FAIRMODE/CT6

Benchmarking low-cost sensors

Plenary Meeting April 28, 2022

- Status and outlook of CT6
- Recommendations wrt sensors

1

Low-Cost sensors benchmarking

- At the FAIRMODE meeting in Berlin (2020) the topic of sensor networks was discussed. It was decided to include this topic in the road map for the next years as a "Benchmarking" topic. The Benchmarking stage is intended as a first step that aims at exploring and comparing results from different approaches, in this case of using/exploiting sensor networks.

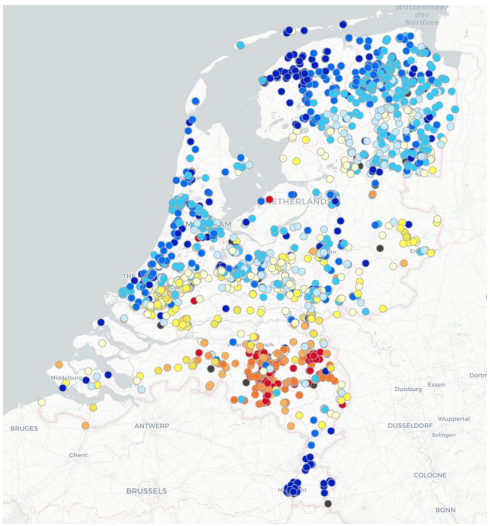
- The FAIRMODE road map describes Benchmarking as:

  "This stage also requires developing and testing a standardized evaluation or inter-comparison methodology (possibly supported by common tools and common datasets) for collecting and reporting model inputs and outputs in a way that enables relevant comparisons.".

2

FAIRMODE

Low-Cost sensors benchmarking

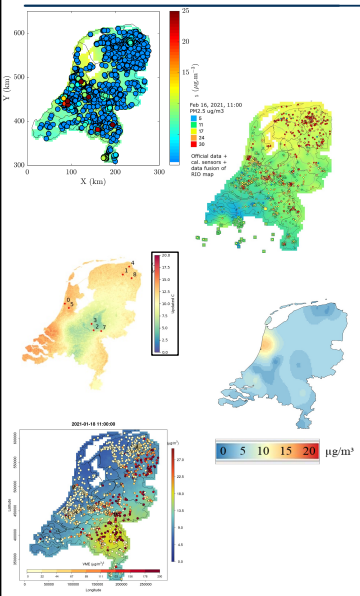


- Use data from low-cost sensors (~ 25€/30$) in the Netherlands providing $PM_{2.5}$, mostly Nova SDS-011.
- Since January 2021, hourly sensor data and official data and model results provided to participants on a real-time basis.
- All interested FAIRMODE participants can use these data to work on:
  - Selection and calibration of sensors;
  - Individual sensors / network;
  - Data fusion/assimilation.
- **New participants still welcome!**

https://sensors.rivm.nl/

FAIRMODE

Different approaches …



Starting from the same $PM_{2.5}$ data set, a number of participants use their own approaches and tools to get the optimal results for calibration and (eventually) data fusion.

- **INERIS** (FR), **VITO** (BE), **RIVM** (NL): network approach, data fusion of existing $PM_{2.5}$ maps with cleaned-up/calibrated data.
- **U. Aveiro** (PT): AI/ ANN as tools to support future methodologies (is there enough data?)
- **ISSeP** (BE): Looking at selected sensors, close to official data.
- **UC. Cork** (IE): looking at correlations between groups of sensors.
- **VMM** (BE): look at hyper-local concentrations.

Results were presented at the FAIRMODE technical Meeting.

Evaluate performance …

- We try to process the raw results of low-cost sensors in such a way that the best estimates of the "real" concentrations can be obtained.
- Usually, we do not know the actual concentrations, so we cannot test the quality of different algorithm's in a simple way.
- Knowing the "real" concentrations makes it possible to:
  - Compare results from different calibration methods to real values;
  - Objectively test the effects of variations in calibration strategies.

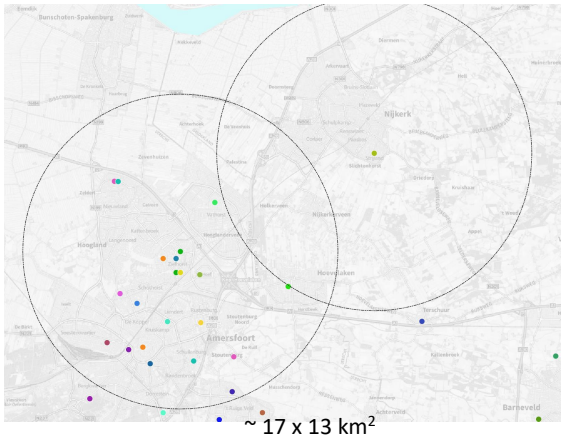
- Alternatively, we can generate **synthetic data** to test different algorithm's.
  - It is essential to take all the (seemingly) chaotic aspects of sensors into account.
  - Analytical distributions will probably not fully describe the behaviour of low-cost sensors.
  - Use behaviour of actual sensors to create synthetic sensor data.

5

Synthetic data, spatial correlations

- For every hour we define the "real" concentrations.
- For every location of official data, we combine the "real" concentration with the variations observed in official measurements in the neighbourhood, thereby producing synthetic official data.
- For every location of a sensor, we combine the "real" concentration with the variations observed in actual sensor data in the neighbourhood, thereby producing synthetic sensor data.

- Actual behaviour drives synthetic behaviour.
- Note: the behaviour of a sensor at time *t* is not related to that at time *t+1*.

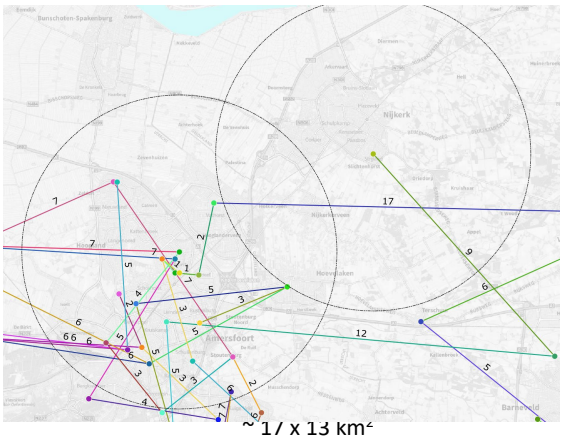


~ 17 x 13 km^2

6

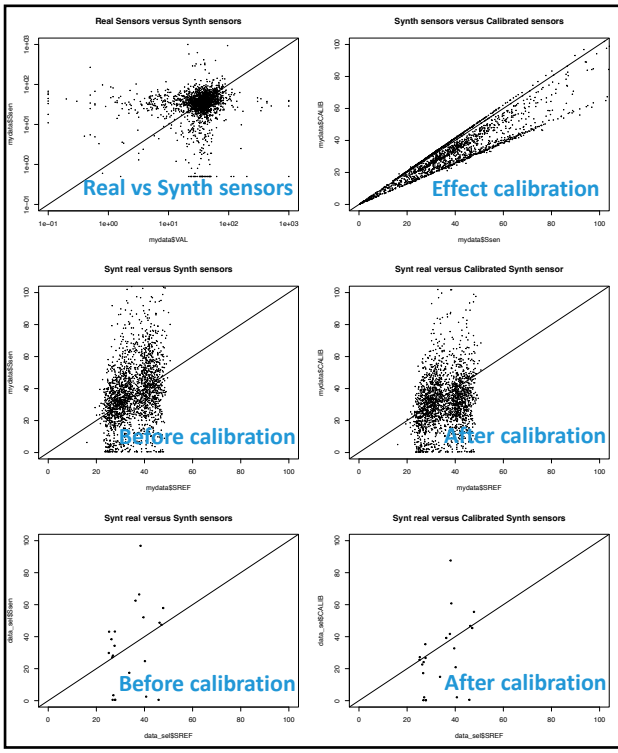
Spatio-temporal correlations

- In order to create more temporal correlations between synthetic sensor values, we link the behaviour of every sensor to that of a nearby other sensor.
- A random component is added.
- Each sensor is linked to the same other sensor as long as that other sensor is available in the hourly data sets.
- Actual spatio-temporal behaviour drives synthetic behaviour.

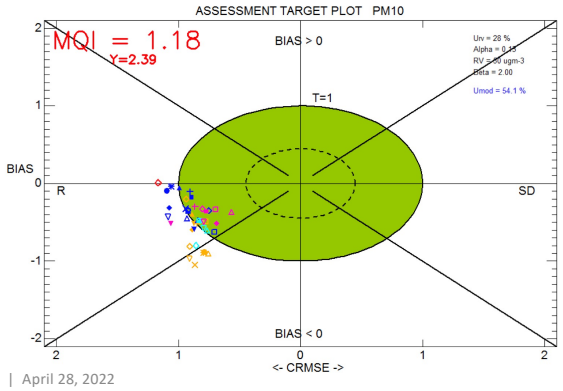


7

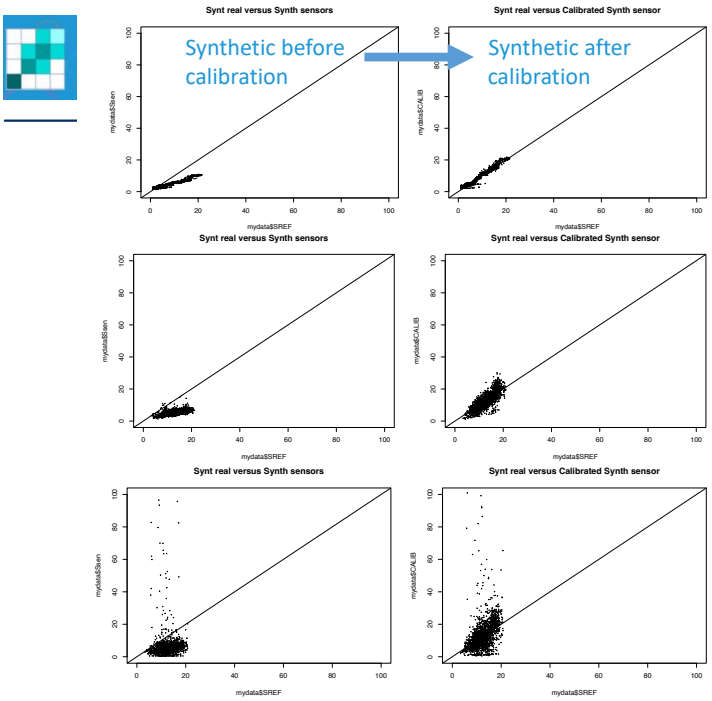
Performance calibration?



- How to evaluate the performance of calibration algorithms?
- Simple scatter diagrams are not very helpful.
- Target diagram?



8



Synthetic data …

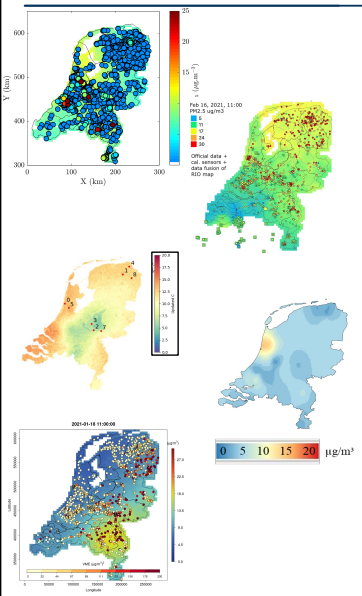
- With synthetic data it is possible to vary the amount of noise in the sensor data and test the effect on calibration methods.
- The figures show, top to bottom, little noise, more noise, real data.
- Statistical (network) calibration can correct for a systematic bias in sensor data, but cannot reduce the noise in individual sensors.

2022

9

FAIRMODE

Using the calibrated data



- Global hourly calibration (statistical calibration) can
  - identify malfunctioning/suspect sensors
  - correct for regional environmental effects
  - …
- It is difficult to improve the hourly results of individual sensors in a detailed way.
- We can combine the results of many uncertain results:
  - Smart interpolation
  - Data Fusion/Assimilation
  - AI/NN techniques
  - …
- Next steps?

10

CT6: Status and Next steps …

1. **Present status**: Work in progress on several different strategies for selection, calibration and data fusion of low-cost sensor data.
   - Further develop the different approaches.
   - Test the use of synthetic data.
   - Define a metric to evaluate sensor calibration.
   - Compare results obtained using different approaches.
2. **Next steps:**
   - Identify and combine/integrate the strong points of the different approaches.
   - Optimal use of the calibrated sensor data in data fusion/data assimilation schemes.
   - Test synthetic data for evaluating different approaches.
   - Use better sensors?

https://fairmode.jrc.ec.europa.eu/activity/ct6

11

Status of work and Next steps …

Publications?

- Abstract submitted for HARMO conference in Aveiro.
- End of 2022: Status overview, best practices, benchmark results as a FAIRMODE/JRC document.
- Publication around the end of 2022 on the overall experiences, guidance, best practices of processing data from low-cost sensor networks.
- If possible, create and release a standard set of (synthetic) data for general use.

12

Recommendations → sensors

3.3 Recommendations

6. Low-cost sensors are an emerging technology that opens opportunities for additional information to be used in assessments. However, proper attention should be given to the QA/QC process. Therefore:

   a. FAIRMODE recommends, in addition to using individually calibrated low-cost sensors, to calibrate/validate groups of low-cost sensors in a network setting. A network calibration, where multiple sensors can be dealt with in batch, can extract useful information from sensors where the individual quality of the sensors is limited or unknown.

   b. FAIRMODE recommends to further develop a QA/QC procedure for low-cost sensor networks to guarantee sufficient added value of the measurement technology. Within a few years, (low cost) sensors in a sensor network are expected to be qualified as indicative measurements for specific pollutants under the AAQDs.

   c. FAIRMODE recommends, once the QA/QC procedure is developed, to integrate sensor data in modelling results via data fusion or data assimilation techniques to improve the overall quality of the air quality assessment methodologies.

13

Recommendations → sensors

Several remarks

- Groups of sensors?
  The recommendation concerns both individually calibrated low-cost sensors as well as calibrated/validated groups of low-cost sensors *in a network setting*.
  (How many sensors make a network?)

- Sensors are indicative measurements?
  Sensors in networks do not necessarily need to qualify as indicative measurements on an individual basis to be useful in fusion/assimilation. When used on an individual based, they indeed should qualify on an individual basis.

- Data assimilation/fusion in general?
  There are no recommendations on general data assimilation/fusion.

14

FAIRMODE

Questions / Discussion

15