FAIRMODE

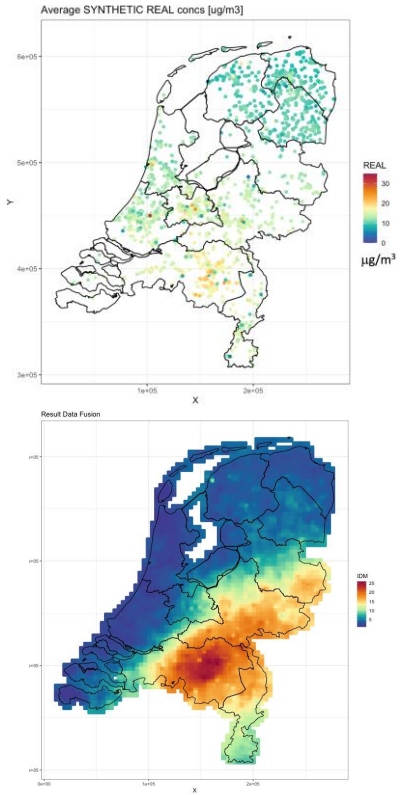
National Institute for Public Health and the Environment
*Ministry of Health, Welfare and Sport*

# Developments around low-cost sensors in FAIRMODE/WG6

**Joost Wesseling**
**Sjoerd van Ratingen**

FAIRMODE

# Low-cost sensors in FAIRMODE




- A recent benchmark of calibration methods for low-cost sensors.
- An ongoing benchmark of data fusion methods using data from low-cost sensors.

FAIRMODE

# Benchmark calibration sensors

- During 2021-2023, benchmark PM2.5 sensor calibration in FAIRMODE, https://fairmode.jrc.ec.europa.eu.
- The work focused on calibration in a **network of low-cost PM2.5 sensors**:
  - No individual a priori calibration or comparison of sensors.
  - Use hourly information from the network of (2000+ in NL) AQ measurements (LCS and official) to estimate calibration of the sensors.
  - Setup and perform a benchmark.
  - Create and use synthetic sensor data.

FAIRMODE

# Benchmark calibration sensors

- During 2021-2023, many discussions and tests.
- INERIS, ISSeP and RIVM used **synthetic data** to develop/test and benchmark their **selection** and **calibration** methods of LCS.
- Data for benchmark: January 2022.

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.

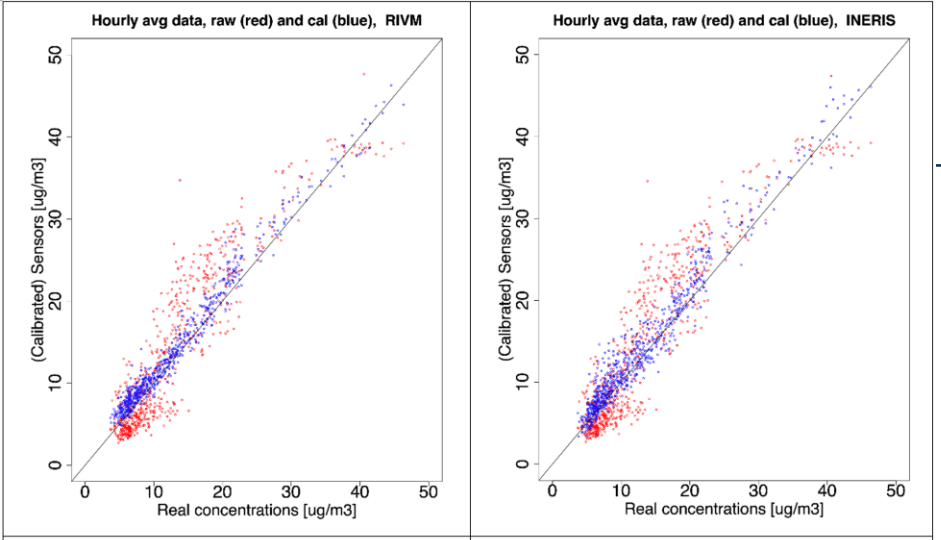
ISSeP

- 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*

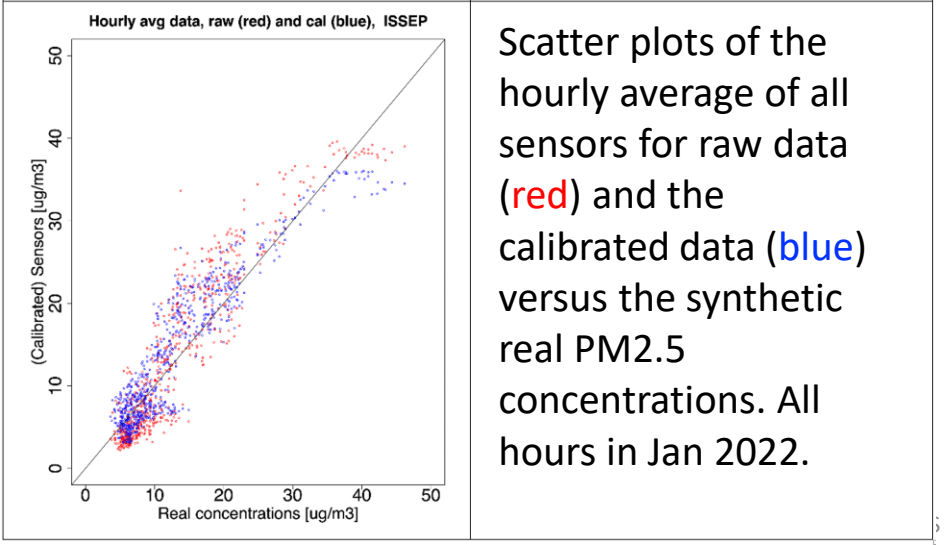
- Outliers detection methodology based on lowest/highest sensors.
- Look for sensors in the vicinity of the reference stations, then estimate local correction factor and interpolation correction field.
- Later: Apply data fusion by Bayesian weighing of sensor data and official map considering data uncertainties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

# Results

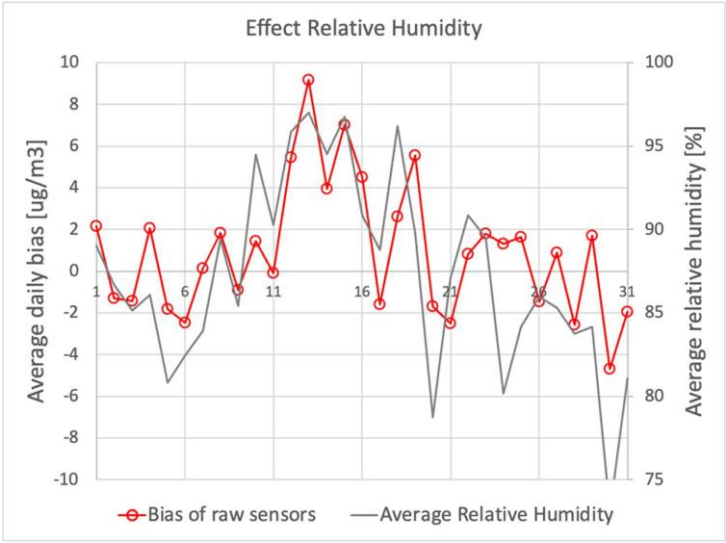









Scatter plots of the hourly average of all sensors for raw data (red) and the calibrated data (blue) versus the synthetic real PM2.5 concentrations. All hours in Jan 2022.




Average daily biases of the raw sensors (red) and the daily average relative humidity (gray), in percent, as a function of the day in the month of January 2022

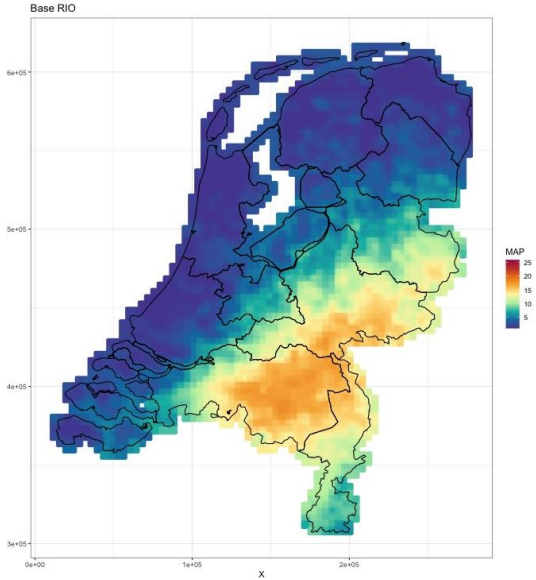
# Conclusions

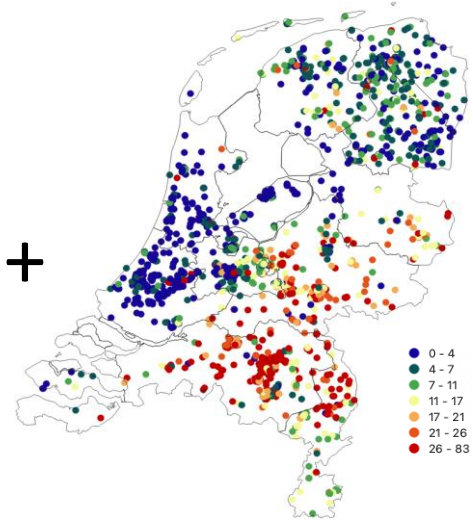
- Sufficiently realistic **synthetic real** concentrations and **synthetic sensor** data can be constructed.
- The algorithms can **substantially correct** the influence of environmental conditions on the performance of the SDS011 PM2.5 sensors.
- The SDS011 sensor, used as a basis for the synthetic data, has a large **random uncertainty** that cannot be corrected by network calibration, which **limits** individual use.
- Combining the calibrated PM2.5-sensor data with existing air quality maps in a **data fusion** approach is expected to improve the quality of the air quality maps.
- "Using synthetic data to benchmark correction methods for low-cost air quality sensor networks", Air Quality, Atmosphere & Health, doi.org/10.1007/s11869-023-01493-z

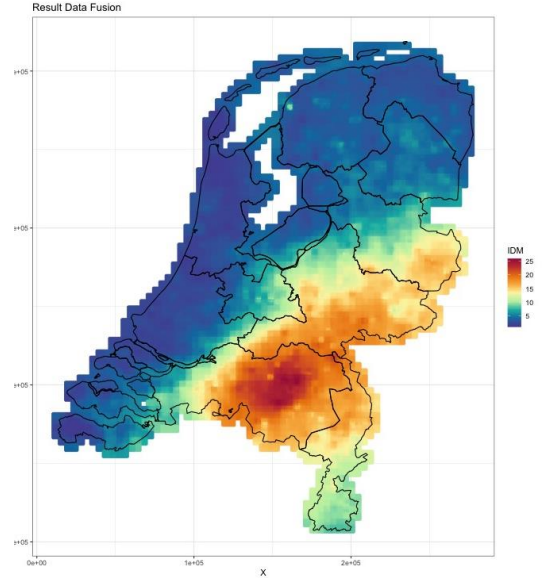
FAIRMODE

# Data Fusion …

Starting from a (PM2.5) concentration field based on official measurements and many sensors, what is the best combination using data fusion?




PM2.5 concentrations calculated with the RIO model based on ~40 official measurements.

Some 2000 low-cost PM2.5 sensors (SDS011), calibrated using official measurements.

Including the sensors in the analysis leads to higher concentrations in parts of the Netherlands. Method used: inverse-variance weighting.

FAIRMODE

# Benchmark Data Fusion

- **FAIRMODE/WG6** is organizing a benchmark on data fusion methods using low-cost measurements/sensor data for PM2.5.
- Data available mid may 2024.
- At least 10 groups from all over Europe want to participate.
- The present setup for low-cost sensors in the Netherlands is used.
- RIVM provides hourly official and (calibrated) sensor data for PM2.5, an initial concentration field for the Netherlands (using RIO model from VITO).
- Different combinations of data will be provided.

- All participants will use the available data and apply their data fusion tools.
- Participants will report the results:
  - Hourly maps of concentrations;
  - Time series on locations that were not used in the data fusion.

# What can we learn?

- A comparison of the results of several different methods/ways to perform data fusion using large numbers of low-cost measurements.

- How are the results of data fusion influenced by the available amount and quality of the input data (official measurements, low-cost measurements, model quality and input).

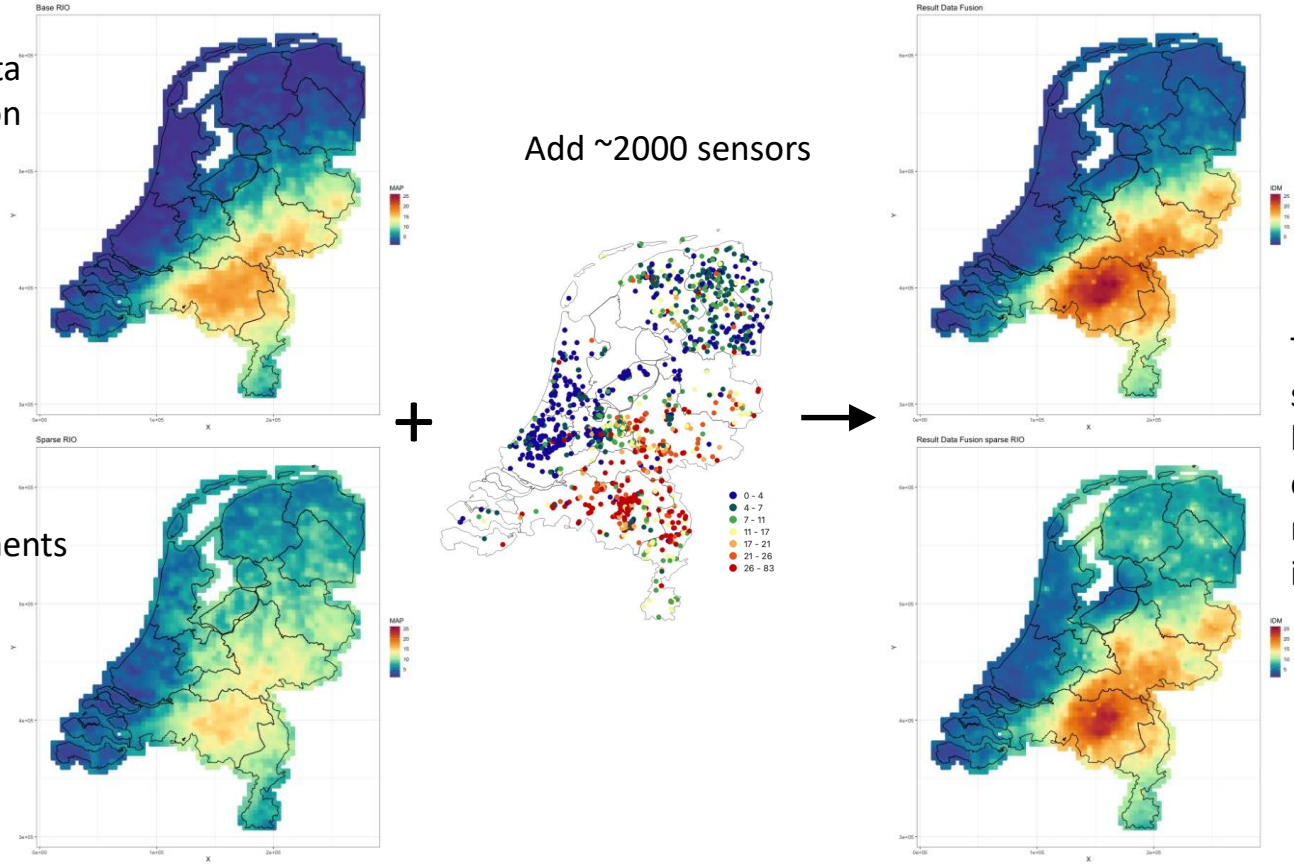
- Can data fusion compensate for less official measurements or an incomplete model?

- What are the (data) requirements for successful use of data fusion?

# Example

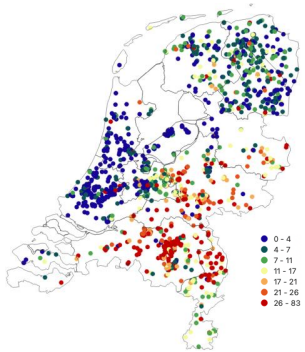
Use **all** official data for a concentration map (RIO).

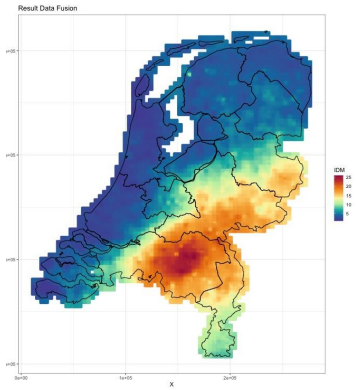
Use **8 out of 40** official measurements for concentration map (RIO).




The available sensors can, to a large extend, compensate for the missing information in RIO.

# Planning 2024




- Mid may: start data stream.
- June: test of correct data and tests.
- September: add one month of winter data.
- Fall 2024: first analyses.

- Interested: sjoerd.van.ratingen@rivm.nl
  joost.wesseling@rivm.nl

FAIRMODE

# Thank You !