WG6 Sensors Benchmarking
Status

Technical Meeting October 2023

Sjoerd van Ratingen (RIVM)

FAIRMODE

Agenda WG6 session

- **Sensor calibration**
  - Recap: Benchmark correction methods for low-cost air quality sensor network (RIVM)

- **Sensor-model fusion**
  - Using sensor networks to improve model reliability (VITO)

  - ~~Enhancing Environmental Monitoring: Leveraging Novel Sensor Technology and Data Assimilation (NILU)~~

  - Short presentation data-fusion methods RIVM

  - Brainstorm on benchmarking data-fusion models

WG6 Sensors Benchmarking

Recap: calibration of low-cost PM2.5 sensors.

Technical Meeting October 2023

Practical work and results

- In the first two years, many subjects and issues were discussed.
- Over the last year, INERIS, ISSeP and RIVM use the available data to (further) develop/test their **selection** and **calibration** methods and, later, for data fusion/assimilation.
- In the first half of 2023, work on article on the results on the selection/calibration benchmark. Submitted for publication in summer.

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.

Issue: validation data

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

- We do not know the actual "real" concentrations at the (majority of the) ~2500 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.
  - We used behaviour of actual sensors to create synthetic sensor data.
- A **data set** with **synthetic data** was created for **January, 2022**, using 50% of the random uncertainty.

Synthetic data

Data base synthetic data:

- PM25 data of 2000+ sensors(SDS011) from the Dutch measure together website (https://sensors.rivm.nl/) and data base.
- Sensors are low-cost (< € 50,--)
- Sensors need to be calibrated because of e.g., sensitivity to humidity.

Synthetic were created for:

1. Real concentrations (unknown to modellers)
2. Reference measurements
3. Sensors

Synthetic data

Creation of **synthetic truth**:

- Represent real concentrations at every PM25 reference and sensor location in the Netherlands
- Use the RIO interpolation model to calculate background concentrations and interpolate results to 1x1 km scale.
- Assume RIO-model to be good proxy for average concentration distribution across country
- On top of this we need **local variations**
  - Use a distribution of differences between the RIO model and the reference measurements at the reference locations.
  - Option 1: Use a national pool of deviations between RIO model and Reference measurements
  - Option2: Use a (more) local pool of deviations between, because sometimes the deviation will typically show spatial variations over the country (draw back: smaller pool)
  - Choice for local variations.

FAIRMODE

Synthetic data

Creation of **synthetic sensor data**:

- Created at PM25 reference sensor location in the Netherlands
- Use the RIO calculation to approximate synthetic real concentrations
- Assume random and systematic errors of the sensors to be zero on average.
- $C_{sensor,k}(x,y,t) = \gamma_k(x,y,t)\,(\,C_{Bgr}(x,y,t) + C_{local}(x,y,t) + \varepsilon_k(x,y,t)\,)$

sensor measurement | meteo-factor | actual background | local contribution | sensor measurement error, zero average, no meteo

Actual sensor value

- Use $\hat{\gamma}_k(x,y,t) = \dfrac{C_{sensor,k}(x,y,t)}{C_{RIO}(x,y,t) + \cancel{C_{local}(x,y,t)}}$ as factor to generate synthetic from RIO

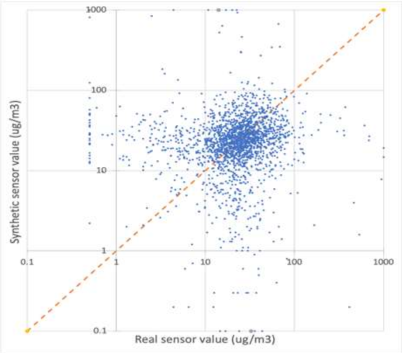
neglecting $C_{local}(x,y,t)$, but yielding even larger factors to generate the synthetic sensor values from the true concentration:

- $C_{synth\_sensor,k}(x,y,t) = C_{synth\_real}(x,y,t)\ \hat{\gamma}_k(\text{neighbourhood}, t)$

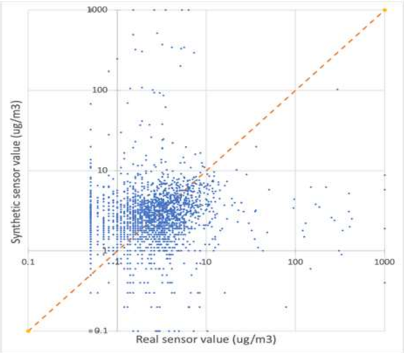
Synthetic data

- Compare **real** and **synthetic** sensor values:

January 11, 2022, 06:00

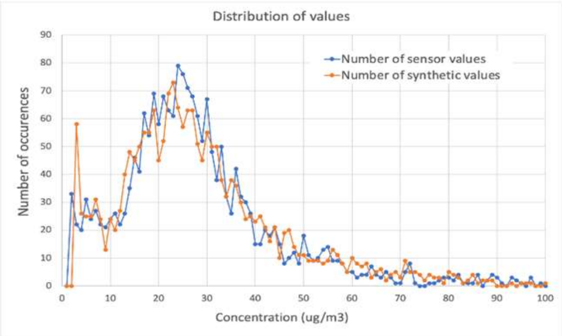


February 08, 2022, 20:00

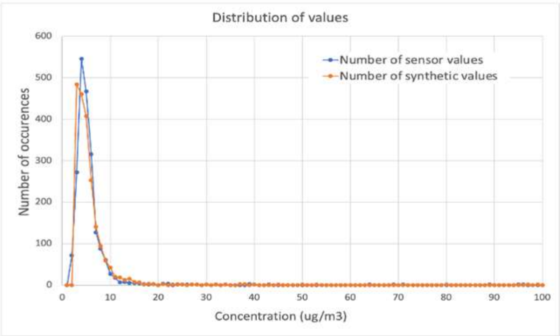


- Negligible correlation between real(x-axis) and synthetic (y-axis) sensors values.

January 11, 2022, 06:00



February 08, 2022, 20:00



- Similar concentration distribution for real (blue) and synthetic (orange) sensor data

Calibration method INERIS

- Perform **data cleaning** on PM25 sensor values
  - No negatives
  - No "frozen" in time
  - Constant positive BIAS (respect to reference measurement)
  - Remove very high peak values

- Perform **outlier detection**
  - First create clusters based on
    - Type of pollution (Urban, Regional, Traffic, Industrial)
    - Distance between sensors
    - Season
  - Outlier detection per cluster based on log-transformed concentrations
  - Eliminate concentrations outside confidence interval

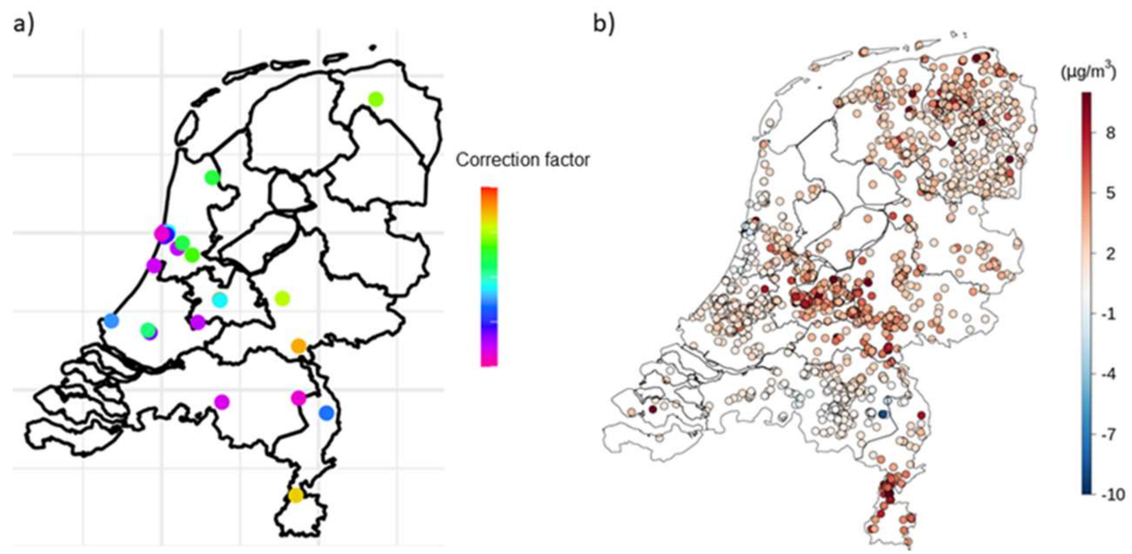
a. clusters defined by site type (n_{type}) b. clusters defined by distance (n_{dist}) c. clusters defined by season (n_{season})

Calibration method INERIS

- Take averages of processed sensor in a radius, surrounding the reference stations,
  depending on pollution type and determine factor as:

$$F_{i\,station} = \frac{C_{i\,station}^{ref}}{C_{i\,station}^{sensors}}$$

- Example of calibration factors calibrated concentrations



Calibration method ISSeP

- Iterative process
  - Use two regression models linking sensor and interpolated field of reference values
    - Linear model
    - Multi-variable non-linear model
    - Regression not only dependent on real-time value but multiple values in time
    - Yielding Corrective parameters, R^2
  - Determine sensor weights from performance index as calculated by above models.
  - Merge reference and weighted sensor values into continuous field of reference values by using DIVA interpolation method, yielding updated interpolated reference field.
- Calibrated sensor values = interpolated field of reference values
  after a few iterations of updating this field with sensor data
- Sensors with a too low performance index drop out.

Calibration method RIVM

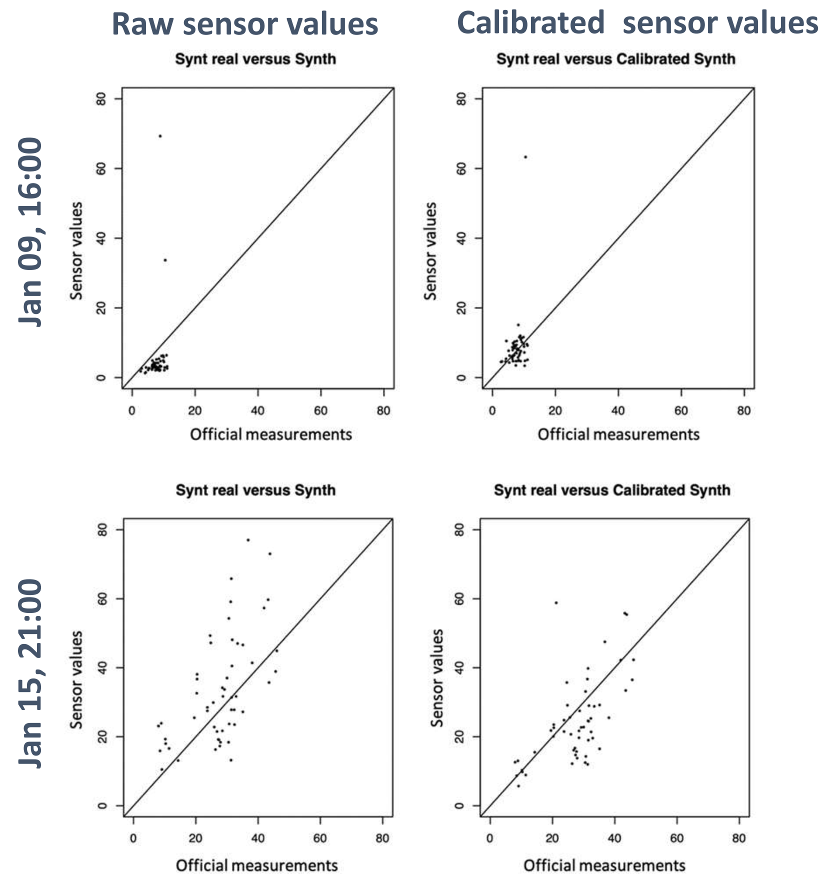
1. Select malfunctioning sensors:
   - Concentrations always almost zero
   - Concentration are very high
   - Don't include lowest and highest 5 percent of raw sensor values

2. Group sensor data. Determine corrections.
   - Group sensor data in clusters with typical max distance = 5 km
   - Also use Germany/Belgium sensors at borders.
   - Depending on number of sensors within group, exclude highest and lowest sensors

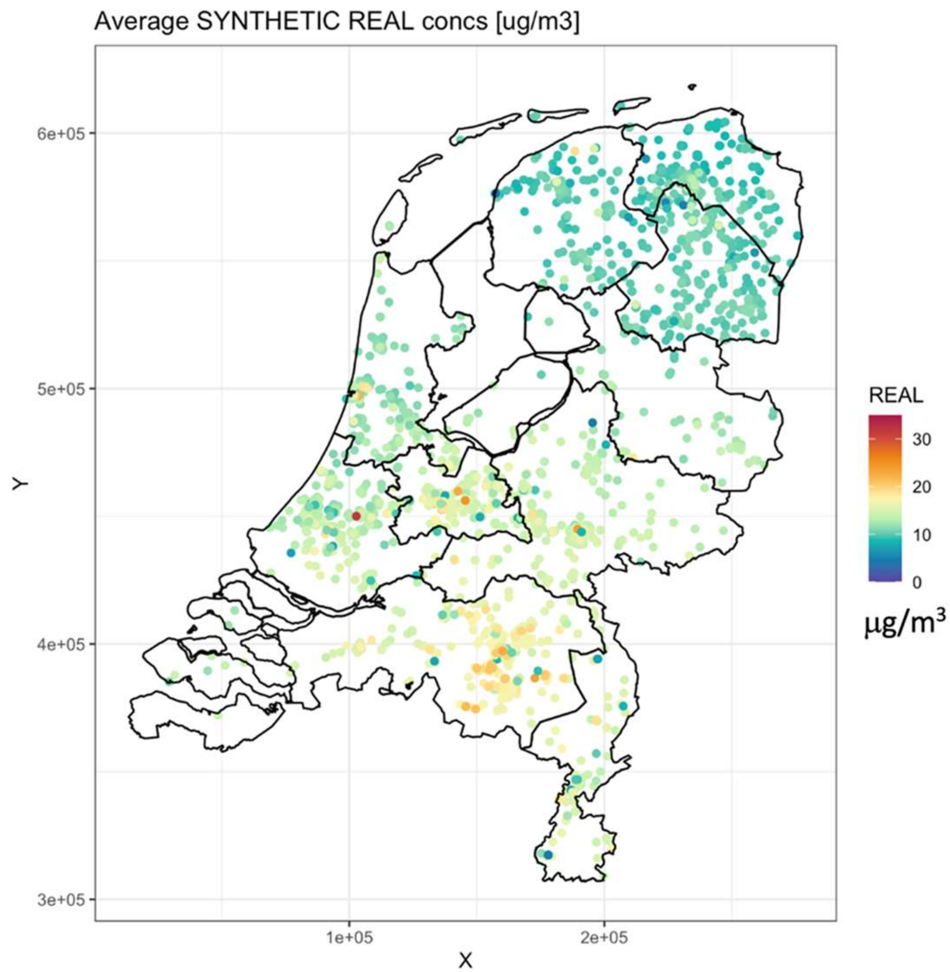
     More sensors → More excluded sensor
   - Divide reference concentration by sensor-group-average

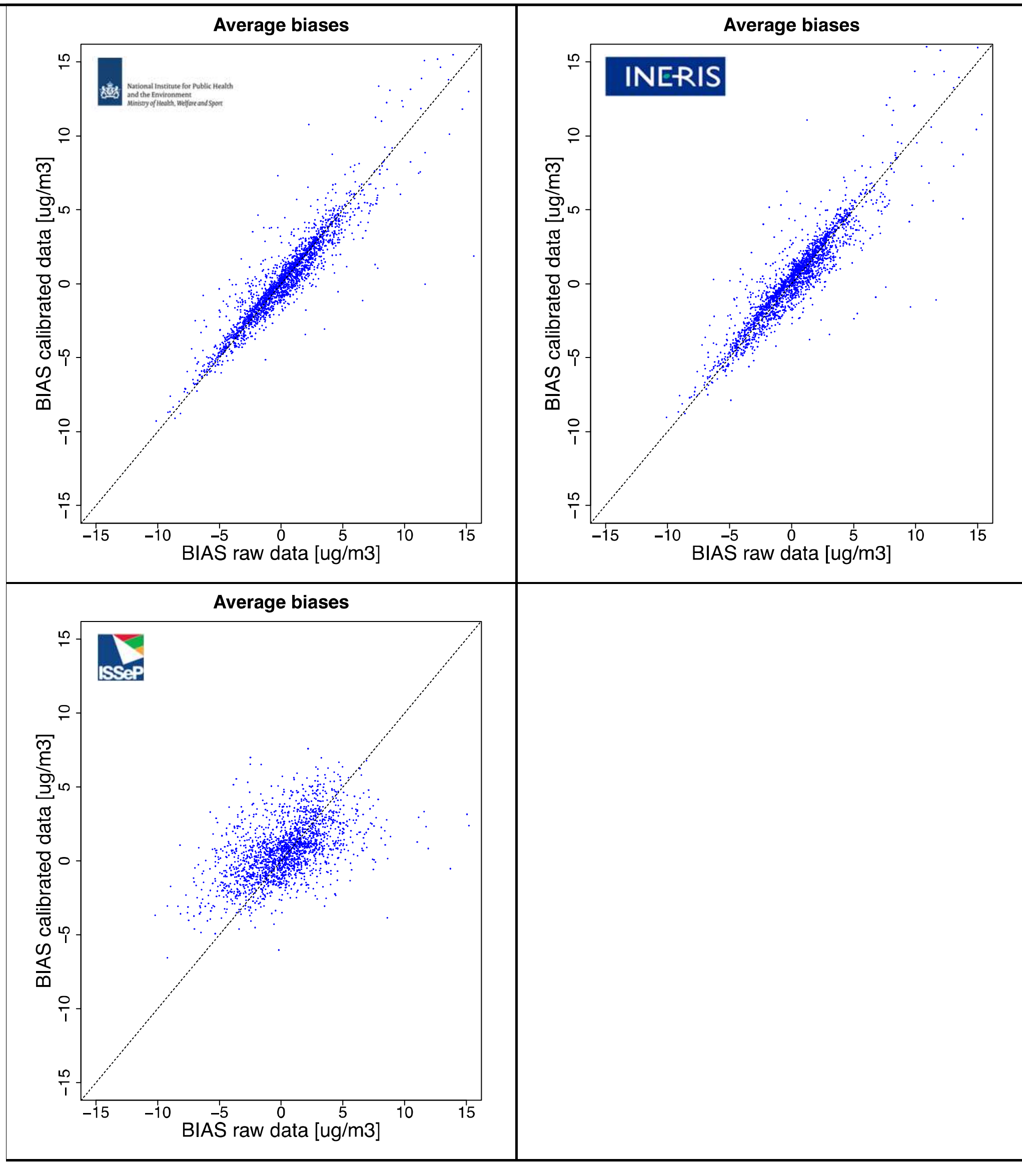
Calibration method RIVM

- Local effect (sensors < 1 km of reference station) calibration on synthetic sensors by comparing with synthetic truth
- Local calibration
  - Reduces average spread
  - Reduces average BIAS
- Interpolate corrections
  - IDW with modified Shepard's method, using only the nearest neighbours
- Calibrate all sensors
  - **All** sensors (including those removed for calculating calibration factor) are calibrated using the interpolated corrections



Synthetic sensors: Average values



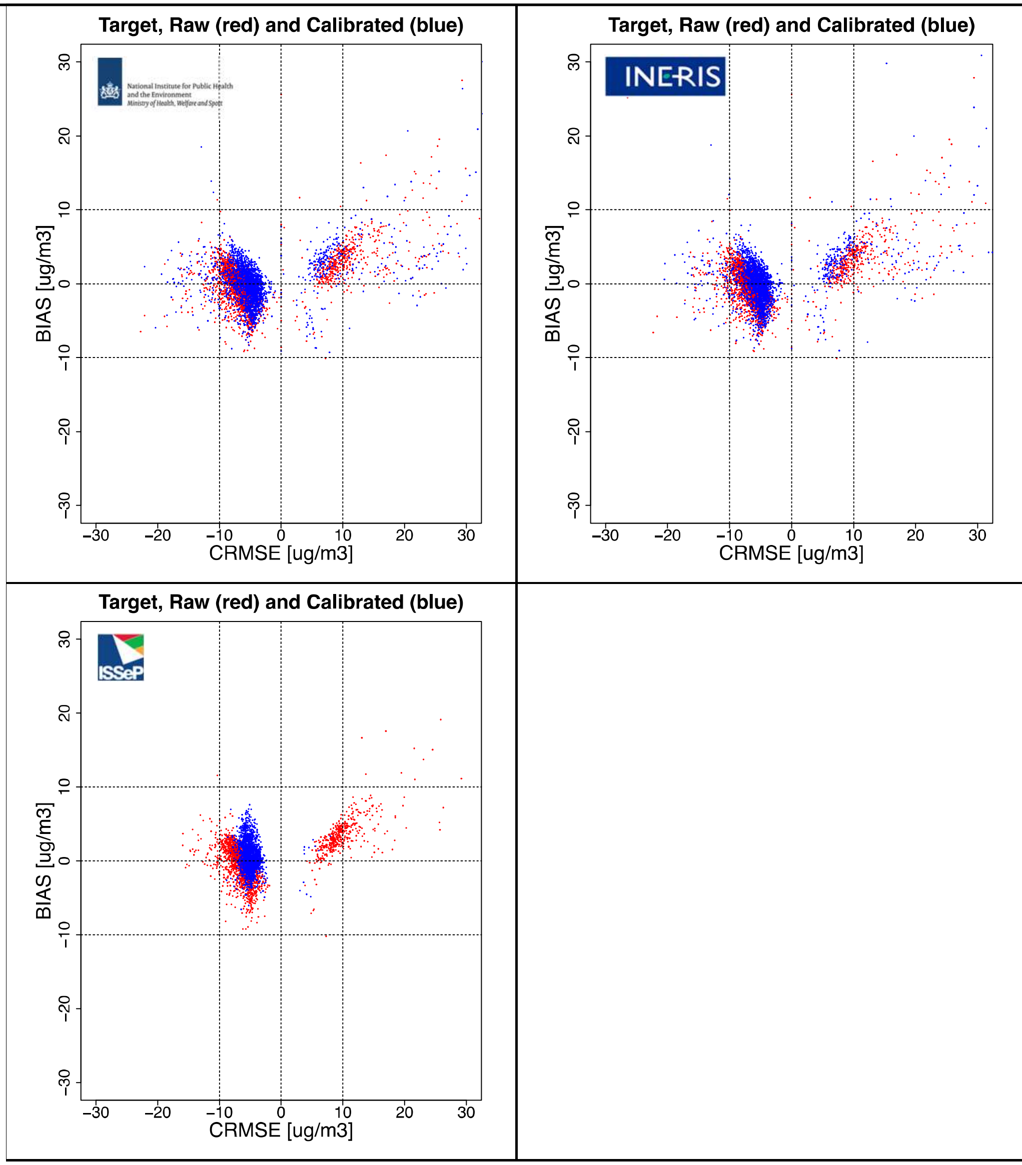


Analysis/visualization

Period Jan 01-31, 2022

Biases of the **monthly averaged** raw PM2.5 data and the averaged calibrated data versus the averaged synthetic real data.

Over the full month, the raw and calibrated concentrations do not differ much.



Analysis/visualization

Period Jan 01-31, 2022

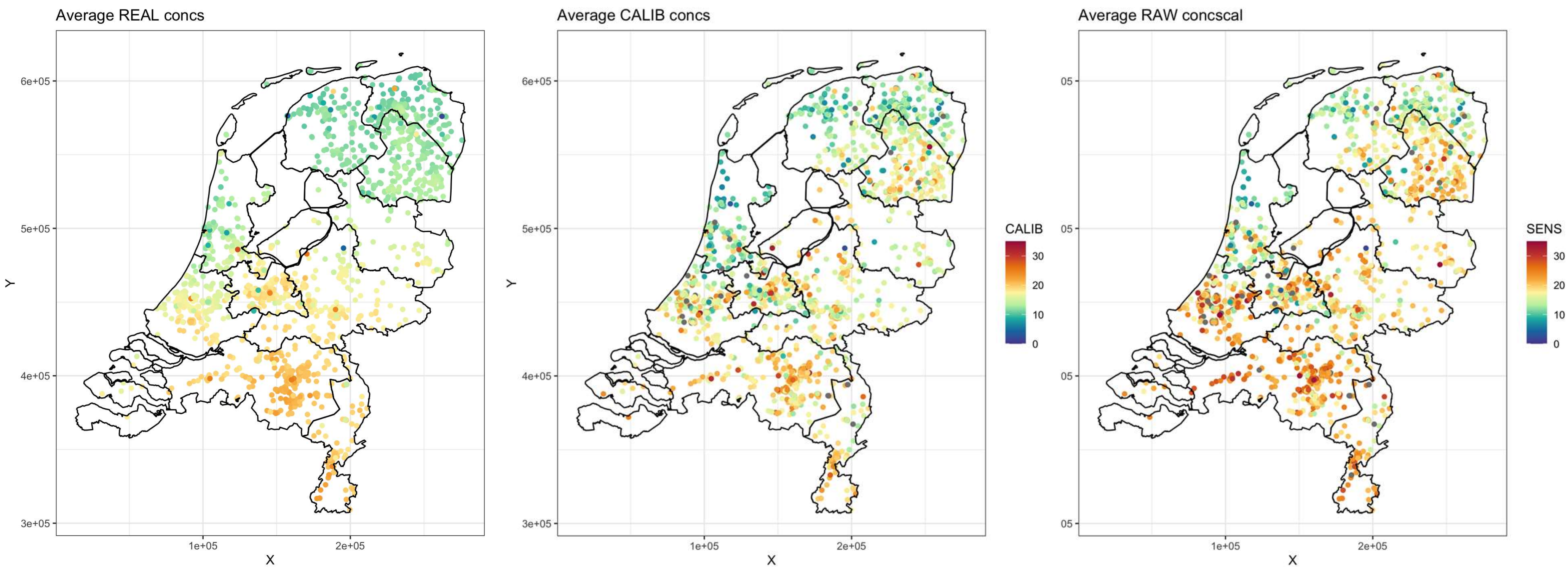
Target plots of the monthly averaged raw data and the averaged calibrated data.

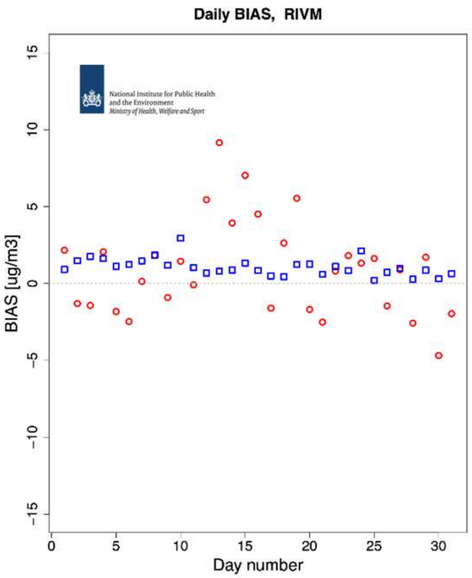
The raw and calibrated data differ in the CRMSE's.

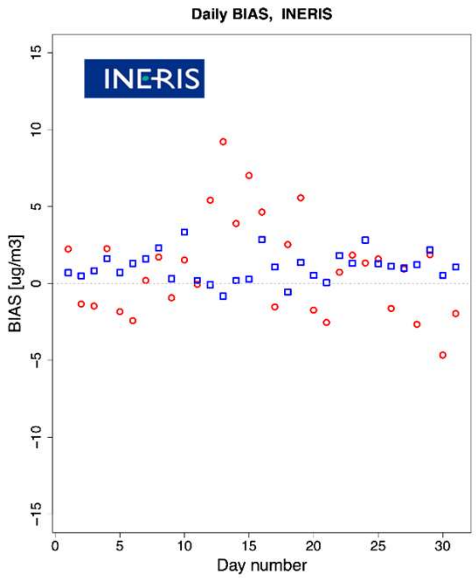
The BIAS and CRMSE are <u>not</u> normalised using the uncertainty of reference PM2.5 measurements.

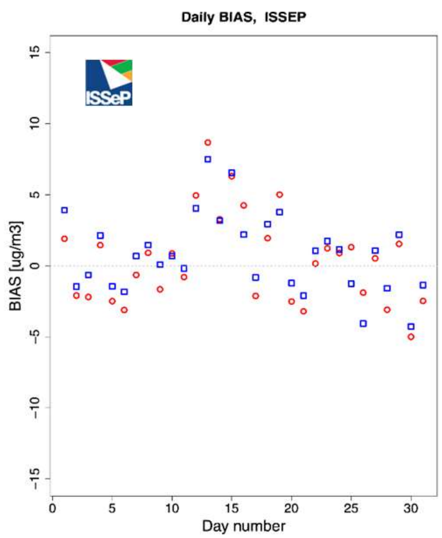
Analysis/visualization …

- **Spatial** effect of calibration in the sub-period 10-19 Jan, 2022







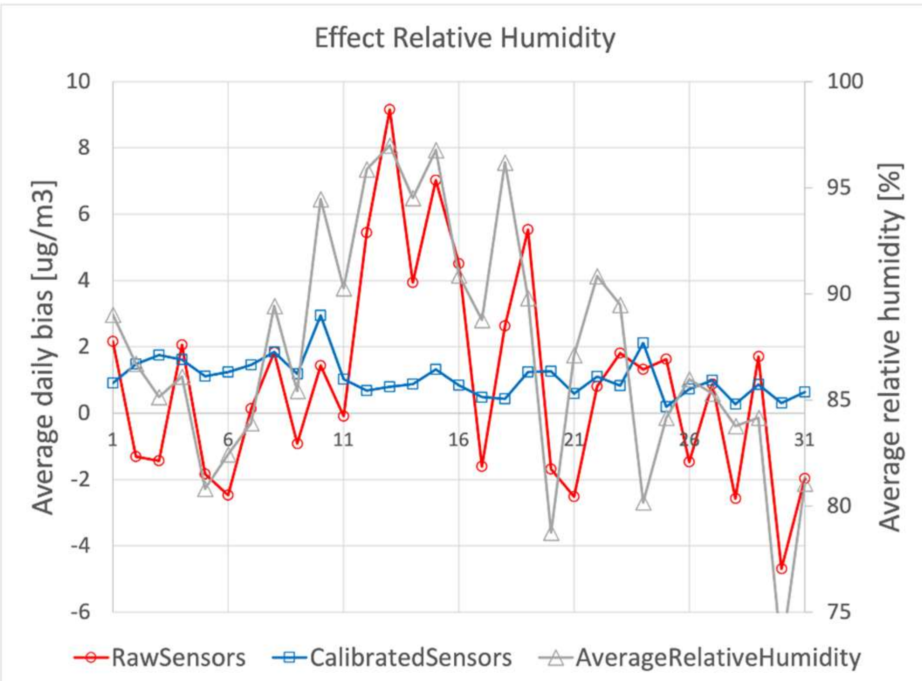


Analysis/visualization

Period Jan 01-31, 2022

Plots of the daily average of all sensors for raw data (red) versus the calibrated data (blue).

Analysis/visualization



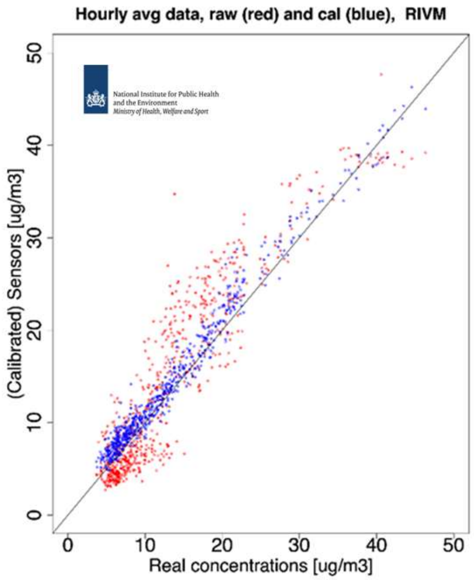
Period Jan 01-31, 2022

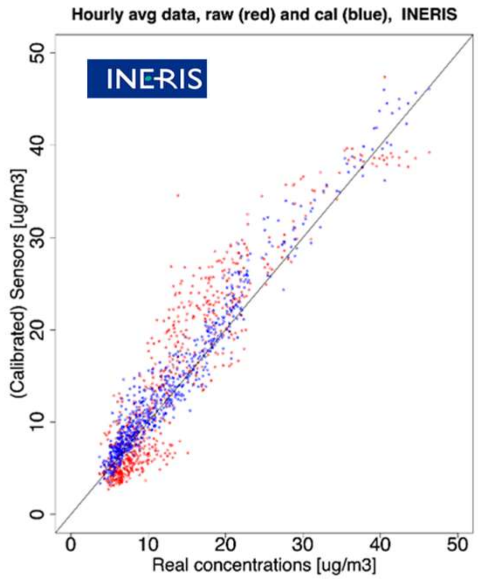
More daily averages for

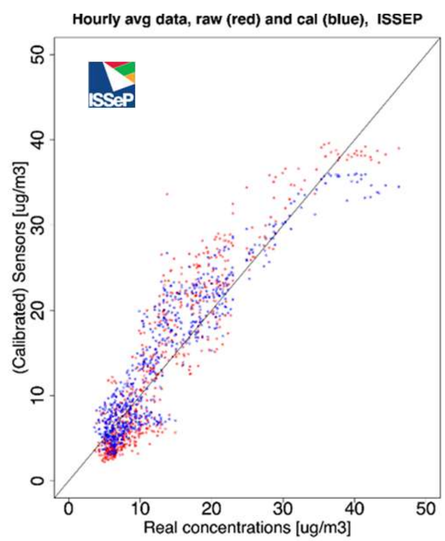
- Bias of raw sensor data
- Bias of calibrated sensors (RIVM method)
- Relative Humidity

Daily bias of raw sensors data and relative humidity show similar behaviour

Analysis/visualization







Period Jan 01-31, 2022

- Scatter plots of the hourly spatial average of all sensors for raw data (red) and the calibrated data (blue).
- Most calibrated values above 1:1 line. Result of including sensors that are not functioning properly and are giving high values.

Conclusions (1)

- Benchmarking is an important process.
- The importance of data cleaning, handling of uncertainty, interpolation and calibration of low-cost sensors is demonstrated and investigated.
- Sufficiently realistic synthetic sensor data can be constructed and these are valuable for an objective test of sensor-processing algorithms.
- 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.

Conclusions (2)

- The results obtained by INERIS and RIVM are quite comparable.
  Based on hourly averaged concentrations, the ISSeP method shows less improvement after calibration. The monthly average BIAS of ISSeP method is better centered around zero. Likely due to discarding more data points.

- 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 has a large random uncertainty that can not be corrected for by network calibration → limits individual use.

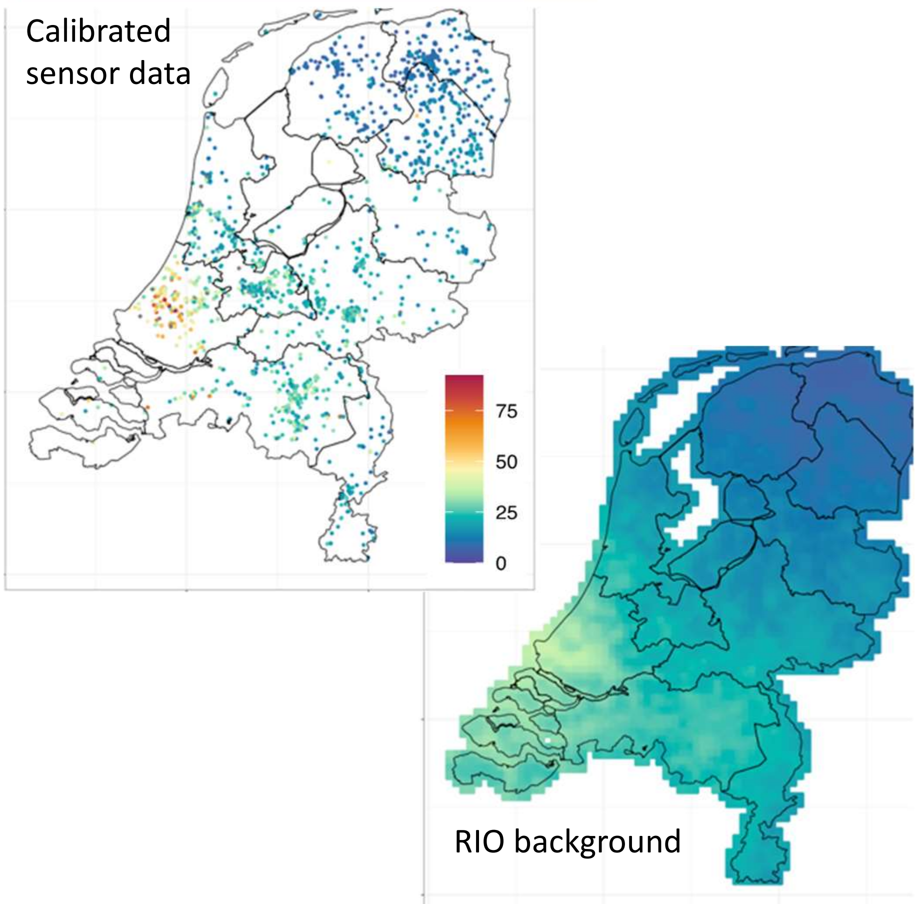
Questions ?

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

Calibration and Data Fusion of PM2.5/PM10 sensors in the Netherlands

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

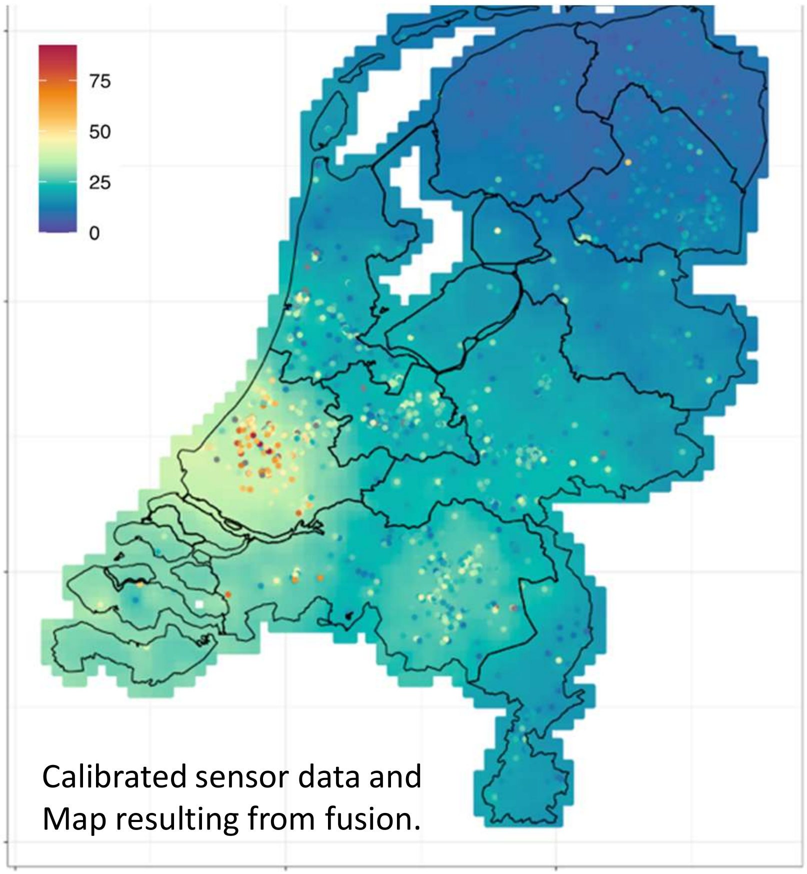
Data Fusion of sensors and maps



- After the calibration of the individual sensors, we have estimated calibrated concentrations for all the sensors.
- The estimated uncertainties are obtained using a bootstrap procedure.
- The sensors field can be extrapolated over all of the country.
- We also have have an official estimate for the hourly background for all of the Netherlands from the RIO model created by VITO. The scale of this background is km^2.

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

Data Fusion of sensors and maps



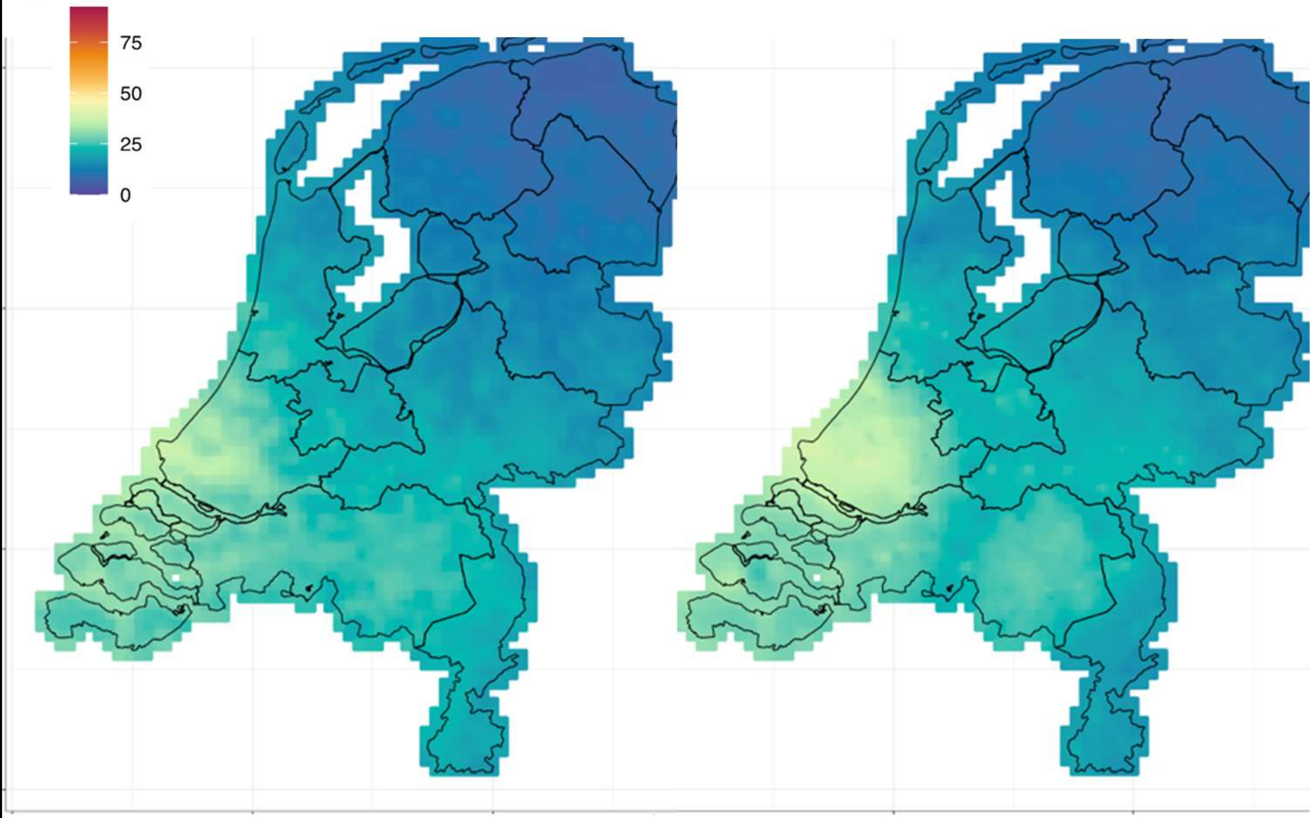
Calibrated sensor data and
Map resulting from fusion.

- Assuming Normal distributed uncertainties in both RIO and the sensor field, we can combine the RIO map and the values of the sensors using inverse variance weighing.
- Better: assume the RIO values as the Bayesian prior *p(ϑ)* for the concentration at a location and the sensor field as likelihood *L(x|ϑ)*.
- Following Bayes, we can then write the combination of results from sensors and RIO as:

$$p(\theta|x) = L(x|\theta)\cdot p(\theta) \,/\, \int L(x|\theta)\cdot p(\theta)d\theta$$

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

Data Fusion of sensors and maps



Original RIO map

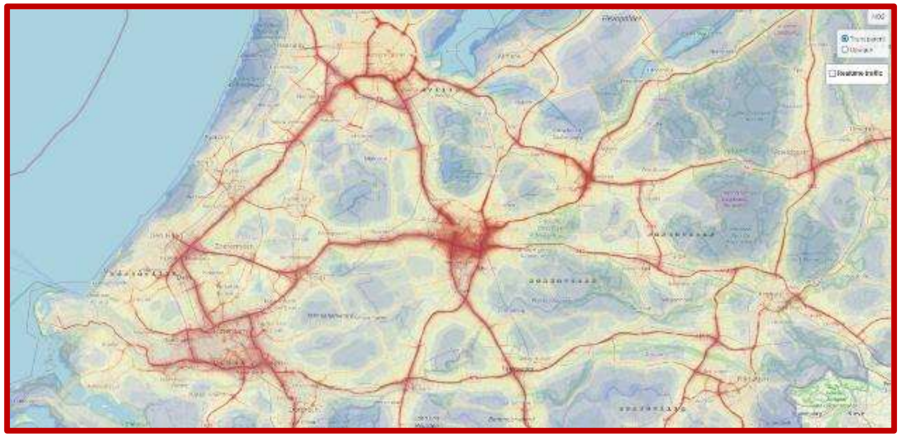
Result of fusion of RIO map and sensors

- After the fusion of RIO and sensors, the new map becomes more similar to the distribution of the sensor values.
- Quite often, the effect of the fusion is a slightly different map than before.
- This indicates that the RIO field already provides a good approximation of the concentration field.

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

NO_2 fusion model

- At this moment implemented using reference data
- Use detailed calculation from dispersion model to calculate an hourly road contribution C_{road}
- Also calculate hourly background concentrations using interpolation model (RIO) C_{bg}
- Define a best estimate $C_{pred} = f_{road} * C_{road} + C_{bg} + \Delta_{bg}$
- f_{road} and Δ_{bg} to be determined from comparing measurements and calculations at ~ 10-20 locations.



- Possible to make local clusters of correction factors instead of two that cover the country?
- Sensor uncertainties sufficiently small for a fusion approach?

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

Questions ?

WG6 Benchmarking
Next steps

Technical Meeting October, 2023

WG6 next steps

- **Collect assimilation/fusion models** that **combine calculations** of urban concentrations on hourly/daily timescale **with low-cost sensor data**.

- Maybe you have such a model for reference measurements and want to test its usefulness for sensor measurements)

- Can your model run outside your outside your usual national/urban use case ?

WG6 next steps

Additional sensor data sets?

- First part of WG6 delivered
  - Method for creating synthetic sensor data. In the Netherlands.
  - Data set for synthetic real concentrations, synthetic reference measurements and synthetic sensor data
  - Propose to continue working with (Dutch) synthetic data → **synthetic model results**

- Apply lessons learned to **new sensor data** sets
  - Searching for 1 or 2 additional sensor data sets (mail to WG6 will follow)
  - Go from national to urban scale. → Increases availability of sensor data sets

FAIRMODE

WG6 next steps

- Aim to have well defined cases by the start of 2024

- Focus on PM2.5, NO_2 (near future). Most common compounds in air quality models.

- Please let us know if you can provide urban sensor measurements and/or data fusion model calculations (sjoerd.van.ratingen@rivm.nl).

- You will be **invited** for a **kickoff** meeting where we can decide on optimal (urban) use cases the different assimilation models.

Sensor/model fusion: Example.

- Example fusion of RIO model with sensor data

In:
Current hourly reference measurements

RIO interpolation model (based on historical reference measurements and land cover data)

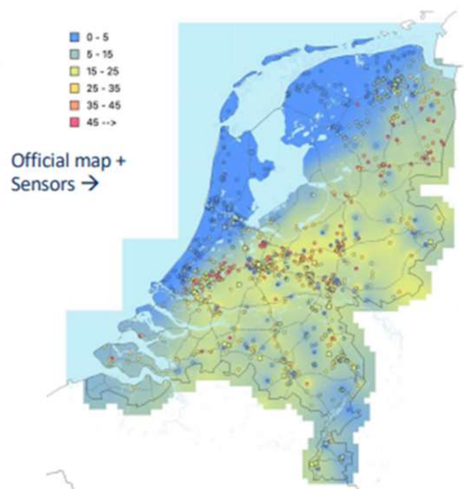
Model result:
Interpolated concentrations
At requested locations.
Rectangular 4x4 km grid

In:
Sensor measurements at 2000+ locations

Data fusion of PM2.5 map and sensor data

Fusion result:
PM2.5 concentration map at 1x1 km grid

PM2.5 concentration map



Data fusion of PM2.5 concentration map and sensor data

Sensor/model fusion: How to compare?

How to compare fusion outcomes based on calibrated sensors.

Brainstorm: Please interrupt / share your ideas!

1. Benchmark for calibration was done using **synthetic**
   - Real concentrations (truth)
   - Official concentrations
   - Sensor values

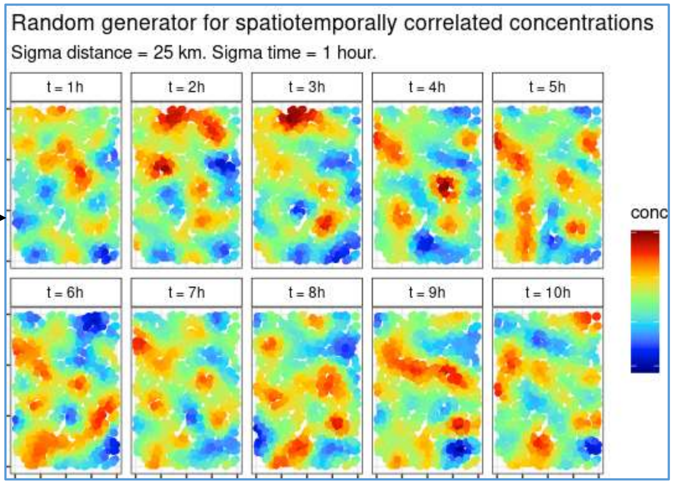
2. Leave one out validation

3. Compare outcomes of different data-fusion methods, without validating.

Sensor/model fusion: How to compare?

Suggestion 1: Continue using synthetic data in data-fusion

- Generate synthetic truth, reference and sensor data.
  Should in principle also be possible for other data sets, besides the previously used Dutch sensor network
- Add model results:
  - The calibrated sensor data **and the model** have to be related to the synthetic truth
  - The synthetic truth was (to contain plausible spatial correlation) derived from model results
  - Use same model that was used in generating synthetic truth
  - Synthetic model by interpolating synthetic reference data
  - Synthetic model by distort output of model.
    E.g., generate random (spatially correlated) distortions
  - Distort input (Land use, emission , meteo) of model
  - Combine distorted model and synthetic (sensor) measurements by **different** data fusion methods



Sensor/model fusion: How to compare?

Suggestion 2: Use leave-one-out validation

- Use real sensor data
- Use reference station for validation
- Leave **one reference** station out
- Reference stations are also used in calibrating the sensors.
  - Leaving out a reference station will affect calibration as well as fusion model.
  - How many reference stations left when going to e.g., urban scale?
- Model should be able to calculate concentration at validation location
- Consider whether model is suited for background locations or also street level locations in choice of validation location.

Sensor/model fusion: How to compare?

Suggestion 3: Compare outcomes of varying data-fusion methods, without validating

- No regret option
- Compare data-fusion outcomes to each other
- Use **different** calculation models and different fusion method (confusing…)
- Use **same** calculation models and different fusion method

Pittfalls?

- Model **too** good. No improvement by sensors.
- ….. ?

Summing up

- Please let us know if you have an interesting (urban) sensor set to share.
- Are you currently using a specific sensor-model fusion method ?
- Would you like to know how this method compares to others ?
- Are you interested in actively participating in a data fusion benchmarking exercise?

sjoerd.van.ratingen@rivm.nl

Questions ?