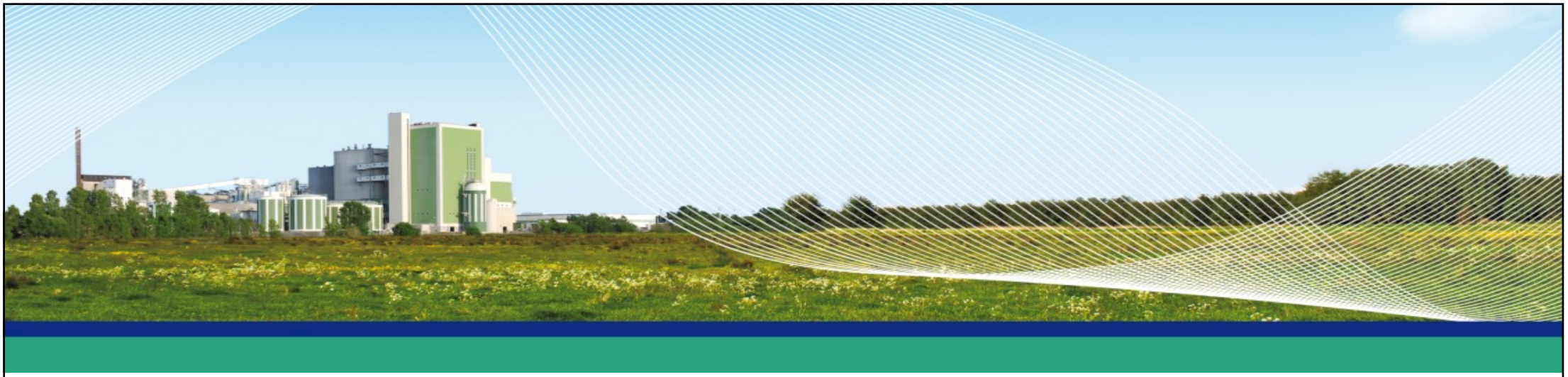


FAIRMODE
CT6: Near real time assessment with low-cost sensors

Processing and calibration of sensor network data

Alicia Gressent (alicia.gressent@ineris.fr)

INERIS
maîtriser le risque pour un développement durable

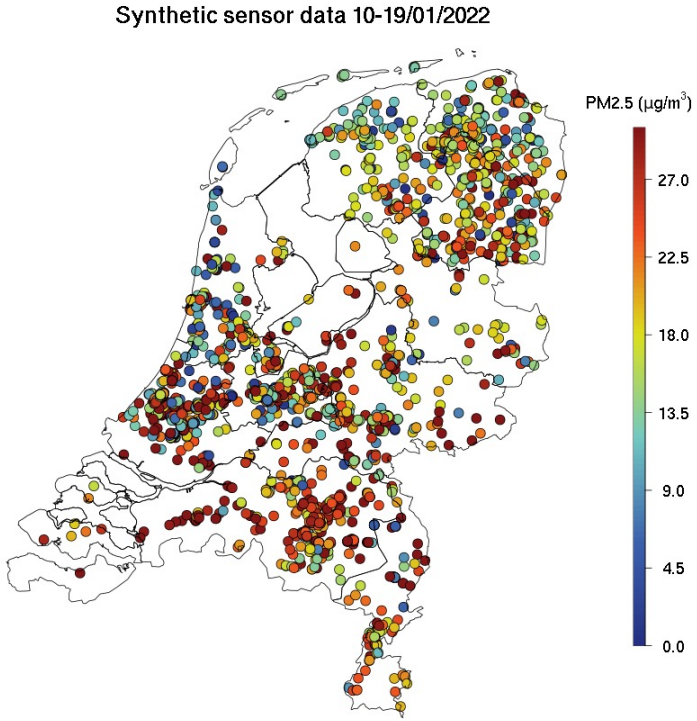
Fairmode Technical meeting – October 18-20, 2022.

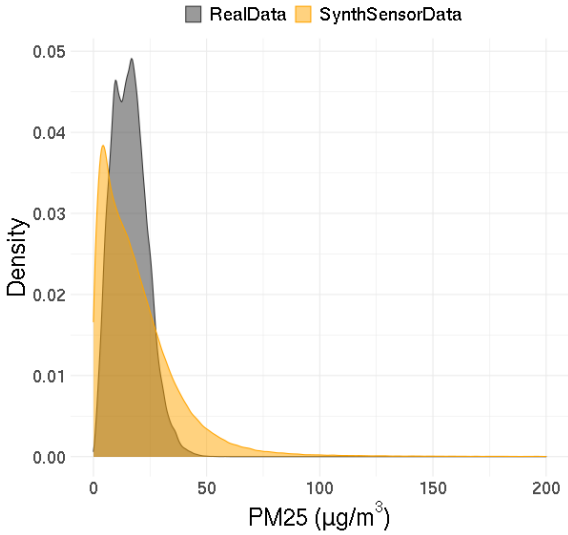
Dataset

Netherlands

Synthetic data

- Full synthetic sensor data for January 10-19, 2022

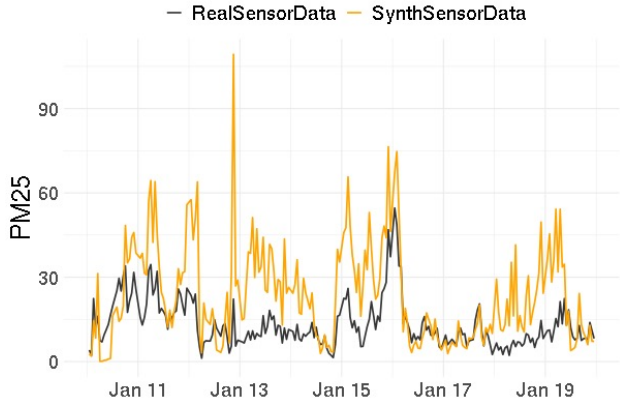




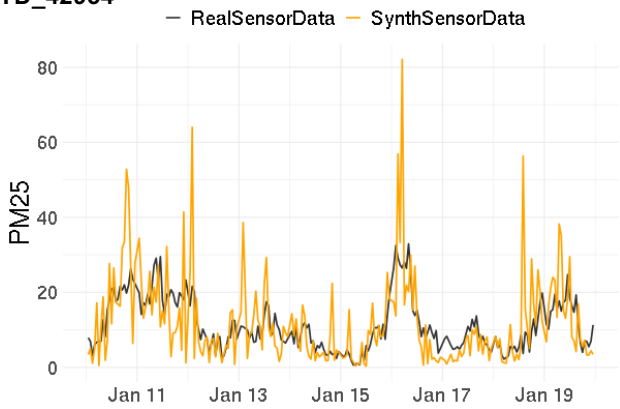
- Real synthetic data at sensor positions
- Full synthetic sensor data

➢ 464561 observations / 2205 sensors

HLL_hl_device199



LTD_42954



INERIS
maîtriser le risque pour un développement durable

Data processing

Outlier detection

Data classification and definition of an interval of validity

Calibration

Calibration factor / RIVM adjusted approach

Performance assessment

Data fusion

SESAM (data fusion with SEnSors for Air quality Mapping)

Outlier detection

Method

Data cleaning

Eliminate negative values
Eliminate values > 2 x (max value of reference stations)
Eliminate frozen concentrations for several hours and days (≥ 3 consecutive hours)
Eliminate sensor with constant positive bias

Clustering & classification

Create groups of data depending on sensor clusters (the nearest neighbors), site typology and season

Outliers' detection

Apply *van Zoest et al., 2018* outliers' detection methodology initially applied to NO_2 in urban areas and adapt to $PM_{2.5}$ at national level.

For each group of data:

Calculate a validity intervals of the data: $\mu \pm z \times \sigma$

Root square transformation that gives a truncated normal distribution, then optimization of a likelihood function to get the mean and the standard deviation of the underlying normal distribution for each observation.

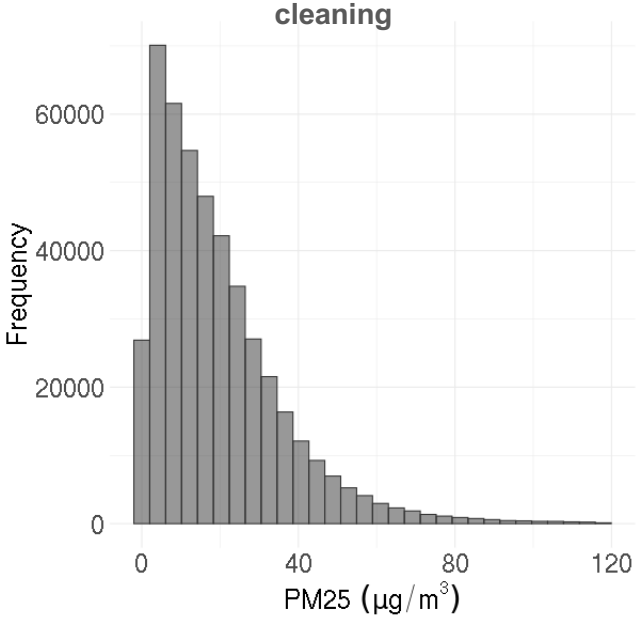
Eliminate values that do not fall within the confidence interval

Outlier detection

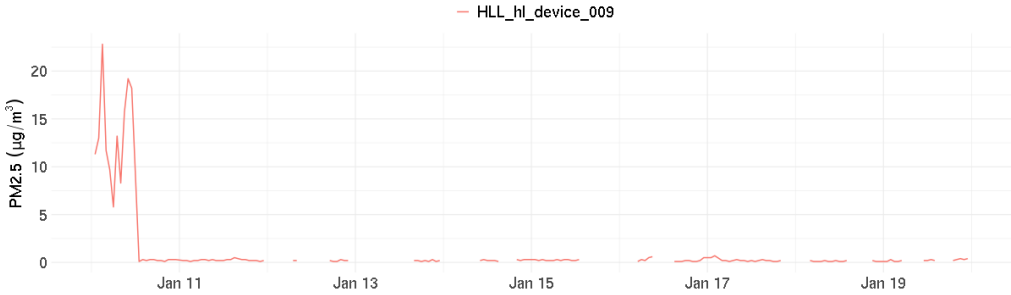
Data cleaning

1. Eliminate negative values
2. Eliminate values > 2 x (max value of reference stations)
3. Eliminate frozen concentrations for several hours and days (≥ 3 consecutive hours)
4. Eliminate sensor with constant positive bias

Distribution of full synthetic sensor data after cleaning



Example of frozen values detection



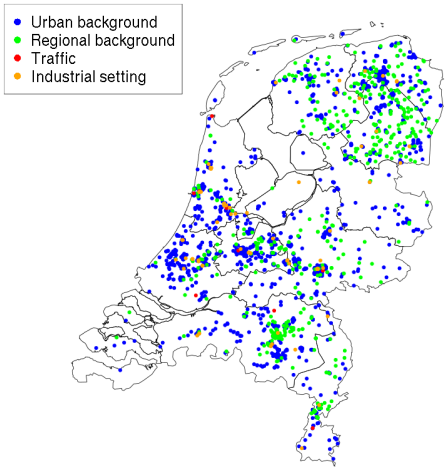
Clean synthetic data from 10/01/2022 to 19/01/2022:

➢ **2% of data is eliminated**

Outlier detection

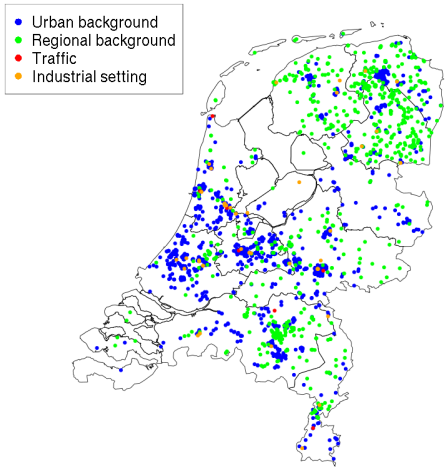
Sensor type

1) Assign typo based on Corine Land Cover data (land use)



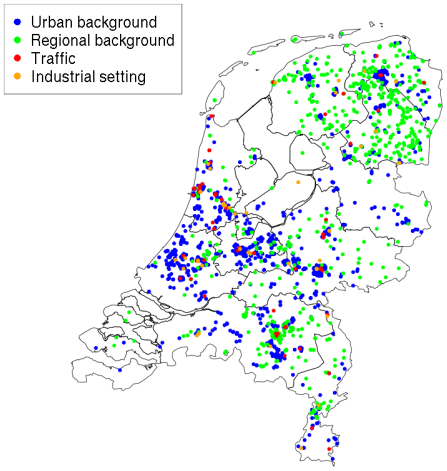
- Typology are assigned to the CLC classes.
- CLC information is extracted within a buffer of 1m around the sensor location.
- Typology is assigned to sensors

2) Adjust typo based on population density



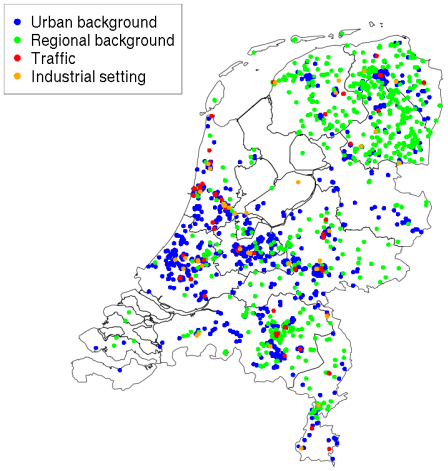
- Population density information is extracted at the sensor location.
- Typology is assigned to the sensor depending on the extracted information.

3) Adjust typo based on road network



- Road information is extracted within a buffer of 5m around the sensor location.
- Traffic typology is assigned to the sensor within the buffer.

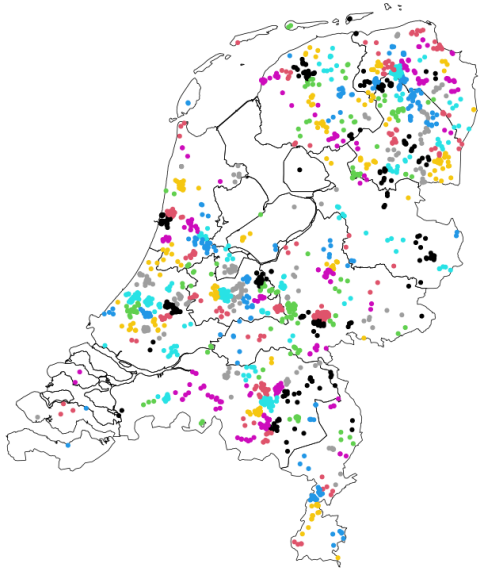
4) Adjust to station typo when in the vicinity of the sensors



Outlier detection

Data classification

Clusters
(the nearest neighbors)



Sensor Typology

- Industrial settings
- Regional background
- Urban background
- Traffic

Season

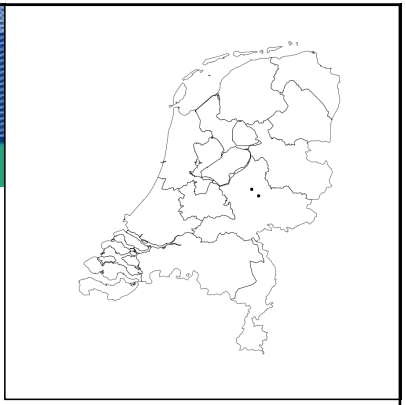
- Winter (DJF)
- Spring (MAM)
- Summer (JJA)
- Fall (SON)

Define clusters of sensors (the nearest neighbors ~ 10km)
251 clusters

Outlier detection

Outlier detection

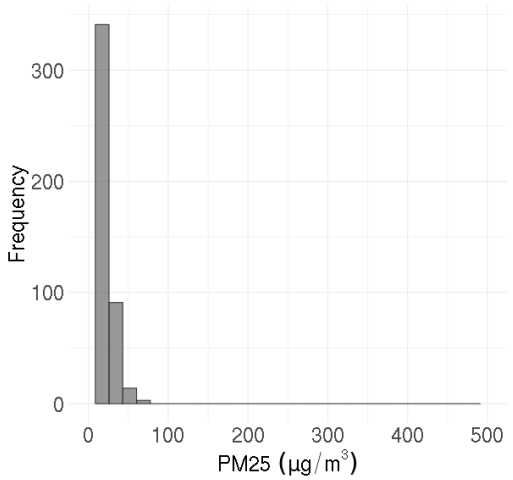
- Define a confidence interval and identify outliers for each group of data



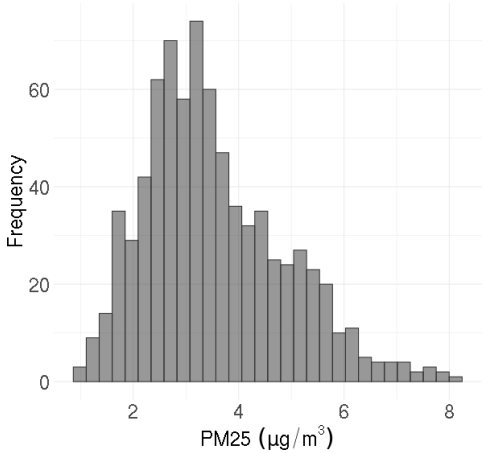
Example for a group of sensors:

Regional background / winter / cluster N°6

Selection of a cluster



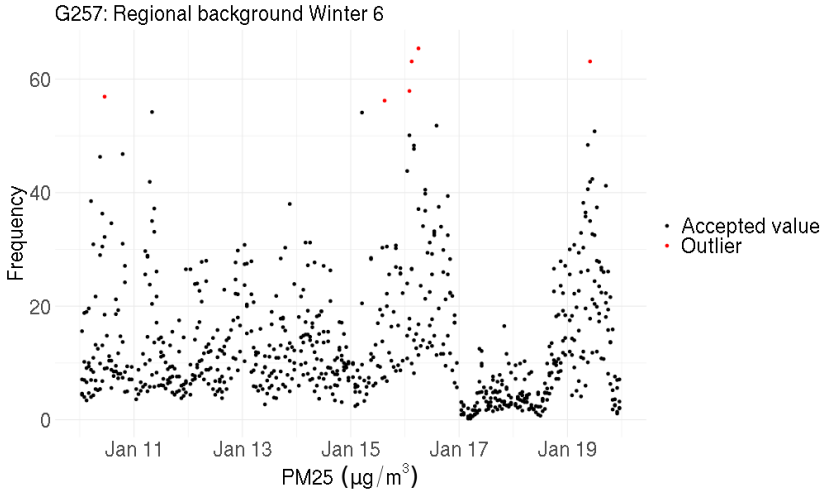
Log normal transformation



Definition of a validity interval

$\Leftrightarrow \mu \pm z \times \sigma$

Identification and elimination of outliers



INERIS
maîtriser le risque pour un développement durable

Calibration

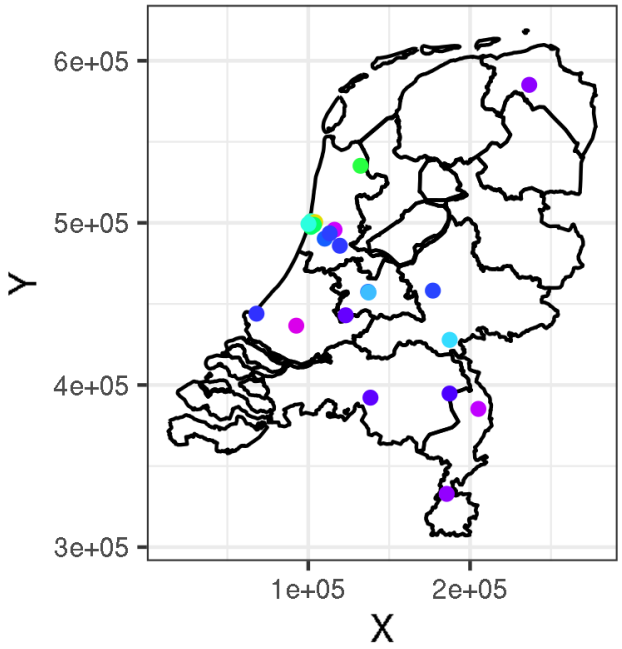
Calibration factor

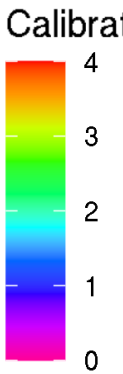
Application of the RIVM adjusted methodology:

- Select sensors in the vicinity of reference stations (distance defined depending on the representativeness of the station), calculation and interpolation of the factor of correction

2021-01-12 14:00:00

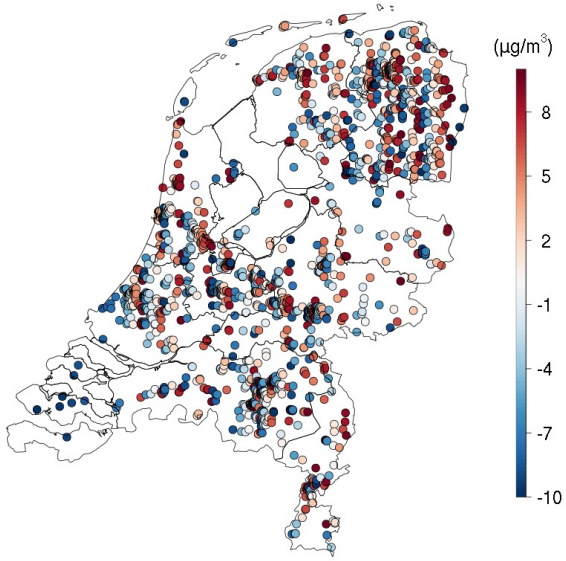
Calibration factors at reference stations





Application of factors on sensor observations

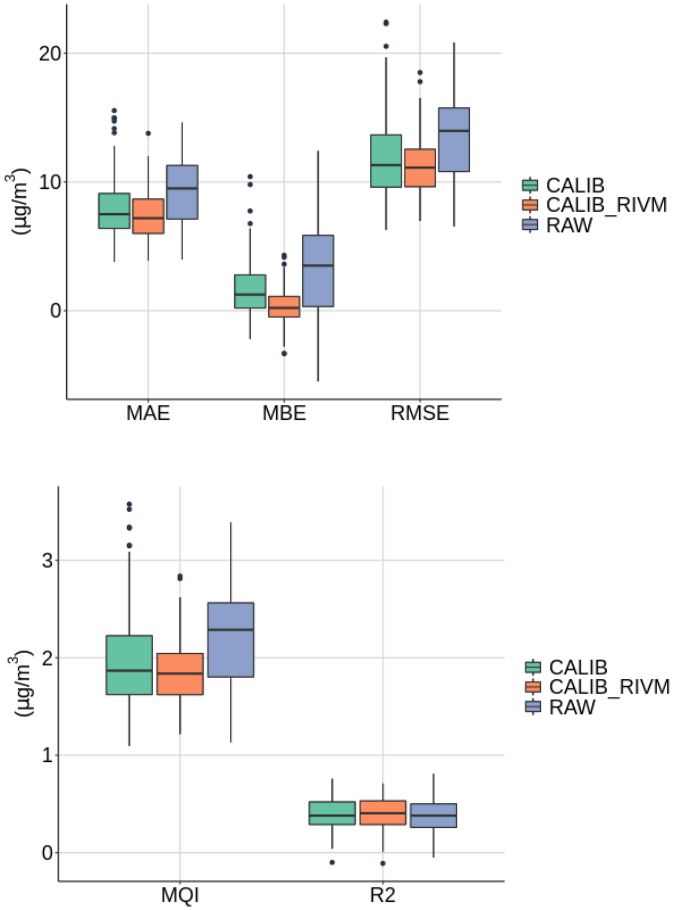
Impact of factors on sensor observations



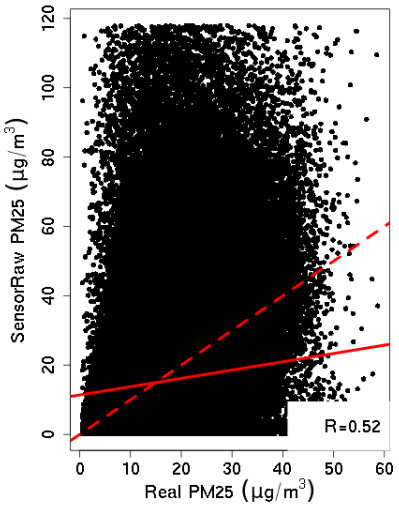
Calibrated – raw sensor observations

Calibration performance

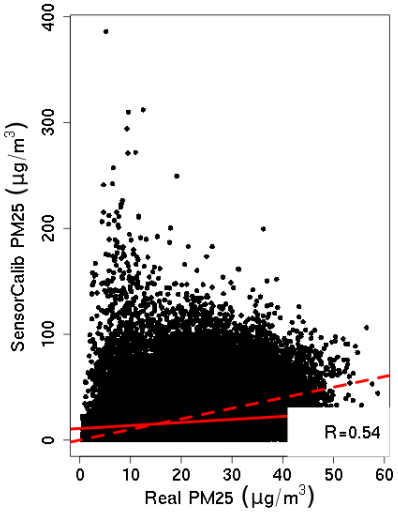
Comparison between raw synthetic sensor data and calibrated synthetic sensor data



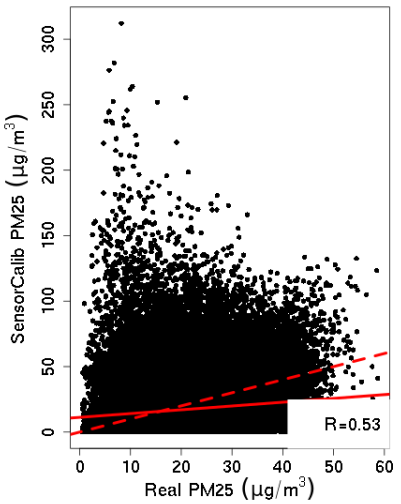
Synthetic real data vs. full synthetic sensor data



Raw full synthetic sensor data



Full synthetic sensor data calibrated with Ineris method



Full synthetic sensor data calibrated with RIVM method

Summary

Contribution to the CT6:

- Test Ineris outliers' detection and calibration procedures on data from a sensor network on a national scale
- Put Ineris method into perspective according to other approaches (RIVM & ISSeP) / benchmarking
- Development of recommendations (next step)

INERIS approach:

1. **Outlier detection**

   Clustering of sensor data

   Definition of an interval of validity / eliminate values that do not fall within the confidence interval

2. **Calibration**

   Calculation of calibration factors based on RIVM approach (adjusted depending on the representativeness of reference stations)

   → To be applied for a longer period

3. **Data fusion – use of the SESAM** tool: data fusion with SEnSors for Air quality Mapping (next step)