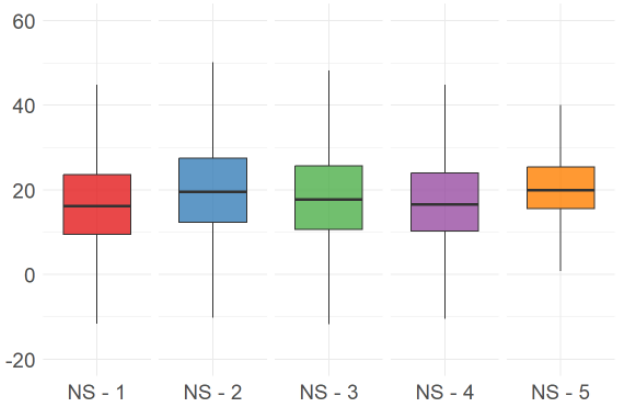


***Figure 6.2*** *The box plots show that the 5 NitroSense sensors co-located at Ghent reference station, represented by the different colors are measuring in a similar range.*

Table 6.2 shows the descriptive statistics of the calibrated device data.

| | $NO_2$ (µg/m3) | | | | |
|---|---|---|---|---|---|
| **ID** | **NS - 1** | **NS - 2** | **NS - 3** | **NS - 4** | **NS - 5** |
| **Mean** | 17.016 | 20.160 | 18.509 | 17.375 | 21.002 |
| **Median** | 16.190 | 19.628 | 17.783 | 16.544 | 20.018 |
| **SD** | 11.399 | 11.806 | 11.525 | 10.733 | 8.191 |
| **Max** | 74.298 | 69.528 | 73.882 | 66.441 | 82.716 |
| **Min** | -57.522 | -69.546 | -62.631 | -73.072 | -27.980 |

**Table 6.2** Descriptive statistics of the device data: mean, median, SD (standard deviation), min (minimum), max(maximum) during the period. Unit is $\mu g/m^3$.

The time series data shows that auto-calibrated NitroSense NO2 data follows the trend of the reference monitor, except for NS-5 in the period until mid September.

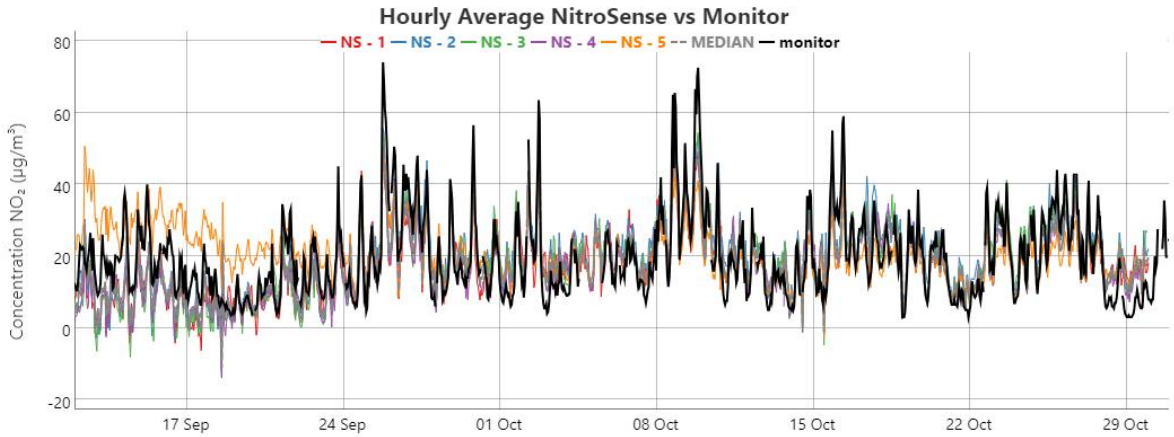

***Figure 6.3*** *Time series comparison of the auto-calibrated data from 5 devices and the data from the reference monitor.*

In Figure 6.4, each sensor is plotted against the reference data, the solid line representing a 1:1 and dashed line representing the orthogonal regression.

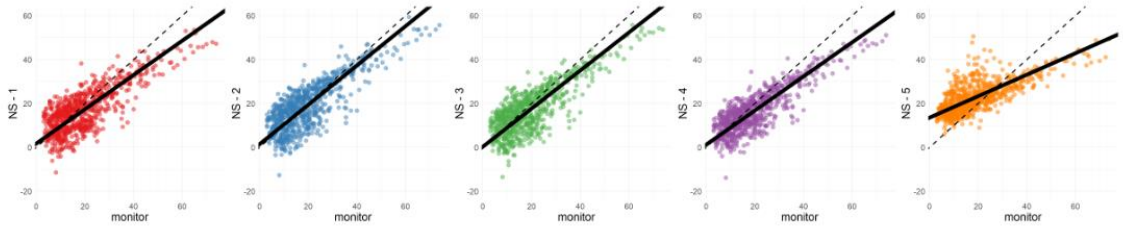


***Figure 6.4*** *Scatter plots where NitroSense data (y-axis) is plotted against the reference monitor data (x-axis).*

Figure 6.4 shows that in general, lower concentrations tend to be somewhat overestimated and high concentrations underestimated. $R^2$ coefficients of determination are calculated for the sensors, using the following equation:

$$R^2_{\mathcal{M}\mathcal{S}_i} = (\rho_{\mathcal{M}\mathcal{S}_i})^2 = \left(\frac{\mathrm{cov}(\mathcal{M}, \mathcal{S}_i)}{\sigma_{\mathcal{S}_i}\sigma_{\mathcal{M}}}\right)^2$$

In Figure 6.6 , $R^2$ coefficients of determination are represented for each sensor.

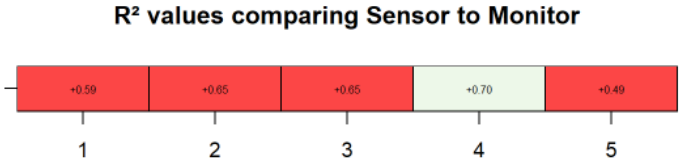


***Figure 6.5*** *$R^2$ values comparing sensor to monitor.*

Figure 6.5 shows moderate $R^2$ values, through NS-1 and NS-4, where only NS-4 passes the acceptability threshold used by VMM (0.70) when tendering for devices for $NO_2$ measurements. One sensor (NS-5 performs) worse than the average (NS-5). This is the same device which had an initial period of drift in Figure 3.5.

### 6.1.3 Comparison of distant calibrated and auto-calibrated data

In the public round report, the distribution plots using distant-calibrated data demonstrate that higher $NO_2$ concentrations measured by the reference station are significantly underestimated

by the low-cost sensor, see Figure 6.7. To compare the two approaches, the same distribution is plotted using auto-calibrated data (Figure 6.8).

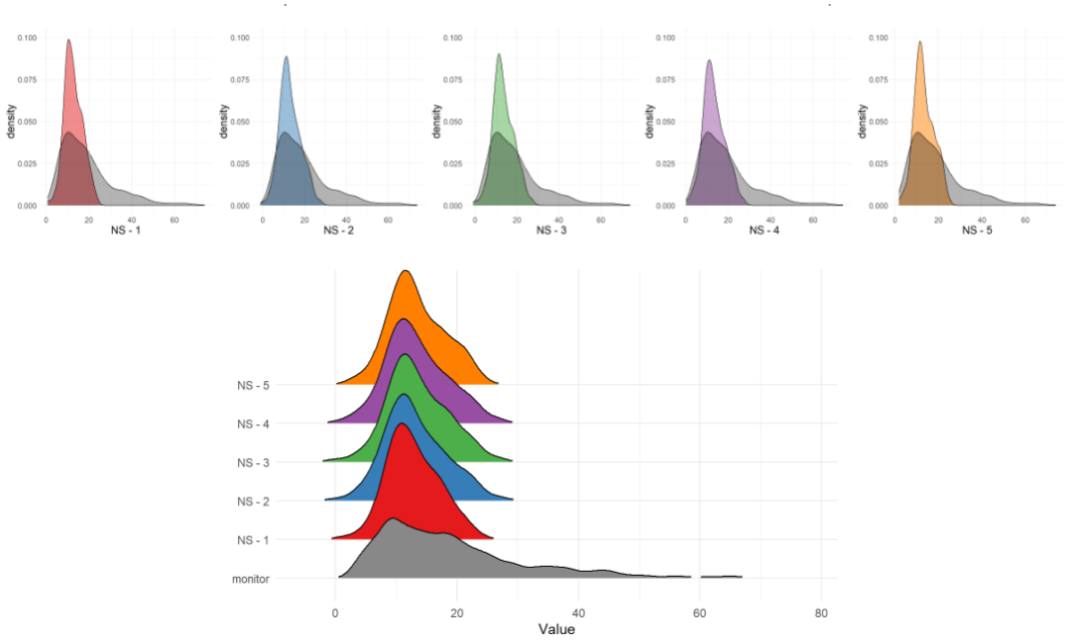


***Figure 6.6*** *Distribution of distant calibrated $NO_2$ concentrations calibrated using the approach described in section 3.3 (source: VMM, D5.6 Public Round Report).*

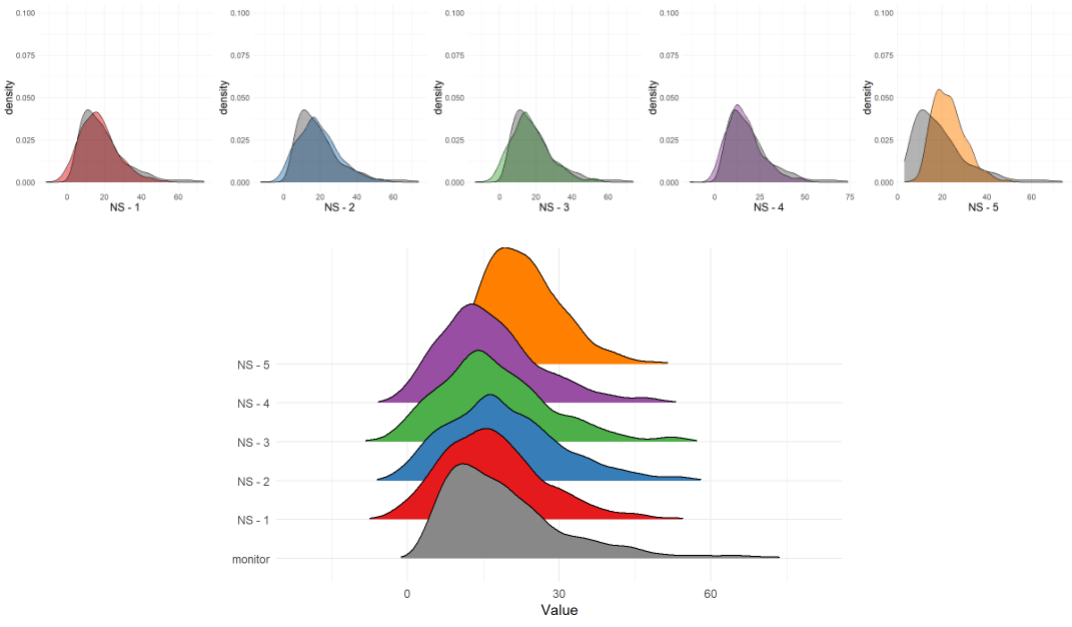


***Figure 6.7*** *Distribution of auto-calibrated $NO_2$ concentrations calibrated using the auto-calibration approach described in section 2.2.*

When Figure 6.6 and 6.7 are compared, it is visible that this underestimation of high concentrations is significantly improved by using auto-calibration. One particular sensor NS-5 (orange) remains an outlier and overestimates the concentrations on average.

# 7. Conclusions and Discussion

In this deliverable we describe the application of a cloud-based calibration approach to improve the air quality sensor output in the COMPAIR project. We evaluate the performance of the different calibration methods by comparing the calibrated data with the results reported by highly accurate reference stations.

After describing the working principles of the low-cost sensors and why they can benefit from calibration, we first describe the novel approaches: (1) distant calibration and (2) auto-calibration along with the standard approaches. Auto-calibration description and results are added to this deliverable in October 2024 when v1.1 version is published.

In section 4, we explain that a new, scalable pipeline was developed to calibrate the sensor data that can cover the various pilot regions. This pipeline supports the cloud-based calibration of 265 (250 SODAQ AIR, 15 OnePlanet NitroSense) devices in the project.

In sections 5 and 6, we evaluate the approaches by comparing the raw sensor data (if available) with differently calibrated data: field calibration where the calibration algorithm is trained based on a period where the LCS and reference station is co-located; plus distant and auto-calibration where co-location is not needed for calibration. The desired result is that the distant and auto-calibration performs somewhere between the raw data and field calibration, preferably closer to field calibration. The results show that on average for $PM_{2.5}$ and $PM_{10}$, the distant calibration introduces moderate reductions in error compared to raw measurements, while correlation remains similar and we observe varying performance in terms of reduction of the measurement uncertainty.

In the case of $NO_2$, distant calibration performance shows room for improvement, even after further optimization of parameters. The analysis is repeated for auto-calibrated data, and while the metrics shows that the results are close to the threshold of acceptability for VMM, auto-calibration performs significantly better than distant calibration. After these findings, the distant calibration algorithm is phased out for $NO_2$ measurements in the COMPAIR calibration pipeline and the data is retroactively calibrated using auto-calibration. The auto-calibrated data is used to analyse the Ghent and St Niklaas experiments in public round use case 6 and 7 (see page 92-06 of COMPAIR D5.6 Public Round Report). This improvement enables a more accurate estimation of how the introduced circulation changes affect air quality.

# 8. References

## Books/Articles

Farquhar, A.K., Henshaw, G.S. & Williams, D.E. (2021), Understanding and correcting unwanted influences on the signal from electrochemical gas sensors. ACS Sensors, 6(3), 1295-1304. https://doi.org/10.1021/acssensors.0c02589

Hofman, J., Nikolaou, M., Shantharam, S.P., Stroobants, C., Weijs, S., Panzica La Manna, V. (2022). Distant calibration of low-cost PM and NO2 sensors; evidence from multiple sensor testbeds. *Atmospheric Pollution Research*, 16. https://doi.org/10.1016/j.apr.2021.101246

Molnár, A., Imre, K., Ferenczi, Z., Kiss, G., & Gelencsér, A. (2020). Aerosol hygroscopicity: Hygroscopic growth proxy based on visibility for low-cost PM monitoring. *Atmospheric Research*, *236*, 104815. https://doi.org/10.1016/j.atmosres.2019.104815

## Websites

Dejan. (2020, December 28). Diy air quality monitor—Pm2. 5, co2, voc, ozone, temp & hum arduino meter. *How To Mechatronics*. Retrieved January 3, 2024, from https://howtomechatronics.com/projects/diy-air-quality-monitor-pm2-5-co2-voc-ozone-temp-hum-arduino-meter/

*Gas sensors types and mechanism*. (n.d.). Tutorials | CO2 Sensors | Products | Asahi Kasei Microdevices (AKM). Retrieved January 3, 2024, from https://www.akm.com/content/www/akm/eu/en/products/co2-sensor/tutorial/types-mechanism.html

Statistics by Jim: Making statistics intuitive. Retrieved February 16, 2024, from https://statisticsbyjim.com/regression

OGC SensorThings API. Retrieved February 16, 2024, from https://www.ogc.org/standard/sensorthings/

# 9. Supplementary Information

**Table 9.1** Comparisons between the original and field calibrated SODAQ AIR measurements on $PM_{10}$ and $PM_{2.5}$ throughout the complete colocation period at the reference station.

| Device | Calibration | RMSE [µg·m$^{-3}$] | MAE [µg·m$^{-3}$] | MAD [µg·m$^{-3}$] | r | kU [%] |
|---|---|---|---|---|---|---|
| *$PM_{10}$* | | | | | | |
| 350457790908220 | Raw | 8.5 | 7.0 | 3.0 | 0.75 | 39.7 |
| | Field | 5.0 | 3.8 | 3.2 | 0.83 | 35.0 |
| 350457790917833 | Raw | 8.6 | 7.2 | 3.1 | 0.75 | 35.9 |
| | Field | 4.8 | 3.6 | 2.9 | 0.85 | 32.2 |
| 350916067032725 | Raw | 11.1 | 8.3 | 3.7 | 0.79 | 50.0 |
| | Field | 5.2 | 3.5 | 2.3 | 0.90 | 24.6 |
| *$PM_{2.5}$* | | | | | | |
| 350457790908220 | Raw | 5.5 | 3.4 | 1.7 | 0.91 | 63.0 |
| | Field | 2.3 | 1.8 | 1.5 | 0.94 | 18.2 |
| 350457790917833 | Raw | 5.1 | 3.2 | 1.6 | 0.91 | 58.4 |
| | Field | 2.2 | 1.7 | 1.4 | 0.95 | 16.5 |
| 350916067032725 | Raw | 7.9 | 4.5 | 1.8 | 0.94 | 64.0 |
| | Field | 2.5 | 1.8 | 1.3 | 0.96 | 18.7 |

**Table 9.2** Comparisons among the original, distant calibrated and field calibrated SODAQ AIR measurements on $PM_{10}$ and $PM_{2.5}$ over the period during which distant calibration was available.

| Device | Calibration | RMSE [µg·m$^{-3}$] | MAE [µg·m$^{-3}$] | MAD [µg·m$^{-3}$] | r | kU [%] |
|---|---|---|---|---|---|---|
| *$PM_{10}$* | | | | | | |
| 350457790908220 | Raw | 8.0 | 6.5 | 2.6 | 0.82 | 29.3 |
| | Distant | 6.6 | 4.9 | 3.9 | 0.82 | 106.6 |
| | Field | 4.8 | 3.6 | 3.0 | 0.88 | 25.2 |
| 350457790917833 | Raw | 8.1 | 6.7 | 2.7 | 0.82 | 27.1 |
| | Distant | 6.5 | 4.8 | 3.8 | 0.84 | 117.4 |
| | Field | 4.7 | 3.5 | 2.9 | 0.89 | 25.0 |
| 350916067032725 | Raw | 10.2 | 6.9 | 2.4 | 0.92 | 46.1 |
| | Distant | 10.4 | 6.5 | 2.9 | 0.93 | 129.0 |
| | Field | 5.3 | 3.5 | 2.4 | 0.95 | 21.4 |
| *$PM_{2.5}$* | | | | | | |
| 350457790908220 | Raw | 4.7 | 2.8 | 1.5 | 0.94 | 52.2 |
| | Distant | 3.2 | 2.4 | 1.8 | 0.93 | 49.0 |
| | Field | 2.2 | 1.6 | 1.3 | 0.96 | 16.4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 350457790917833 | Raw | 4.1 | 2.5 | 1.4 | 0.94 | 45.8 |
| | Distant | 3.1 | 2.3 | 1.7 | 0.94 | 50.8 |
| | Field | 2.1 | 1.6 | 1.3 | 0.96 | 15.6 |
| 350916067032725 | Raw | 10.6 | 4.9 | 1.8 | 0.97 | 64.2 |
| | Distant | 6.5 | 4.8 | 2.1 | 0.95 | 77.0 |
| | Field | 2.6 | 1.8 | 1.2 | 0.98 | 19.9 |

**Table 9.3** Comparisons among the distant, factory and field calibrated NitroSense measurements on $NO_2$ over the colocation period in Ghent, Belgium.

| Indicator | Estimate[1] | Distant | Factory | Field |
|---|---|---|---|---|
| RMSE [µg·m$^{-3}$] | mean | 10.03 | 8.49 | 5.80 |
| | std. | 0.65 | 1.99 | 0.95 |
| | min | 9.13 | 3.38 | 4.47 |
| | median | 10.02 | 7.94 | 5.74 |
| | max | 10.80 | 12.92 | 7.01 |
| MAE [µg·m$^{-3}$] | mean | 7.05 | 6.92 | 4.13 |
| | std. | 0.48 | 1.99 | 0.76 |
| | min | 6.40 | 5.13 | 3.46 |
| | median | 6.93 | 6.24 | 4.48 |

| | | | | |
|---|---|---|---|---|
| | max | 7.71 | 11.52 | 5.56 |
| MAD [µg·m$^{-3}$] | mean | 4.72 | 4.58 | 3.59 |
| | std. | 0.57 | 0.46 | 0.61 |
| | min | 4.03 | 3.85 | 2.74 |
| | median | 4.54 | 4.40 | 3.58 |
| | max | 5.47 | 5.17 | 4.60 |
| r | mean | 0.721 | 0.743 | 0.833 |
| | std. | 0.069 | 0.075 | 0.077 |
| | min | 0.631 | 0.643 | 0.707 |
| | median | 0.710 | 0.746 | 0.842 |
| | max | 0.837 | 0.837 | 0.920 |
| kU [%] | mean | 370.0 | 48.7 | 45.5 |
| | std. | 108.9 | 14.8 | 22.3 |
| | min | 214.0 | 30.8 | 23.5 |
| | median | 366.9 | 41.8 | 38.5 |
| | max | 592.3 | 70.5 | 86.6 |

[1] Estimated from the performance indicator of individual $NO_2$ sensors.

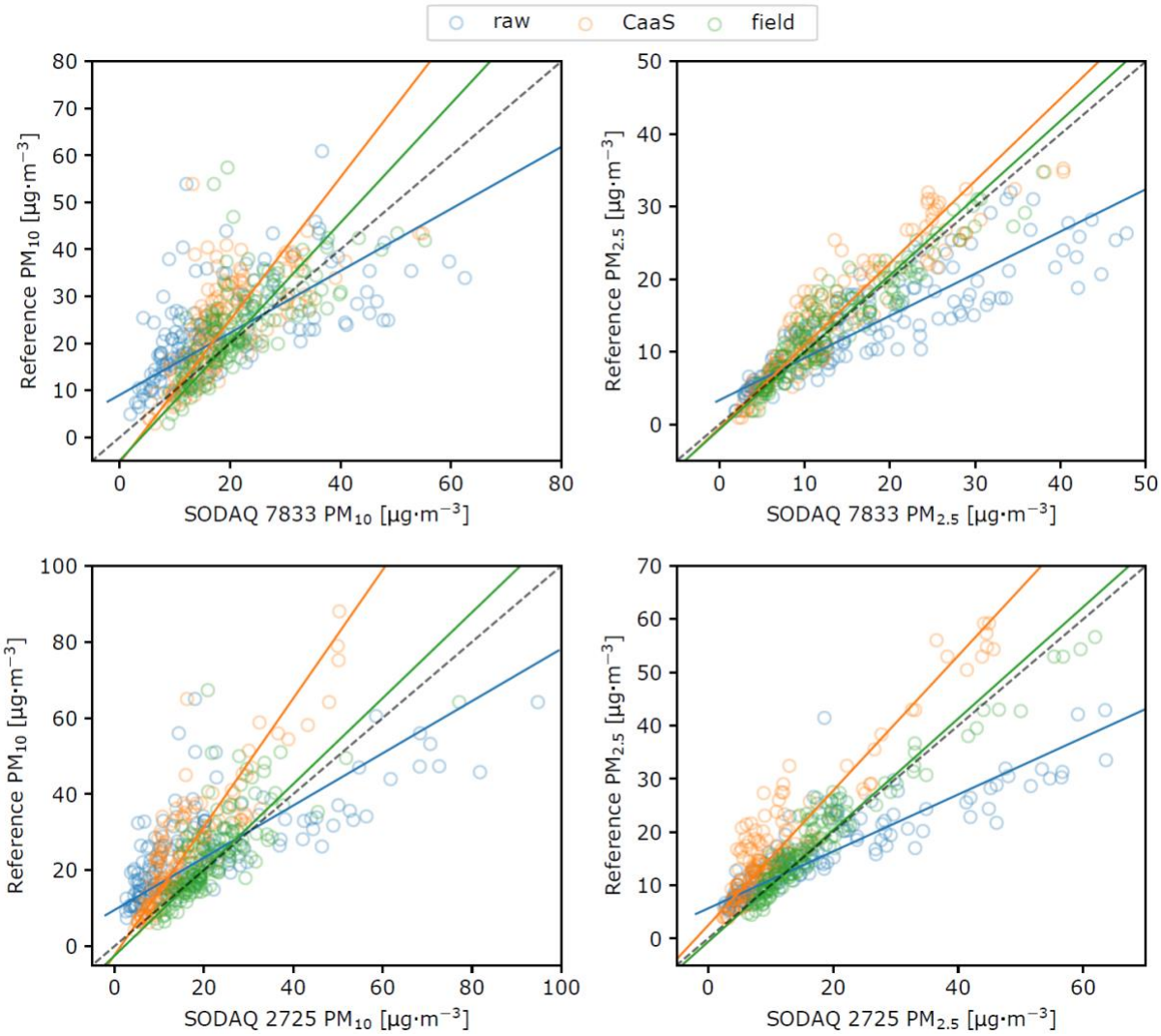


***Figure 9.1*** *Scatter plots of SODAQ AIR device deployed at* Antwerp *(ID ***7833) and Berlin (ID ***2725) against the colocated reference station on $PM_{10}$ and $PM_{2.5}$ measurements. Solid lines denote the regression line of total least squares, and dashed lines the perfect match.*