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

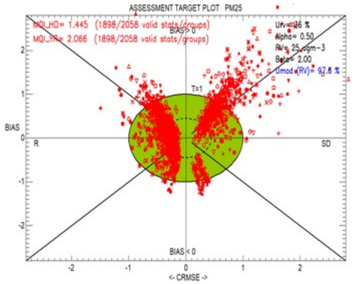
Work done in CT6
"Low-cost sensors"

In the period 2020-2022

FAIRMODE

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

Contents



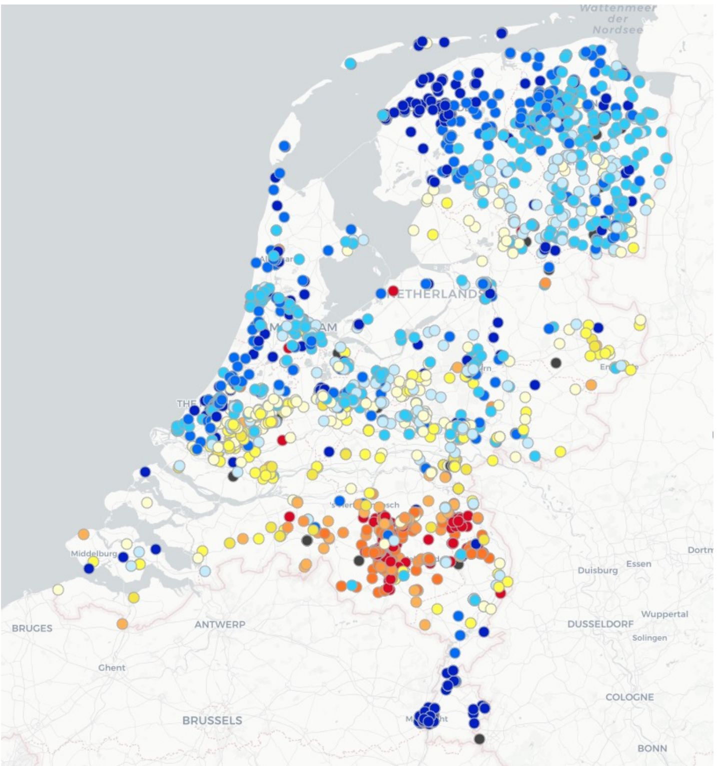
- CT6
- Low-cost PMxx sensors for air quality
- Benchmark
- Analysis and processing of sensor data, comparing results
- Synthetic sensors
- Results
- (Preliminary) Conclusions and Next steps

FAIRMODE

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

CT6 Benchmark

At the FAIRMODE meeting in Berlin (2020) the topic of <u>sensor networks</u> was discussed. It was decided to include this topic in the road map for the next years as a “Benchmarking” topic.



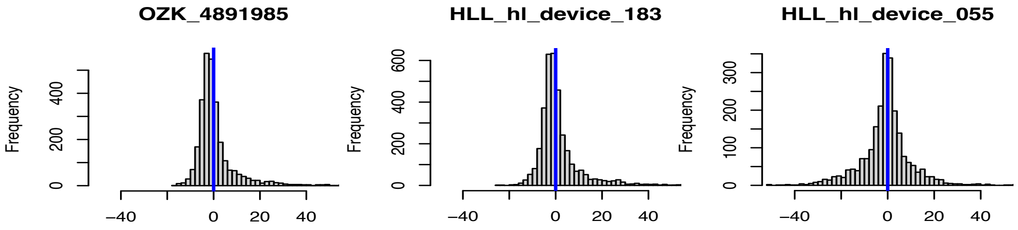
- Exchanging potential concepts and best practices about the integration of sensor network data in air quality mapping methods (Calibration vs DF).
- Exploring how air quality modelling can contribute to the exploitation and validation of an air quality sensor network.

https://sensors.rivm.nl/

FAIRMODE

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

Low-cost sensors for air quality



The quality of sensors can be assessed on an individual basis, work done in CEN WG42.

Alternatively, sensors can be analysed in a <u>network approach</u>; try to calibrate sensors using information obtained from other (nearby) sensors and other measurements.

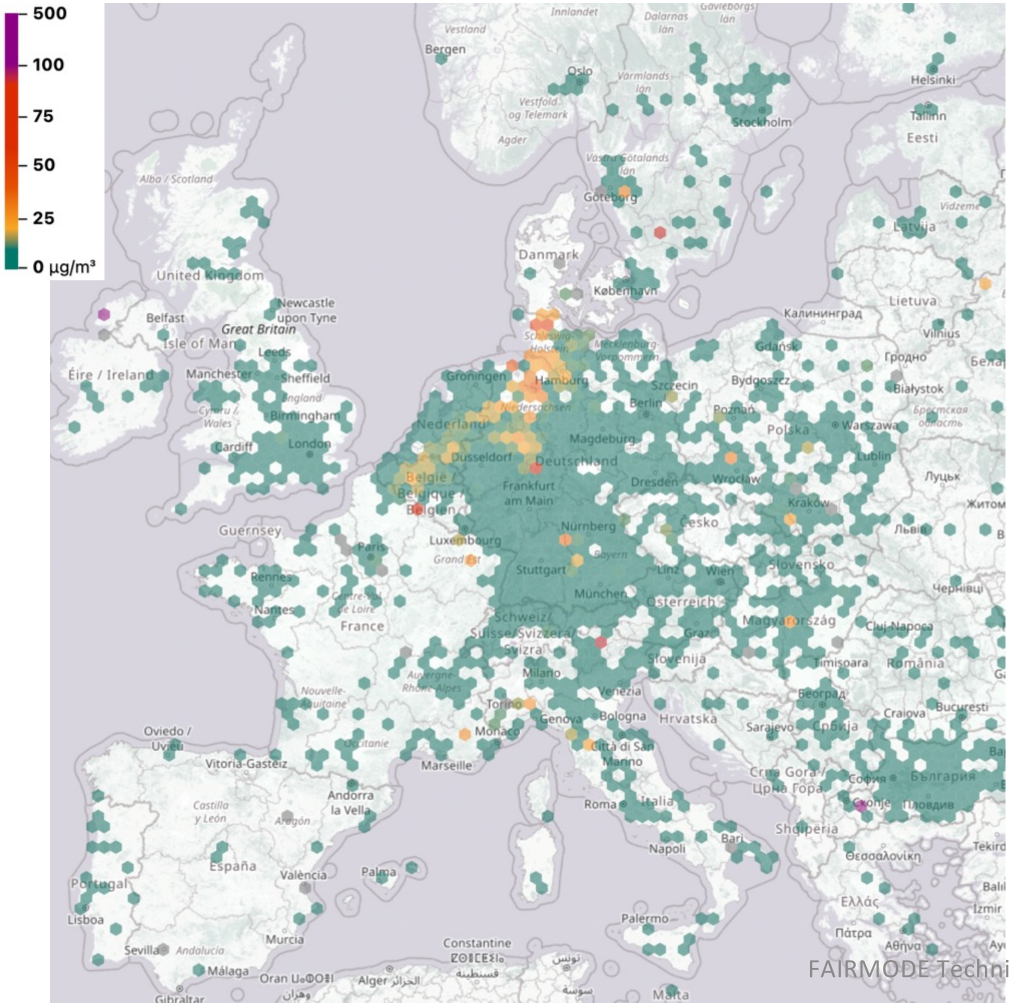
- Low-cost PM10/PM2.5 sensors (~ 25€/30$) have become very popular in (citizen science) networks.

- A popular PM10/PM2.5 sensor is the Nova SDS011.
- Easy to use and connect to any microcomputer.
- Big drawback: quite sensitive to environmental conditions, mainly humidity (is counted as PMxx).

FAIRMODE

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

Sensors for air quality

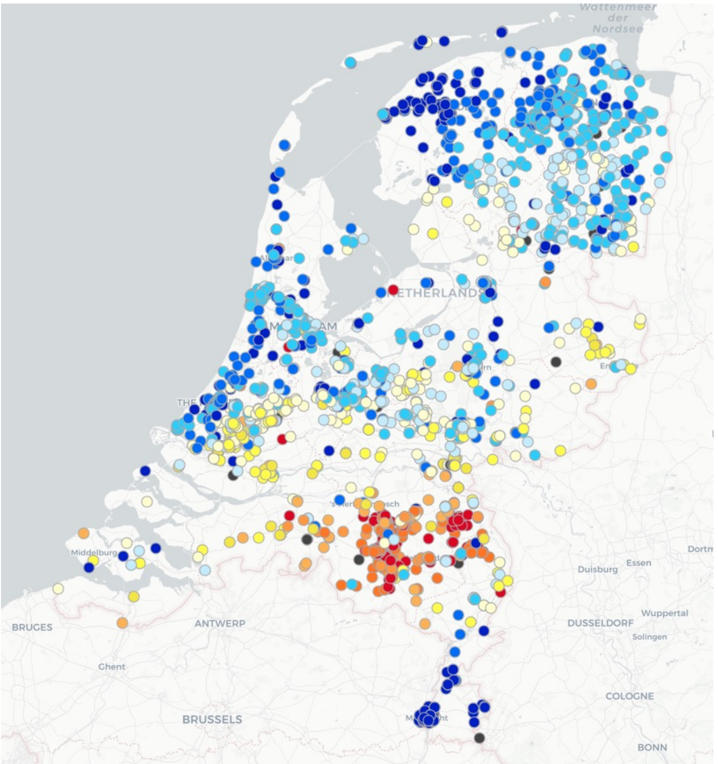


- Several PM_{xx} sensor networks are operational in Europe.
- Biggest citizen driven network is the German Sensor.Community (https://sensor.community/en/).
- In Europe, roughly 10000 sensors of the type Nova SDS011 operational.
- Real concentrations / humidity?

FAIRMODE

Benchmark

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



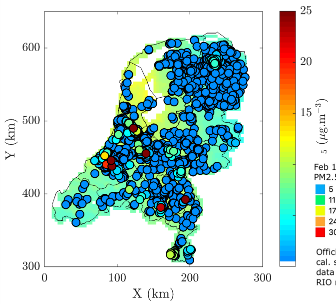
https://sensors.rivm.nl/

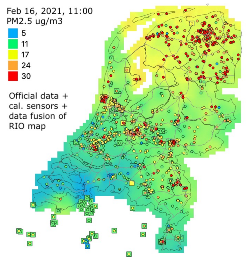
- Use data from low-cost sensors in the Netherlands providing $PM_{2.5}$, type Nova SDS-011.
- Since January 2021, hourly sensor data, official data and model results are provided to participants on real-time basis.
- All interested FAIRMODE participants can use these data to work on:
  - Selection and calibration of sensors;
  - Individual sensors / network;
  - Data fusion/assimilation.

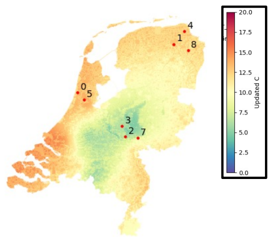
FAIRMODE

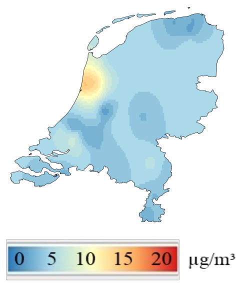
National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

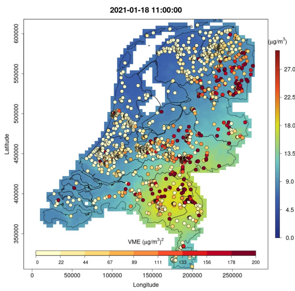
Different approaches











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.

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

Analysis and processing

INERIS, ISSeP and RIVM use the available data to develop/test their selection and calibration methods and, later, for data fusion/assimilation.

INERIS

- Categories of sensor observations: clustering based on distance between sensors, their typology and season.
- Estimate local correction factor and interpolation by kriging.
- Later: Apply SESAM (data fusion with **SE**n**S**ors for **A**ir quality **M**apping) tool: fusion of sensor data and official map considering data variability.

ISSeP

- Measurements from reference stations are used to produce interpolated [PM_{xx}] fields for the studied area. Interpolations are done using the DIVA tool.
- Selected sensor measurements are compared to co-located interpolated reference values
- Sensor values are corrected using linear parameters.

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

- Outliers detection methodology based on lowest/highest sensors.
- Look for sensors in the vicinity of the reference stations, then estimate local correction factor and interpolation correction field.
- Later: Apply data fusion by Bayesian weighing of sensor data and official map considering data uncertainties in both.

Different results → Comparison?

- So, we have many results from different analyses, what now?

- We do not know the actual “real” concentrations at the (majority of the) locations of the sensors, so we cannot test the quality of different algorithm’s in a simple way.
- Knowing the “real” concentrations would make 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 sensor 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.
  - Idea: use behaviour of actual sensors to create synthetic sensor data.

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

Synthetic sensor data

For every hour we have Dutch field-data:
- Official PM2.5 measurements (~55).
- Interpolated official data using the RIO model (1x1 km^2).
- Data from low-cost PM2.5 sensors (2000+).

→ Assume that the set of differences between the sensors and results from RIO in a specific hour is a sufficient approximation for the set of deviations of the sensors.

Create synthetic data for a sensor at a specific location:
- Construct a realistic "real" concentration.
- Add a sample from the pool of deviations of nearby sensors for this hour.
- Option: consistently link a synthetic sensor to deviations from the same nearby real sensor in the pool to create consistent behaviour over time.

We use actual behaviour from real sensors in the field to create the synthetic sensors.

☺ Realistic behaviour of sensors.

☹ Only works if you have actual sensors in the field.

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

Result analysis/calibration

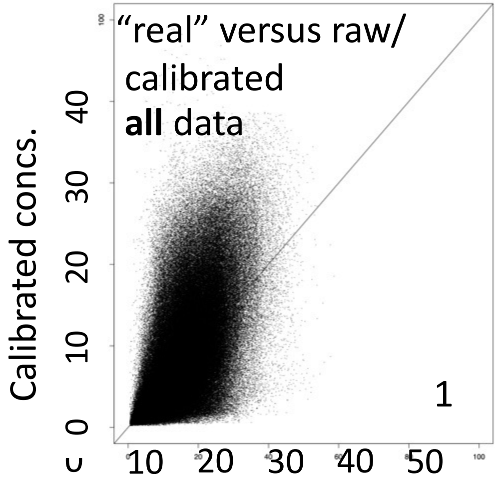
Benchmark test

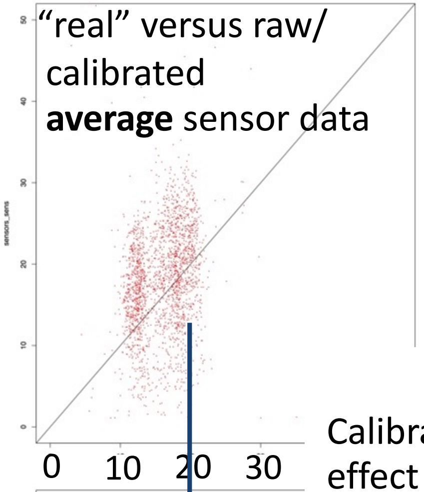
Synthetic sensors

Jan 10-19, 2022

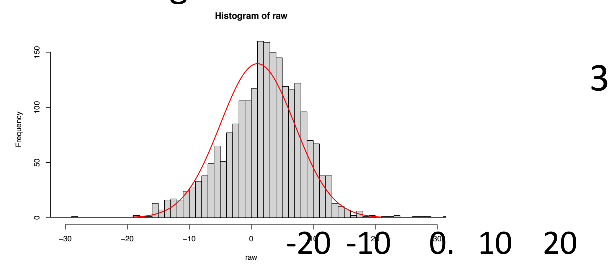
Analysis RIVM

1. All individual hours and sensors (~465000).
2. Average values per sensor (~2100).
3. Histogram of differences synth-real.

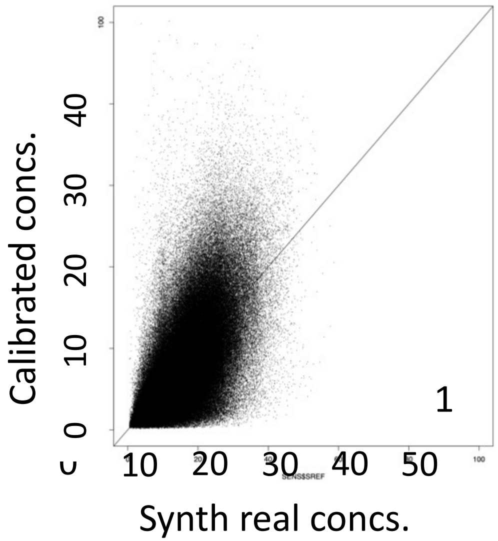


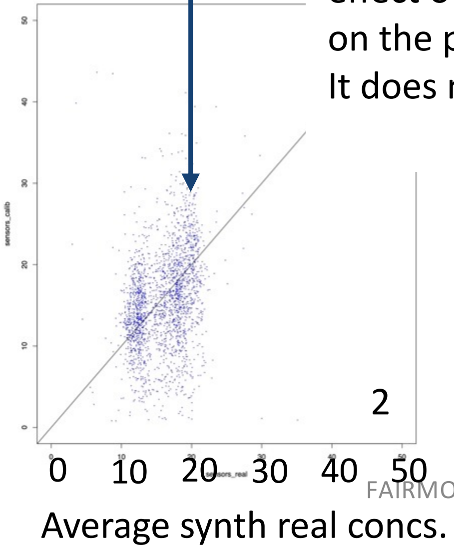


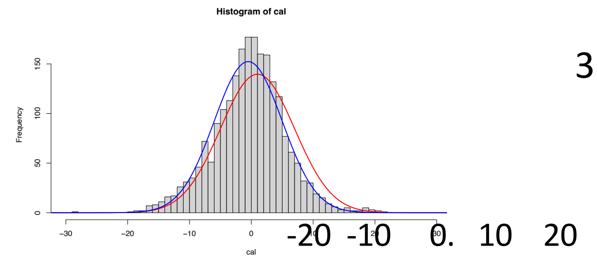
Histogram differences



Calibration reduces the systematic effect of environmental conditions on the performance of the sensors. It does not change the random noise.



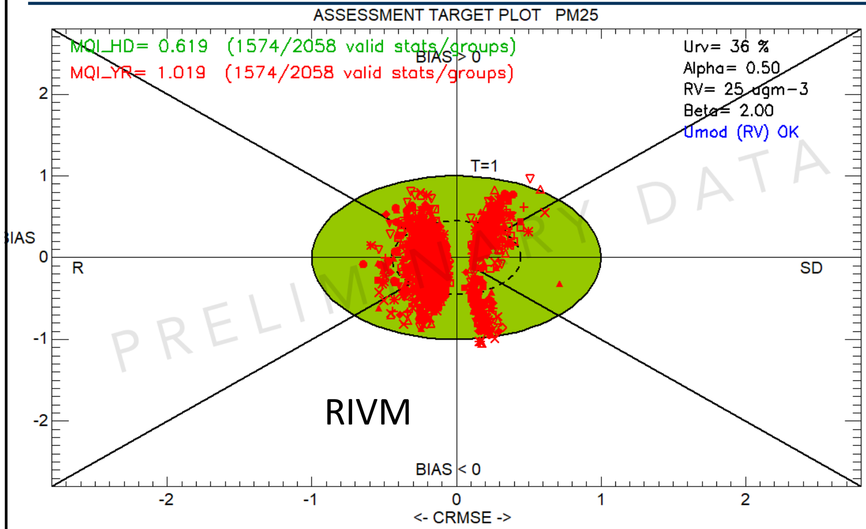


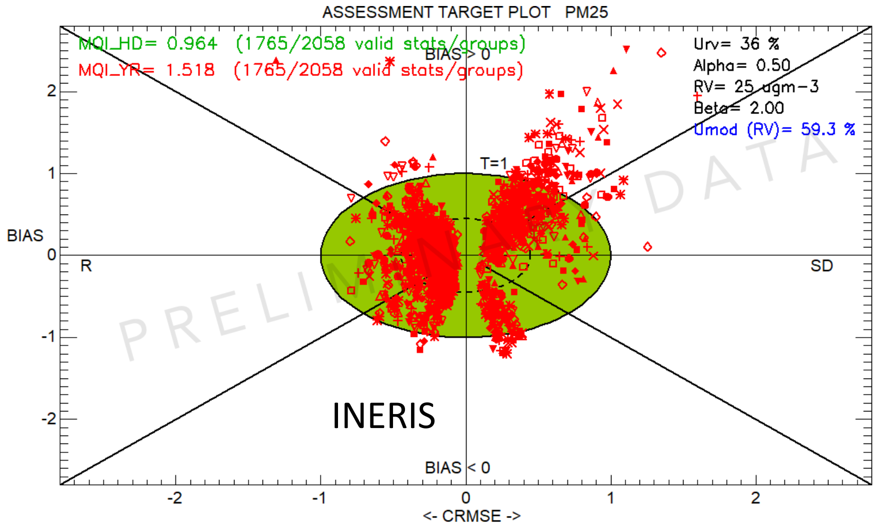


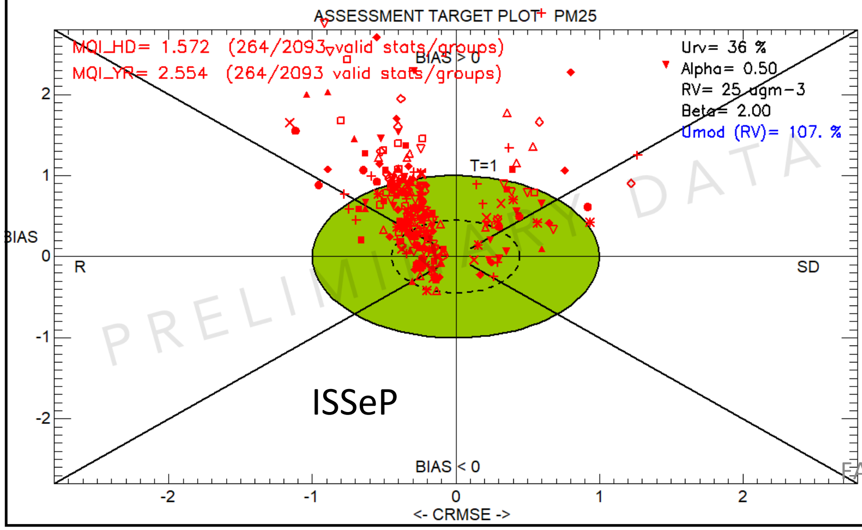
FAIRMODE

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

Comparison using Target plot





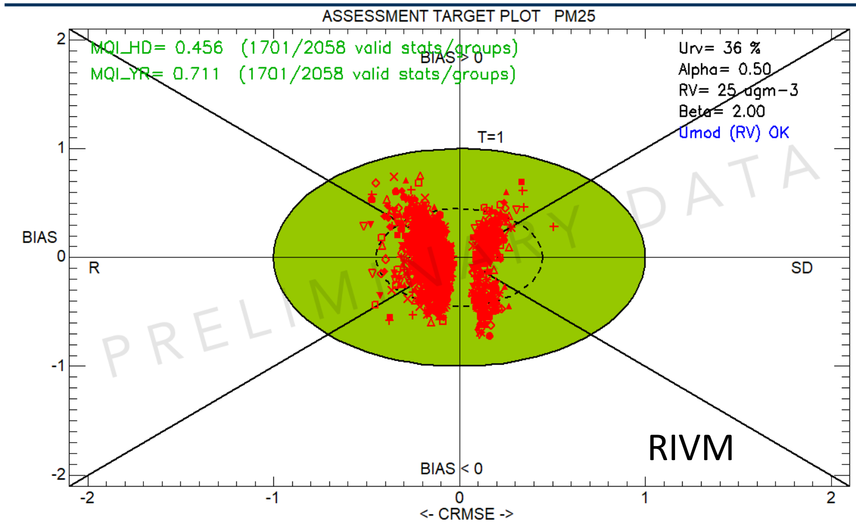


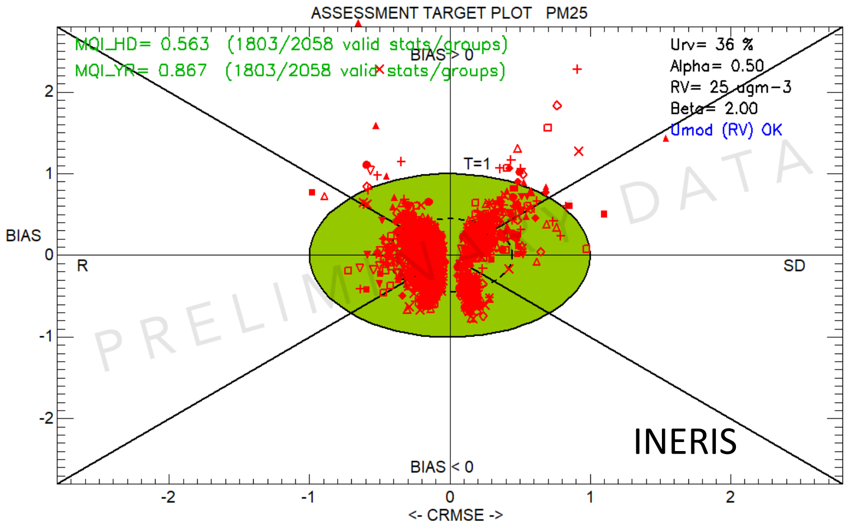
- The results obtained by INERIS and RIVM show similar structures.
- The results from ISSeP are for January 10-16, 2022.
- The different number of data points outside of the green area are partly due to choices in outlier selection.
- The BIAS and CRMSE are normalised by the PM2.5 measurement uncertainty.

FAIRMODE

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

Adjust sensor noise





- The algorithms applied in the benchmark can correct for the influence of environmental conditions on the performance of the sensors.
- The substantial random uncertainty of the SDS011 can not be corrected for.

- What is the effect when the random uncertainty of the SDS011 is reduced by 50%.
- Using synthetic sensors, we can change the random uncertainties of the sensor data.

- The normalisation of the BIAS and CRMSE in the Target diagram by the PM2.5 measurement uncertainty may lead to a biased conclusion regarding the quality of the sensor calibration.

FAIRMODE

Conclusions 1/2

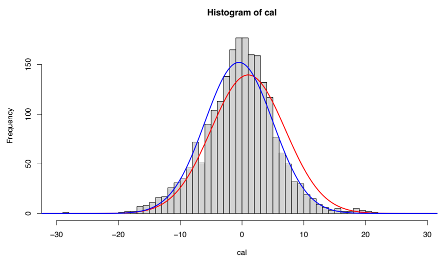
- Benchmarking is an important step in reconciling and comparing results from different approaches of deploying low-cost sensor networks.

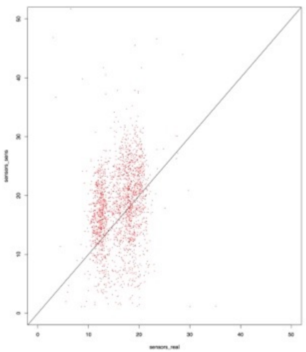
- The importance of data cleaning, handling of uncertainty, interpolation and calibration is demonstrated and investigated.
- Several methods are implemented and tested.

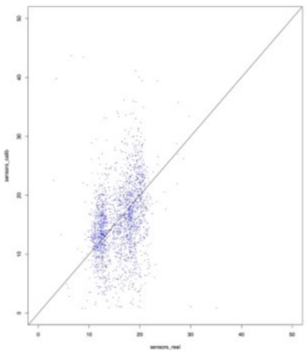
- Sufficiently realistic synthetic sensor data can be constructed and these are valuable for an objective test of sensor-processing algorithms.
- Using synthetic sensor data, it is possible to investigate the effect of different characteristics of the sensors.

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

Conclusions 2/2







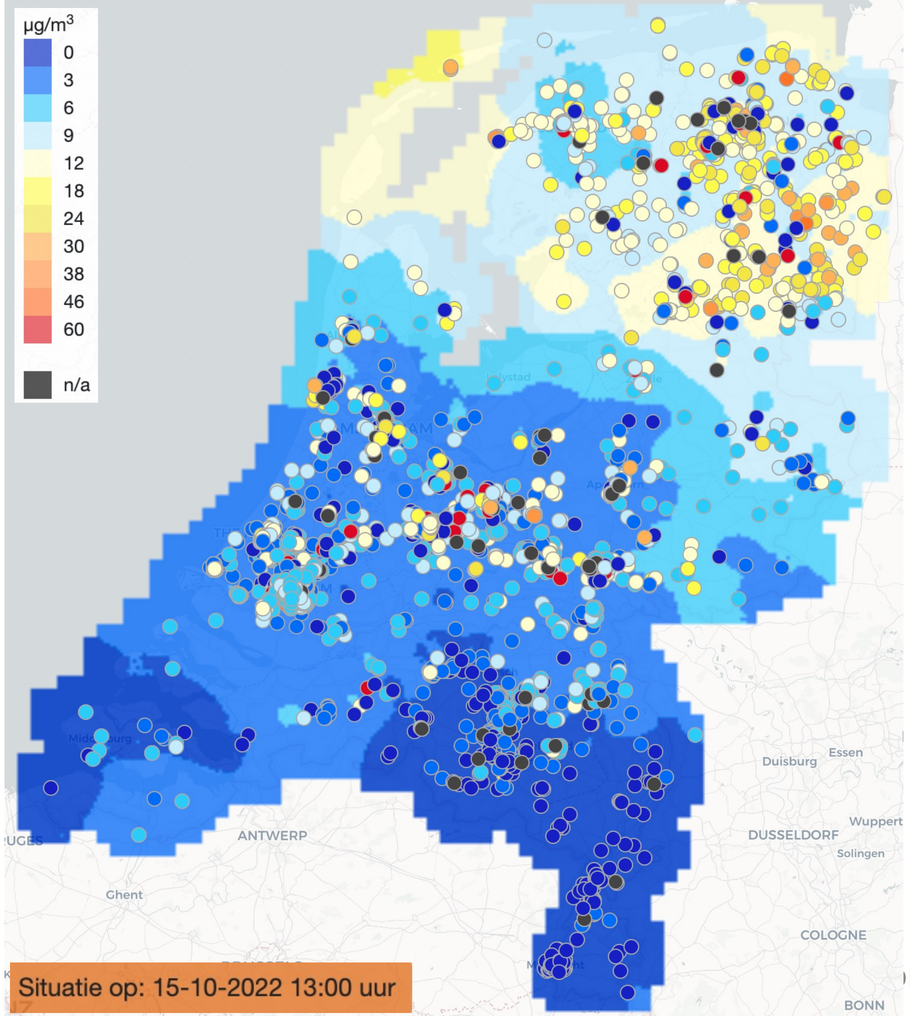
- The algorithms applied in the benchmark for network-calibration can, to a large extent, correct for the influence of environmental conditions on the performance of the SDS011 PM2.5 sensors.
- The SDS011 sensor has a large random uncertainty that can not be corrected for by network calibration → limits individual use.

Sensor network	Little random noise	Much random noise
NOT Sensitive to env. conditions	☺☺☺	😐😐
IS Sensitive to env. conditions	☺☺	😐

Usually better than no information at all.
Many sensors are good for use in maps, DF/DA.

FAIRMODE

Next steps



- We will **extent the test period** to (at least) one month.
- **Analyse** the differences in approaches.
- The results of the calibrated (synthetic) sensor data will be used in **data fusion algorithms** (INERIS, ISSeP, VITO, RIVM, …).
- The results and recommendations will be presented in a **report** and a **publication**.

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

Overall …

- We discussed, implemented and tested several methods for sensor selection and calibration.
- A method to generate and use realistic synthetic sensor data was discussed, implemented and tested.
- Several different approaches for sensor selection and calibration were tested using the synthetic data.
- We are starting to compare the results, obtained so far, in a test of data fusion methods.

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

Thank You !

FAIRMODE