WG6 Sensors Benchmarking

Calibration of low-cost PM2.5 sensors.

Plenary Meeting February 2024

WG6 Timeline activities

- **2022:** Discussions about sensor calibrations. Brainstorming on sensible approach for calibration benchmark
  → Benchmark based on synthetic sensor data.

- **2022/2023:** INERIS, ISSeP and RIVM use the generated synthetic data to develop/test their **selection** and **calibration** methods.

- **Jan-July 2023:** work on article comparing the results on the selection/calibration benchmark. Submitted for publication in summer.

- **Last half year:**
  - Review sensor calibration
    - Long review process during second half of 2024.
    - Minor revisions.
    - Processing well-reasoned substantive commentary.
  - Discussions and thinking out approach for benchmarking sensor fusion methods (next presentation).

- **End 2023:** Publication on benchmarking sensor calibrations

Publication

- Using synthetic data to benchmark correction methods for low-cost air quality sensor networks
- Air Quality, Atmosphere & Health
- 21 December 2023
- RIVM, INERIS, VITO, ISSeP, CESAM & UAveiro, JRC
- Benchmark for low-cost **PM2.5** sensors

Air Quality, Atmosphere & Health
https://doi.org/10.1007/s11869-023-01493-z

Check for updates

Using synthetic data to benchmark correction methods for low-cost air quality sensor networks

Joost Wesseling[1] · Derko Drukker[1] · Alicia Gressent[2] · Stijn Janssen[3] · Pascal Joassin[4] · Fabian Lenartz[4] · Sjoerd van Ratingen[1] · Vera Rodrigues[5] · Jorge Sousa[3] · Philippe Thunis[6]

Received: 22 July 2023 / Accepted: 21 December 2023

Abstract
A benchmark was performed, comparing the results of three different methodologies proposed by three institutions to calibrate a network of low-cost PM2.5 sensors, on an hourly basis, using synthetically generated real concentrations and sensor measurements. The objective of the network calibrations was to correct the 2000+ sensor measurements in the Netherlands for the sensitivity to (local) environmental conditions. The option to use real measurements was dropped because the number of low-cost sensors sufficiently close to the 40 reference measurement locations was assessed to be spatially insufficient to benchmark the proposed approaches. Instead, synthetic real concentrations were generated to enable validation at all sensor locations. Hourly actual sensor and actual fixed concentrations, as well as interpolated concentration maps, were used as underlying data to generate the synthetic data sets for the period of 1 month. The synthetic sensor measurement errors were constructed by sampling from a collection of differences between actual sensor values and actual measurements. Of the three tested calibration methods, two follow a similar approach, although having differences in, e.g., outlier analyses and method of grouping sensors, leading also to comparable corrections to the raw sensor measurements. A third method uses significantly stricter rules in outlier selection, discarding considerably more sensors because of insufficient quality. Differences between the methods become most apparent when analyzing data at a smaller time scale. It is shown that two network calibration methods are better at correcting the hourly/daily bias.

Keywords Low-cost sensors · Air quality · Benchmarking · Calibration

Introduction

Air pollution is recognized as the world's biggest environmental health risk (Lelieveld et al. 2020; WHO 2021). In Europe, many cities are affected by poor air quality levels and ambient concentration levels regularly exceed both the European standards prescribed by the Ambient Air Quality Directive (AAQD) and the guidelines recommended by the World Health Organization (WHO) (EEA 2022). For fine inhalable particles (PM2.5), the EU limit value is generally met (EEA 2022), but only a few cities manage to keep concentrations below the levels recommended by the WHO (González Ortiz et al. 2020; EEA 2022; Thunis et al. 2018; Rodrigues et al. 2021). As a reference, the AAQD establishes an annual limit value of 25 µg/m^3 for PM2.5 concentrations, while the WHO sets an annual guideline of 5 µg/m^3. According to the latest report released by the European Environment Agency (EEA) on air quality in Europe, the WHO annual guideline level for PM2.5 was exceeded by all reporting countries, except Iceland, in 2022 (EEA 2023). According to the latest estimates, at least 238,000 premature deaths were associated with the exposure to PM2.5 (EEA 2023).

To reduce the impacts of air pollution, particularly in cities where most of the population lives (e.g., more than two-thirds of European population lived in an urban area in

✉ Joost Wesseling
joost.wesseling@rivm.nl

[1] National Institute for Public Health and the Environment (RIVM), Utrecht, The Netherlands
[2] INERIS, Verneuil-en-Halatte, France
[3] VITO, Mol, Belgium
[4] Public Service Scientific Institute (ISSeP), Liège, Belgium
[5] Centre for Environmental and Marine Studies (CESAM) & Department of Environment and Planning, University of Aveiro, 3810-193 Aveiro, Portugal
[6] Joint Research Centre (JRC), Ispra, Italy

Published online: 16 January 2024

Springer

Three different calibration methods

INERIS

- First define clusters of sensor observations based on:
  - distance between sensors
  - Typology / land use
  - season
- Estimate local correction factor for clusters close to reference measurement
- Interpolation by kriging

ISSeP

- Measurements from reference stations are used to produce interpolated [PM_{xx}] fields for the studied area.
- Determine sensor weights using regression with the field above
- Use sensor weights to update the field
- Iteration: Repeat the steps above at least twice
- Updated field at sensor location = calibrated sensor

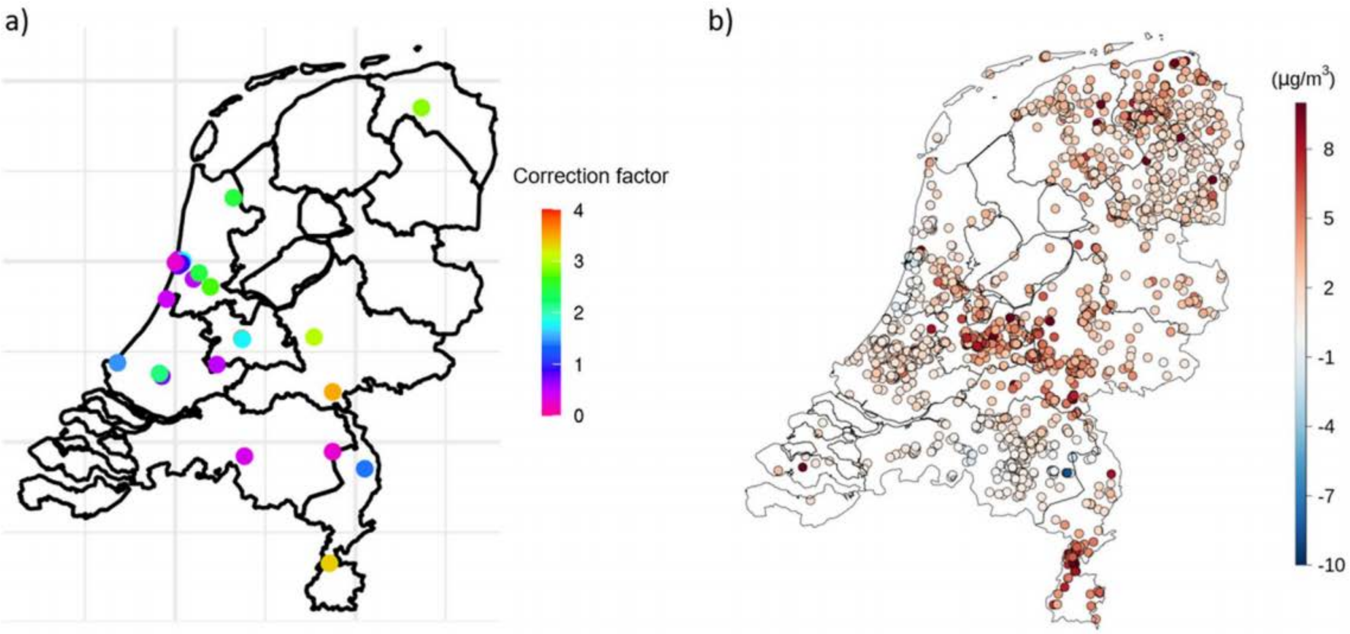
National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport

- Outlier detection methodology based on lowest/highest sensors.
- Define sensor groups in the vicinity of the reference stations
- Estimate local correction factor for these groups.
- Interpolation correction factors to correction field.

Example calibration methods

Correction factors

Correction of sensor concentration (µg/m³)

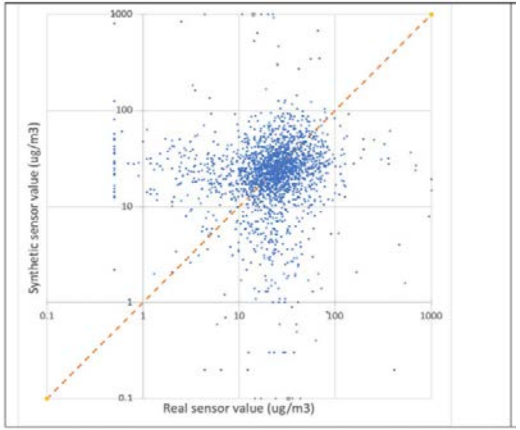


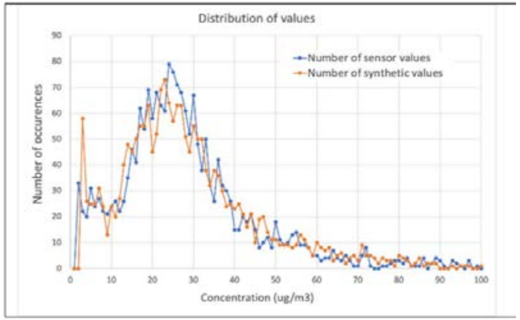
Validation approach

- How to compare the three difference calibration methods?
- We wanted to know the behaviour of the calibration models at all ~ 2000 sensor locations (Not only at a few locations, close to reference measurement)
- Decision to **generate synthetic data** for **validation**.
  - Compare results from different calibration methods to synthetic truth at ~ 2000 locations
  - Objectively test the effects of variations in calibration strategies
  - Generate **synthetic truth** (real concentrations) as well as **synthetic sensor data**
  - Synthetic sensor data should reflect chaotic aspects of low-cost PM sensors. Use real sensor data as bases for synthetic data
  - Synthetic sensor data enables varying sensor uncertainties
  - A **data set** with **synthetic data** was created for **January, 2022**, using 50% of the random uncertainty.

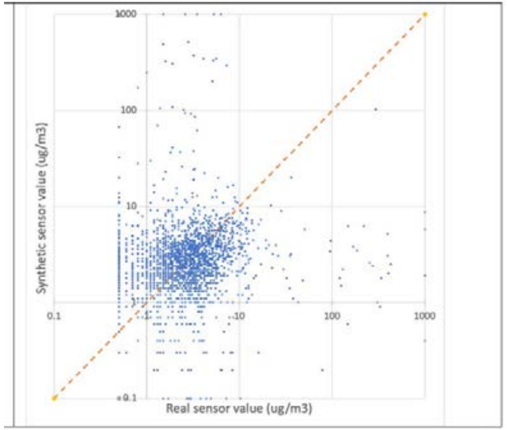
Synthetic vs real sensor data

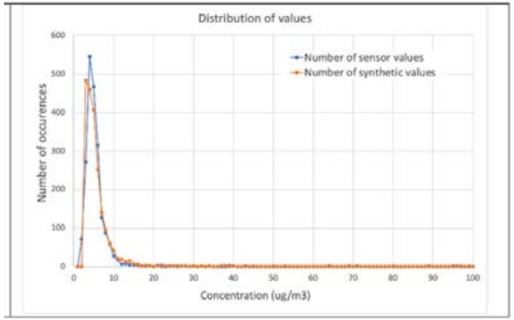
January 11th, 2022, 06:00





February 08th, 2022, 20:00

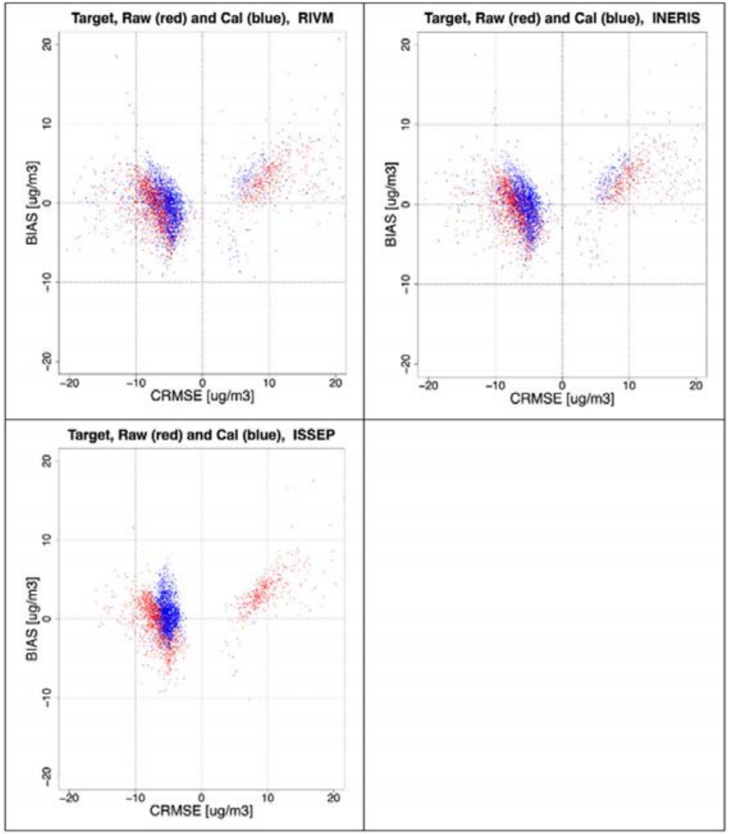




- No correlation when compared real synthetic to real concentrations

- Same distribution for synthetic and real data

Results (1)



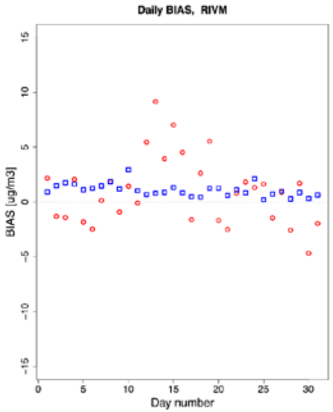
Period Jan 01-31, 2022

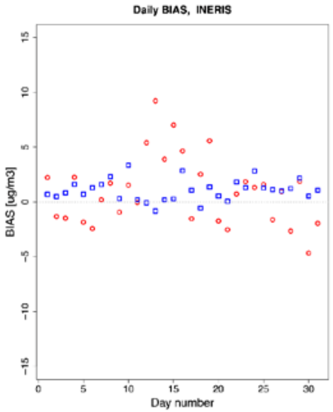
Target plots of the monthly averaged raw data and the averaged calibrated data.

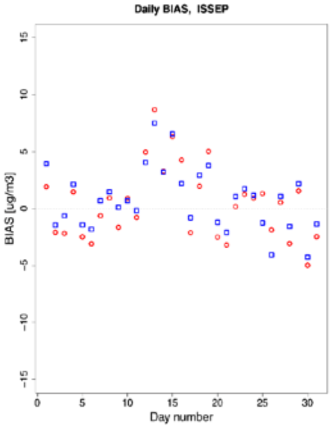
The calibration reduces the CRMSE for all three calibration methods.
The BIAS is less affected.

(The BIAS and CRMSE are <u>not</u> normalised
using the uncertainty of reference PM2.5 measurements)

Results (2)





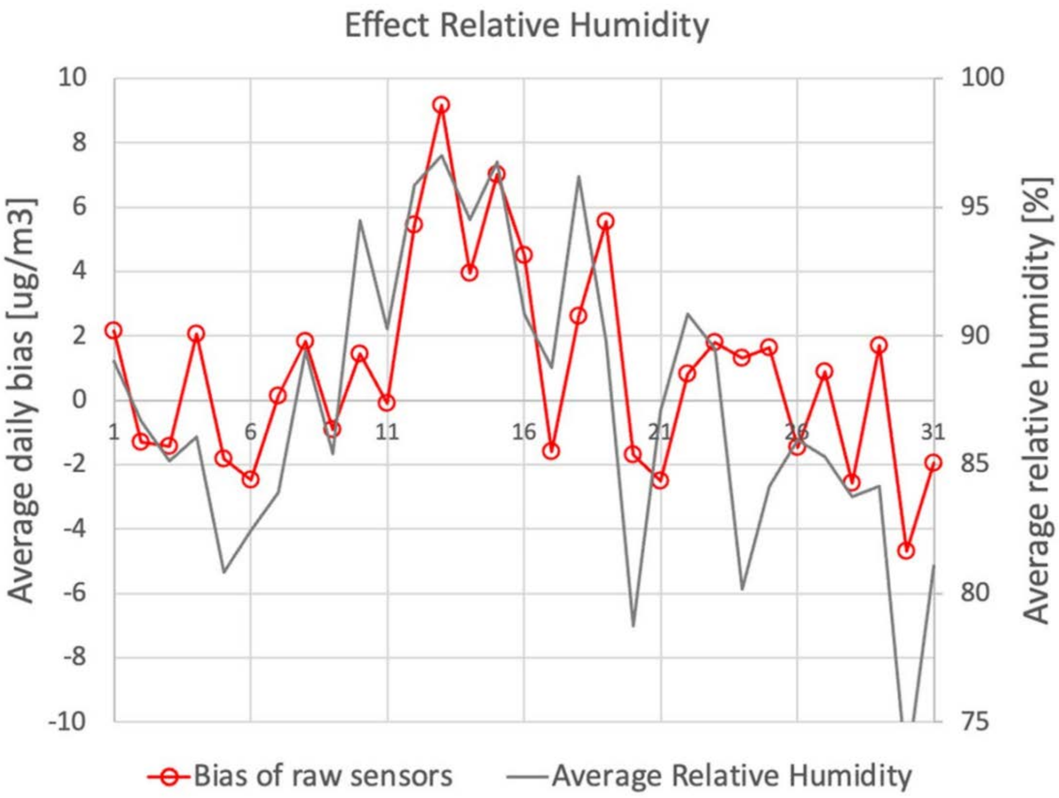


Calibration	median R		median RMSE µg/m^3		mean abs day bias µg/m^3		stdev day bias µg/m^3	
	Raw	Cal	Raw	Cal	Raw	Cal	Raw	Cal
RIVM	0.84	0.89	7.68	6.02	2.53	1.07	3.14	0.58
INERIS	0.84	0.88	7.65	6.15	2.56	1.15	3.15	0.97
ISSEP	0.84	0.87	7.26	5.43	2.53	2.21	3.17	2.73

Period Jan 01-31, 2022

- Daily average of all sensors for raw data (red) versus the calibrated data (blue)
- Significant improvement of daily BIAS by RIVM and INERIS method

Sensor error & humidity



Discussion and Conclusions (1)

- Benchmarking is an important and useful process to study the effects of different approaches in the calibration of data from large networks of low-cost sensors.

- Sufficiently realistic synthetic real concentrations and synthetic sensor data can be constructed, and these are valuable for an objective benchmark of different sensor network processing algorithms.

- The importance of data cleaning, handling of uncertainty, interpolation, and calibration of low-cost sensors was demonstrated and investigated.

- The algorithms applied in the benchmark for network calibration can substantially correct the influence of environmental conditions on the performance of the SDS011 PM2.5 sensors.

FAIRMODE

Discussion and Conclusions (2)

- The methods employed by RIVM-INERIS are suited for a calibration approach looking for a robust good mean calibration, with tolerance for a few “bad” corrected sensors, whereas the ISSeP method is suited for calibrations with low tolerance for badly corrected sensors.

- The SDS011 sensor, used as a basis for the synthetic data, has a large random uncertainty that cannot be corrected by network calibration.

- Combining the calibrated PM2.5-sensor data with existing air quality maps in a data fusion approach is expected to improve the level of detail and the quality of the air quality maps, especially when zooming in spatially and in time.

Questions ?