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

Alicia Gressent (alicia.gressent@ineris.fr)

Fairmode Technical meeting, October 6-8, 2021.

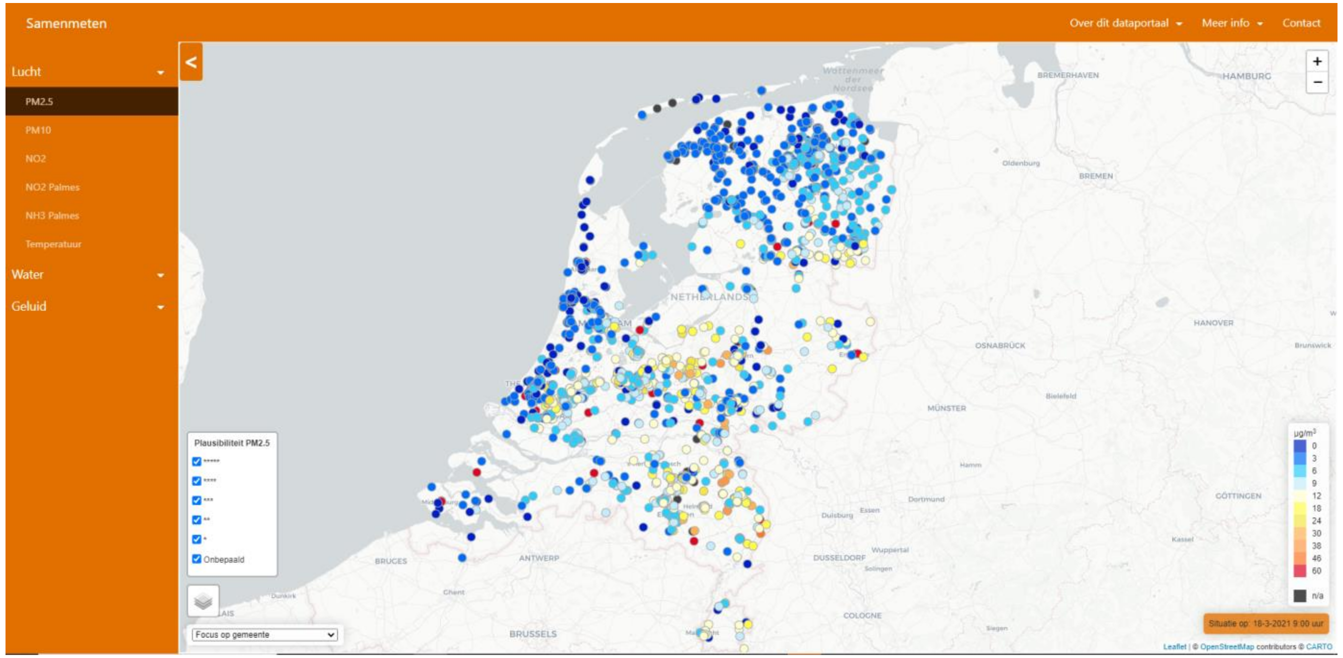
INERIS
maîtriser le risque
pour un développement durable

Dataset

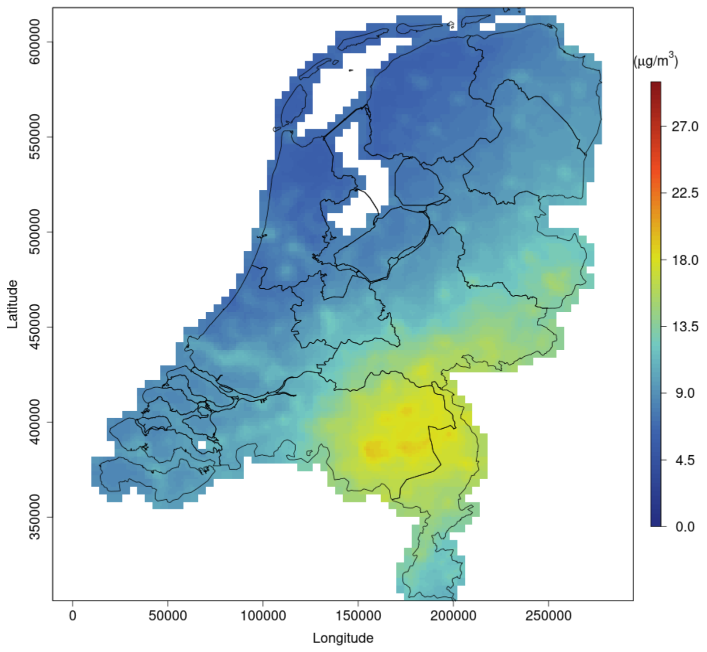
Netherlands

Observation data:

- Reference station measurements of $PM_{2.5}$ concentrations
- Fixed sensors $PM_{2.5}$ data from 18/01/2021 to present



RIO PM2.5 18/01/2021 11am



Model data:

- RIO estimate for the $PM_{2.5}$ concentrations in all of the Netherlands on a 1x1 km^2 grid and the estimated uncertainty in this concentration, both in $\mu g/m^3$

Data processing

Outlier detection

Data classification and definition of an interval of validity

Calibration

Calibration factor / RIVM approach

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 hours)

Eliminate sensor with constant positive bias

Clustering & classification

Create groups of data depending on sensor clusters (the nearest neighbors), site typology and season (> 4000 groups of data)

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: $\mu \pm z \times \sigma$

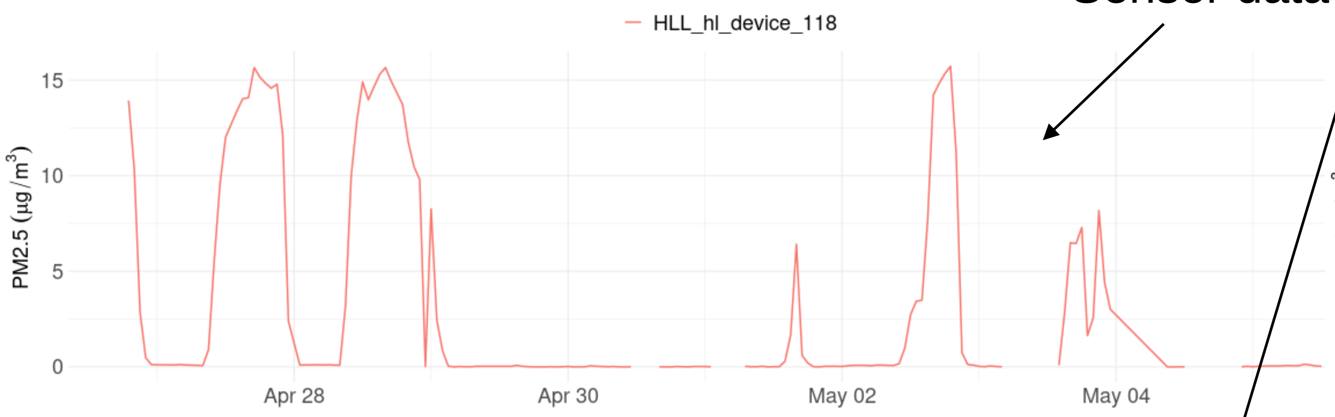
Root square transformation that gives a truncated normal distribution, then optimization of a likelihood function to get the mean and the standard deviation of the normal distribution for each observation.

Eliminate values that do not fall within the confidence interval

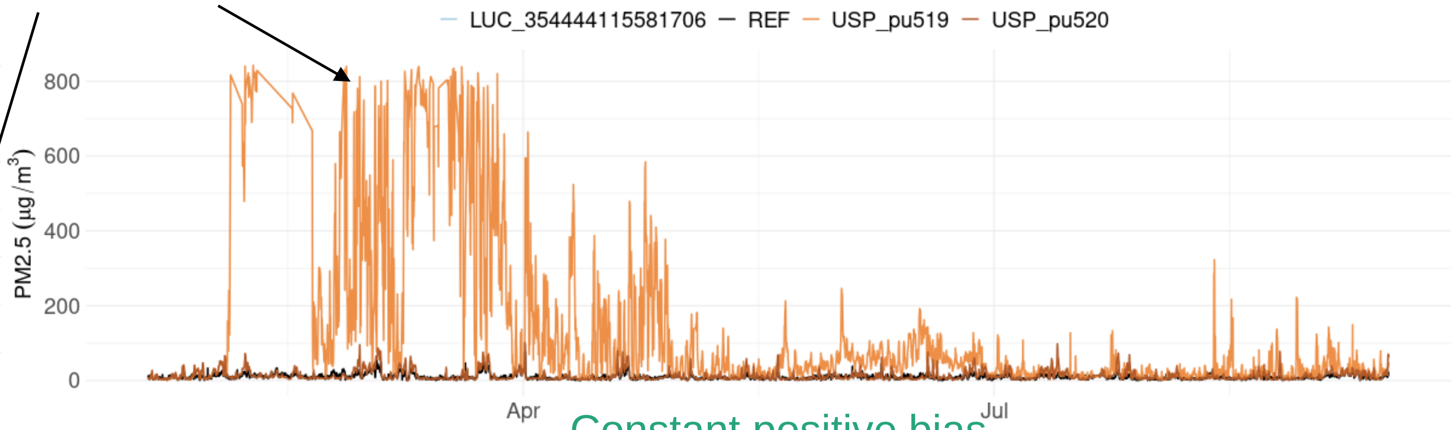
Data cleaning

Example of data cleaning

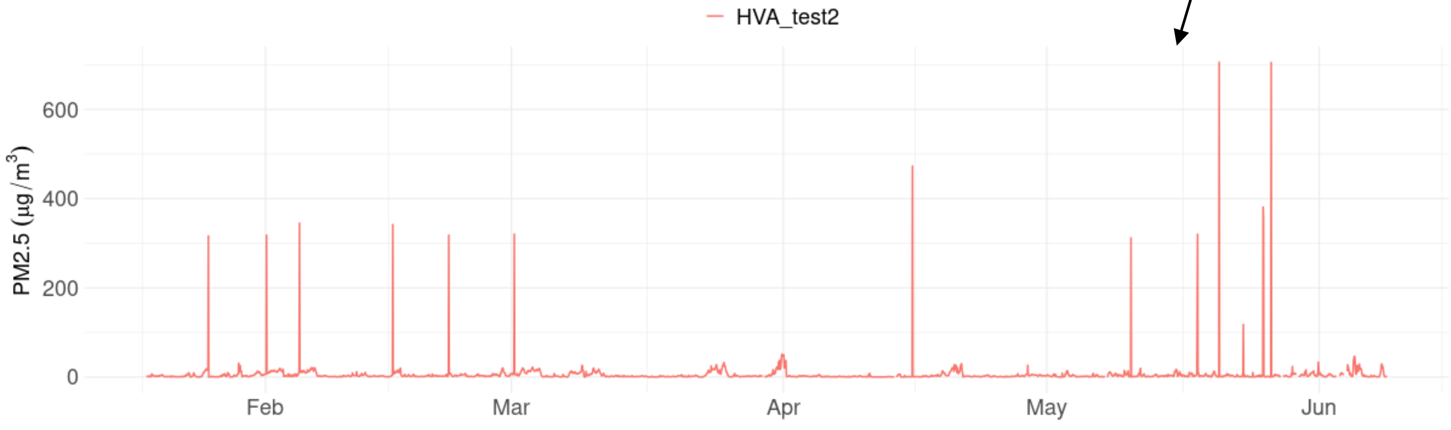
Sensor data to be removed



Series of frozen values



Constant positive bias



Very high peaks

Clean data from 18/01/2021 to 15/09/2021:

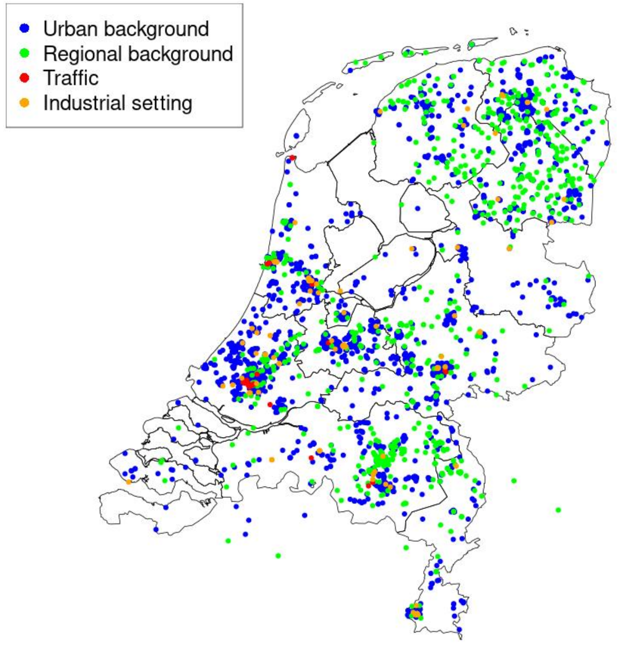
- **1^{e7} data**
- **3598 sensors**

maîtriser le risque pour un développement durable

Outlier detection

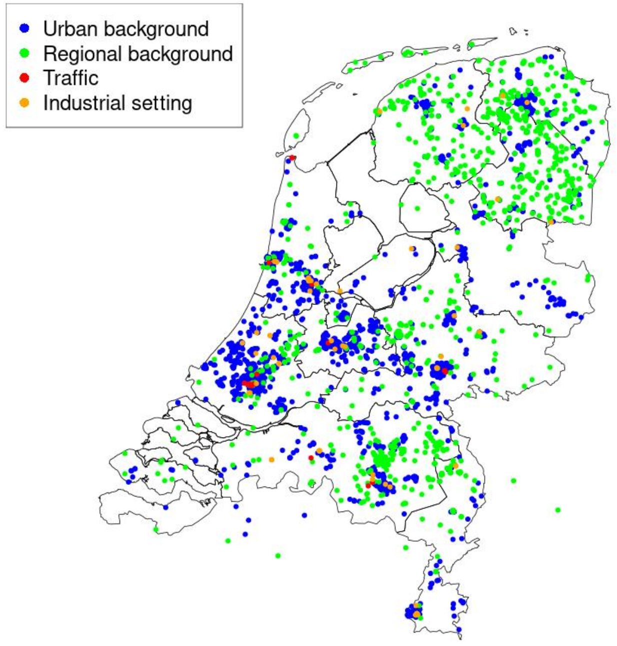
Sensor type

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



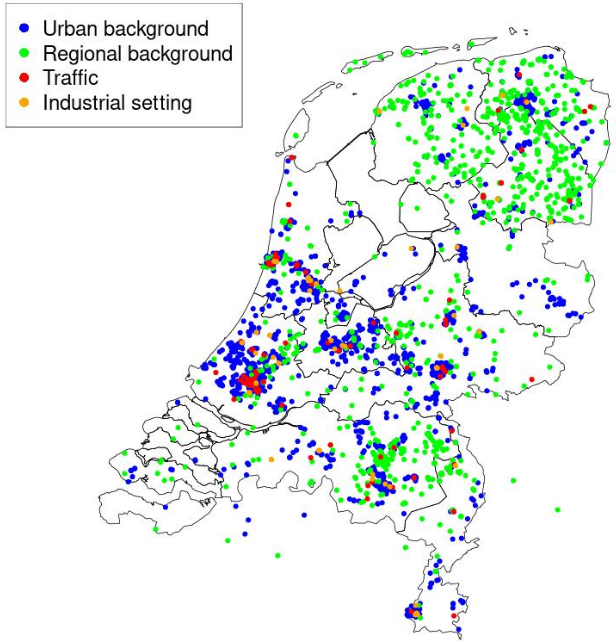
- Typology are assigned to the CLC classes.
- CLC information is extracted within a buffer of 1m around the sensor location.
- Typology is assigned to sensors

2) Adjust typo based on population density



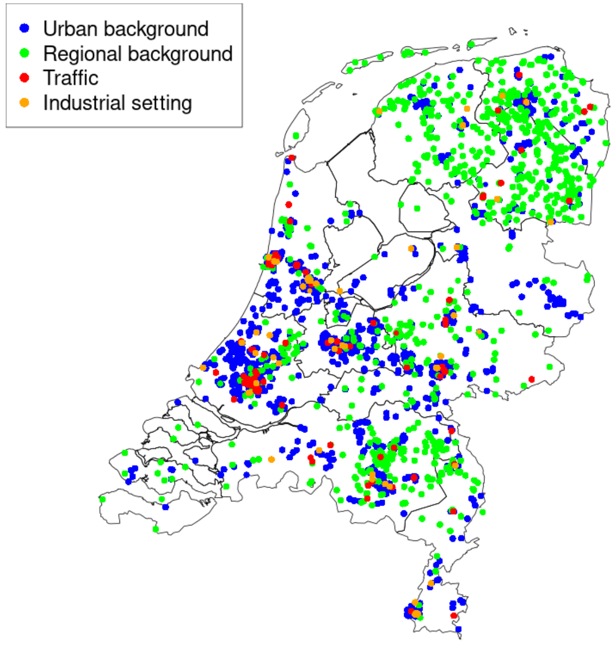
- Population density information is extracted at the sensor location.
- Typology is assigned to the sensor depending on the extracted information.

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

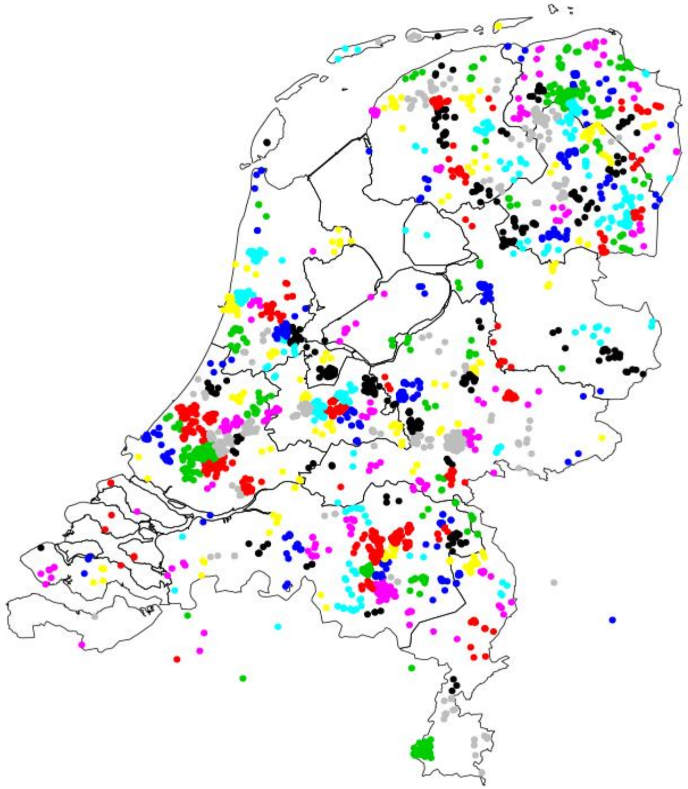


maîtriser le risque
pour un développement durable

Outlier detection

Data classification

Clusters
(the nearest neighbors)



Define clusters of sensors (the nearest neighbors ~ 10km)
→ 284 clusters

×

Sensor Typology

- Industrial settings
- Regional background
- Urban background
- Traffic

×

Season

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

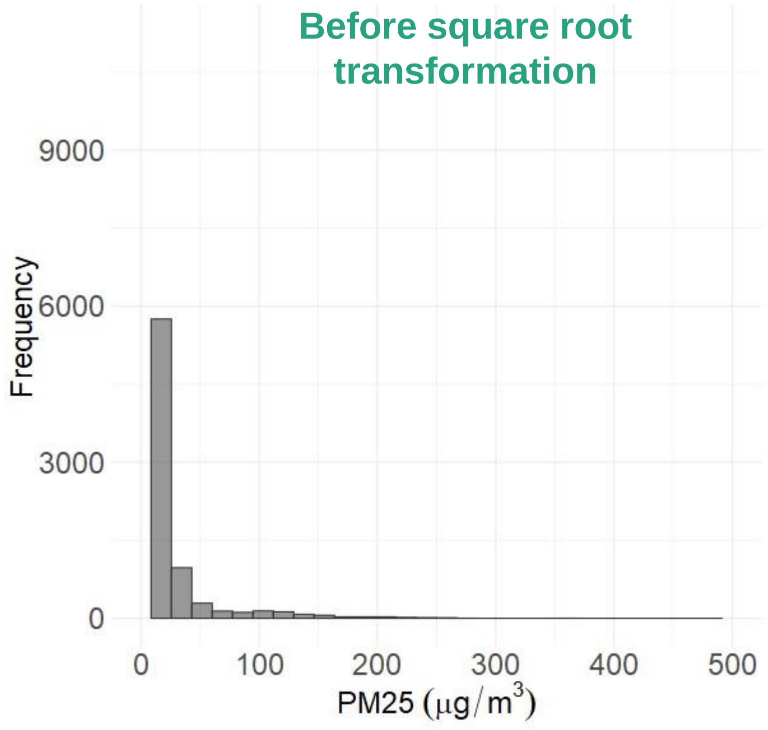
INERIS
maîtriser le risque
pour un développement durable

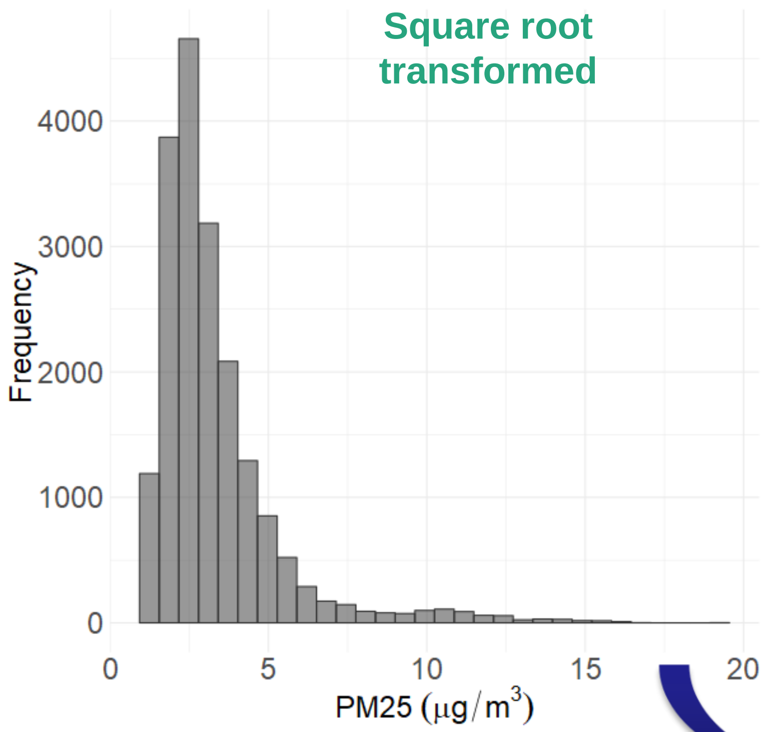
Outlier detection

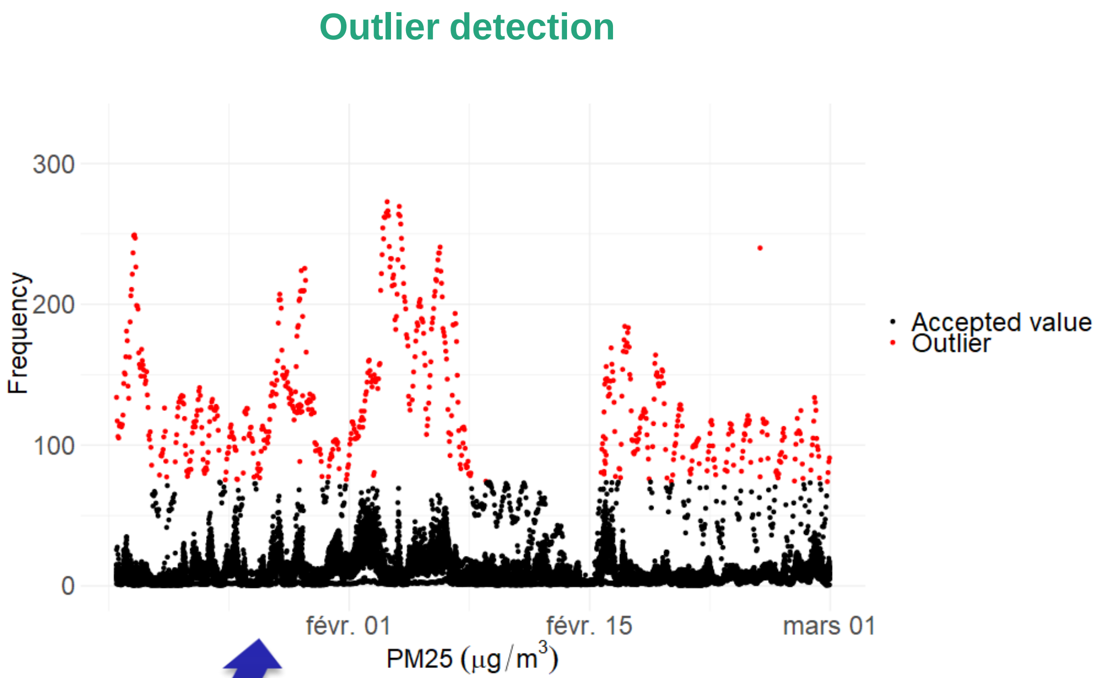
Outlier detection

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

Group Urban background Winter Cluster N°1







Validity interval ⇔
$$\mu \pm z \times \sigma$$

INERIS
maîtriser le risque
pour un développement durable

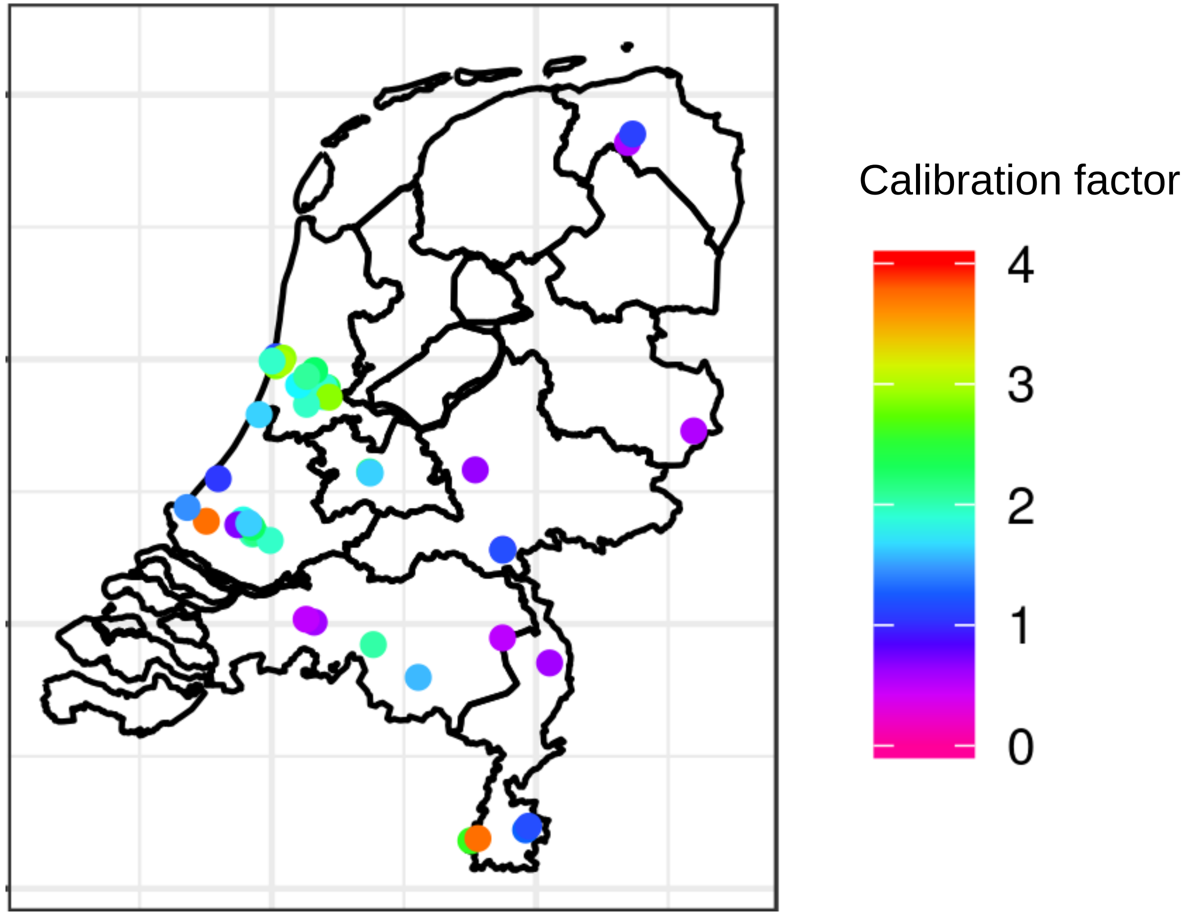
Calibration

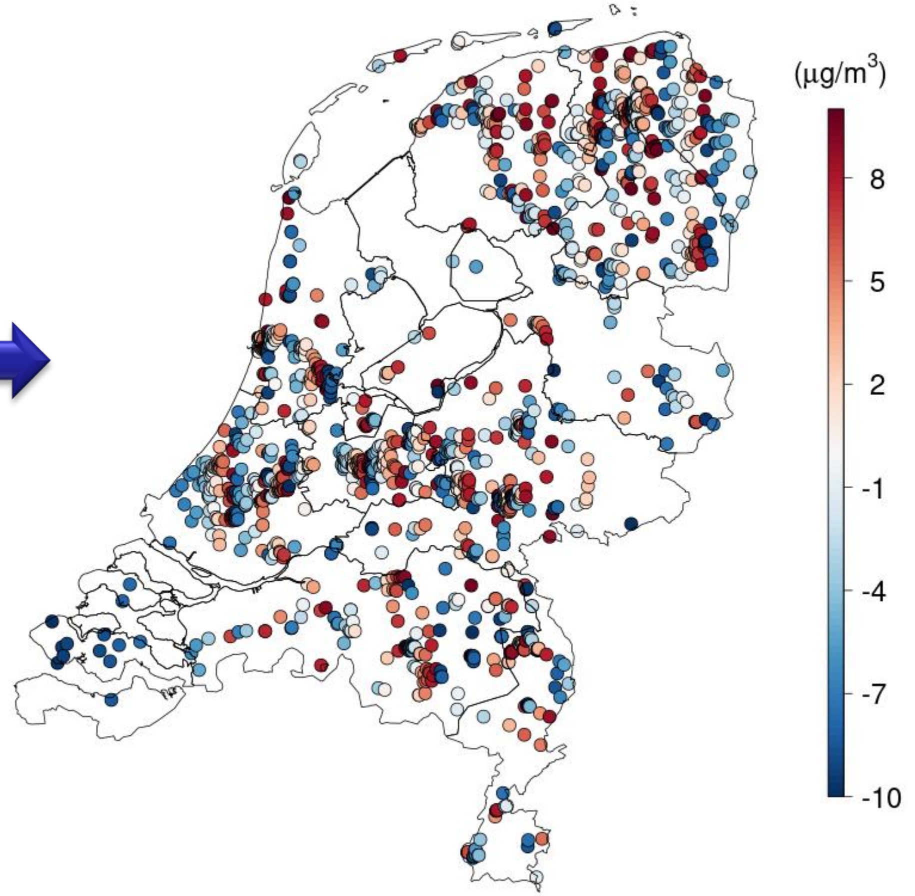
Calibration factor

Application of the RIVM methodology:

- Select sensors in the vicinity of reference stations, calculation and interpolation of the factor of correction

2021-01-18 11:00:00





Impact of calibration on sensor data

Data fusion for AQ mapping

SESAM (data fusion with **SE**n**S**or for **A**ir quality **M**apping**)**

ELSEVIER
Environment International
Volume 143, October 2020, 105965

Data fusion for air quality mapping using low-cost sensor observations: Feasibility and added-value

Alicia Gressent [a], Laure Malherbe [a], Augustin Colette [a], Hugo Rollin [a], Romain Scimia [b]

https://github.com/AliciaGressent/SESAM

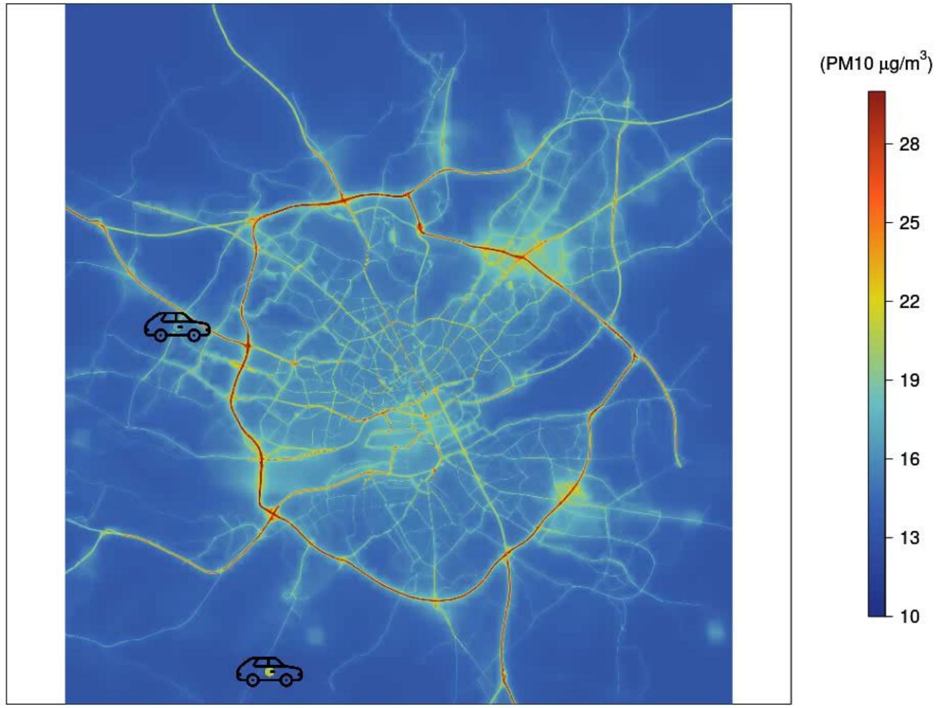
Geostatistical approach → **universal kriging with an external drift**

- ➢ **Merge fixed and mobile sensor data with model outputs at hourly resolution**

Take into account uncertainty and variability of sensor data by introducing the **Variance of Measurement Errors (VME):**

$$\textbf{VME}= \left[\left(\frac{\sigma}{\sqrt{N}}\right)^2 + \frac{{\nu_r}^2}{N}\sum_{j=2}^{N}(C_j)^2\right]_i$$

- σ is the standard deviation of the pollutant observations at the position i;
- N is the number of observations at the position i;
- ν_r is the constant relative type uncertainty (which depends on the type of sensor: **50% fixed sensor observations** and **75% mobile sensor observations**);
- C_j is the j^th pollutant concentration at the position i.



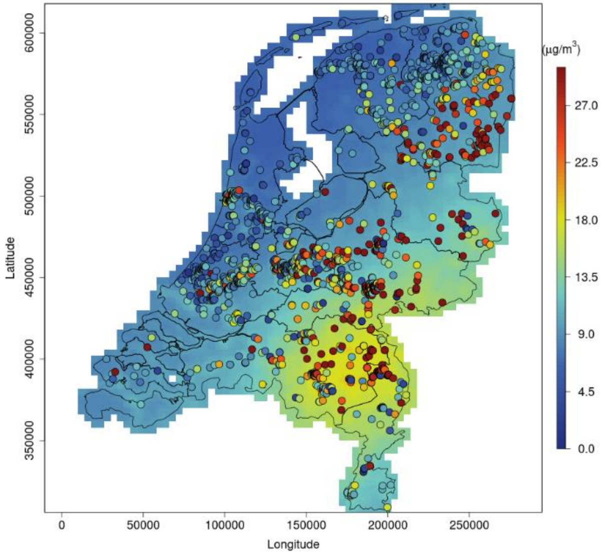
Application in Nantes (French city) for PM_{10} based on AtmoTrack sensors and ADMS-Urban simulations.

LCSQA
INERIS
maîtriser le risque pour un développement durable

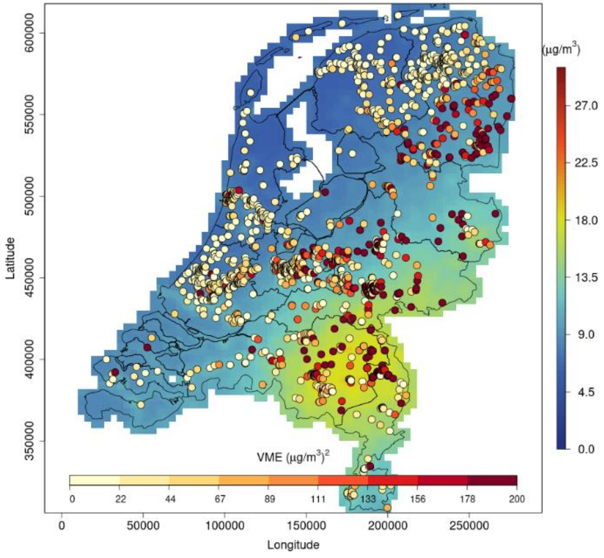
Data fusion for AQ mapping

SESAM (data fusion with **SE**n**S**or for **A**ir quality **M**apping**)**

RIO + calibrated sensor data

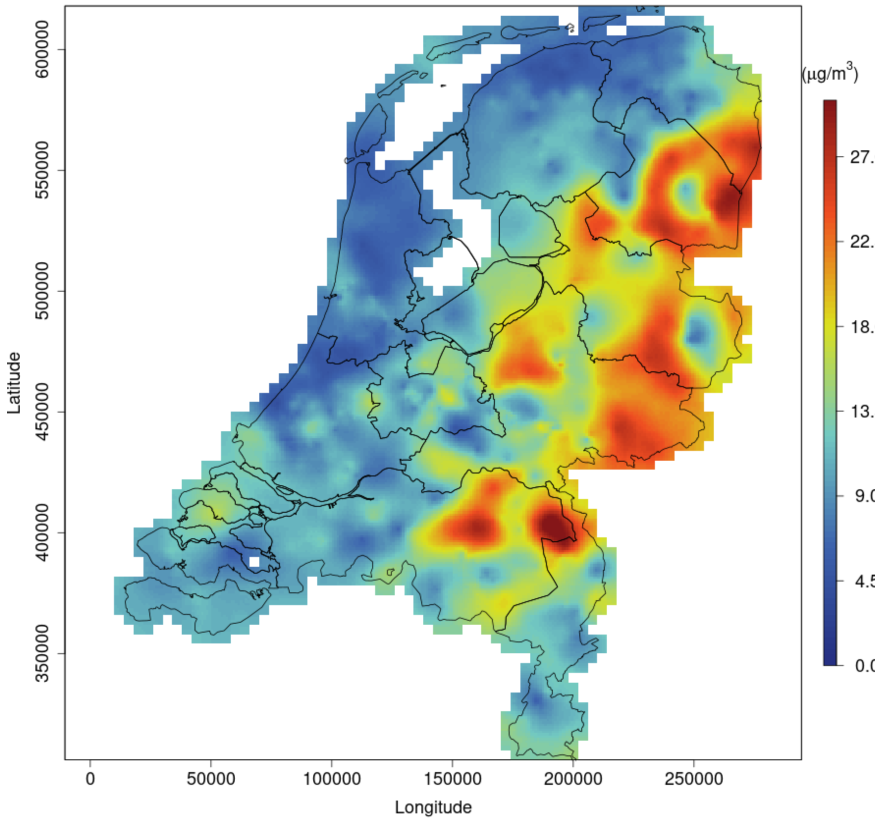


RIO + VME

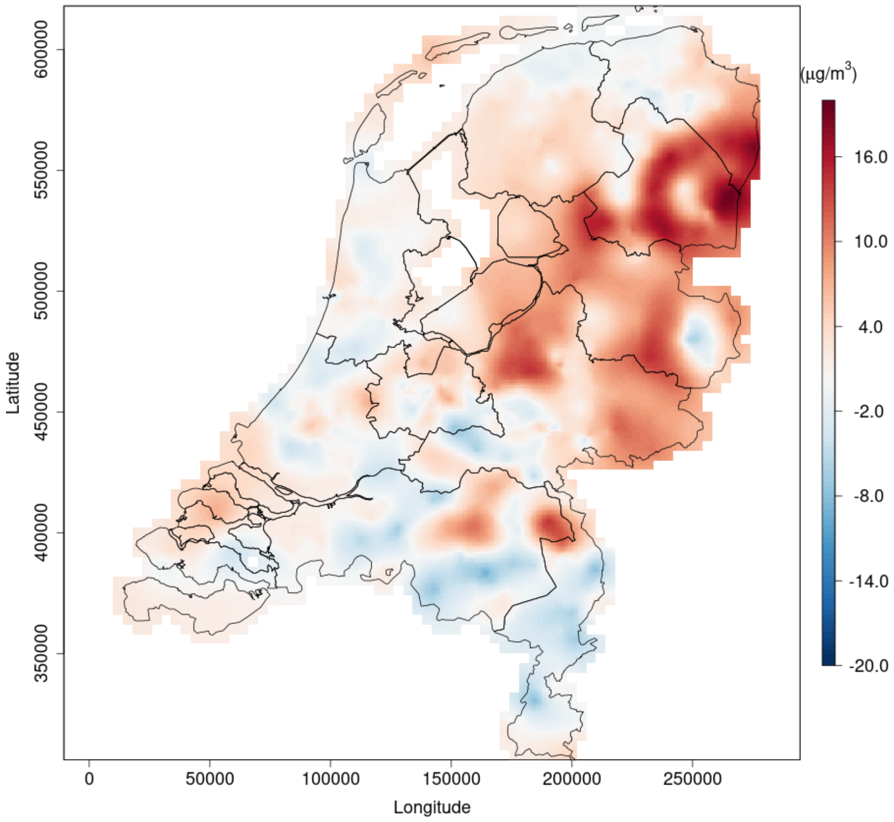


2021-01-18 11:00:00

Fused map



Model – Fused map



Conclusions

- Large number of $PM_{2.5}$ fixed sensors in Netherlands → **opportunity for air quality mapping**
- Questioning about the sensor data quality → outlier detection/calibration are necessary

1. **Outlier detection**

   data classification and definition of an interval of validity

2. **Calibration**

   calibration factor / RIVM approach (to be revisited)

Validation to be done based on synthetic data

3. **Data fusion - SESAM** (data fusion with SEnSors for Air quality Mapping)

- Room for improvement:
  - Estimate sensor uncertainty to be specified in the VME
  - "Rendez-vous" calibration based on the graph theory, adapt geostatistical models: spatio-temporal kriging, stochastic partial differential equations and numerical variogram for AQ mapping (PhD H. Rollin, Ineris)

Thank you for your attention!

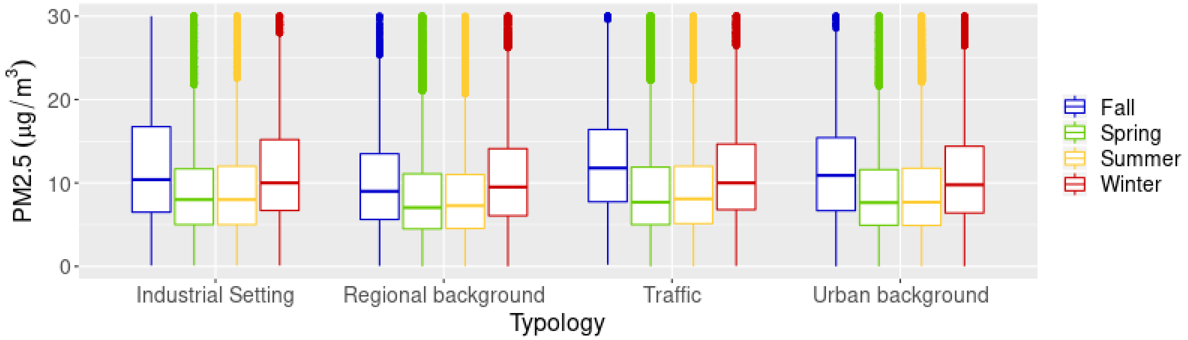
alicia.gressent@ineris.fr

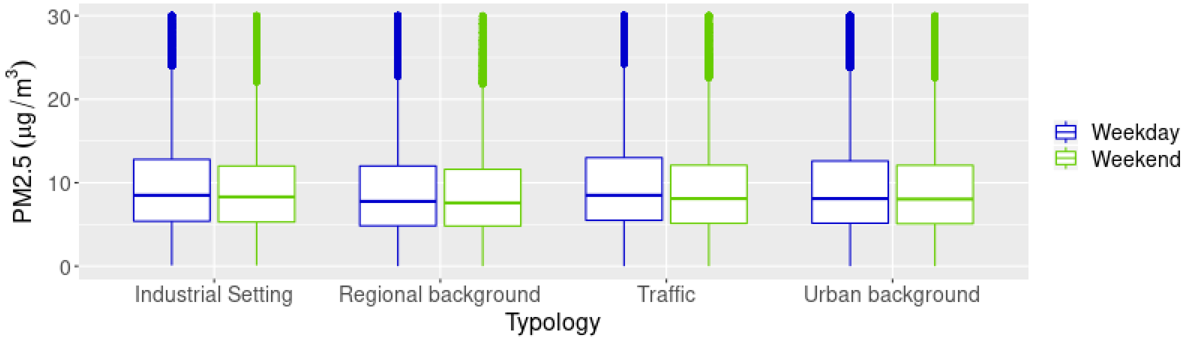
Extra slides

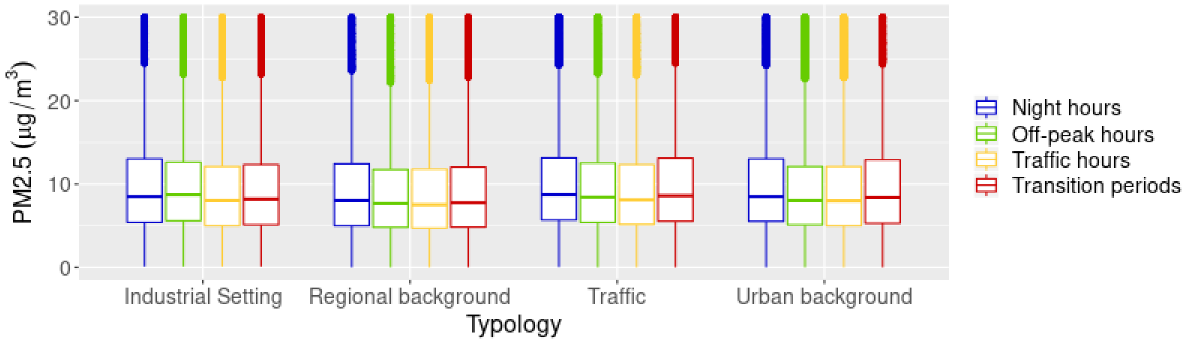
Outlier detection

Data classification: season

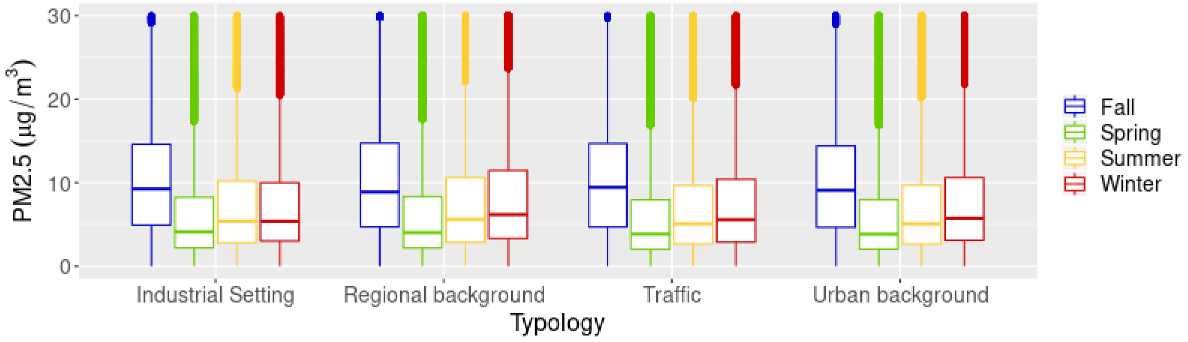
Reference stations

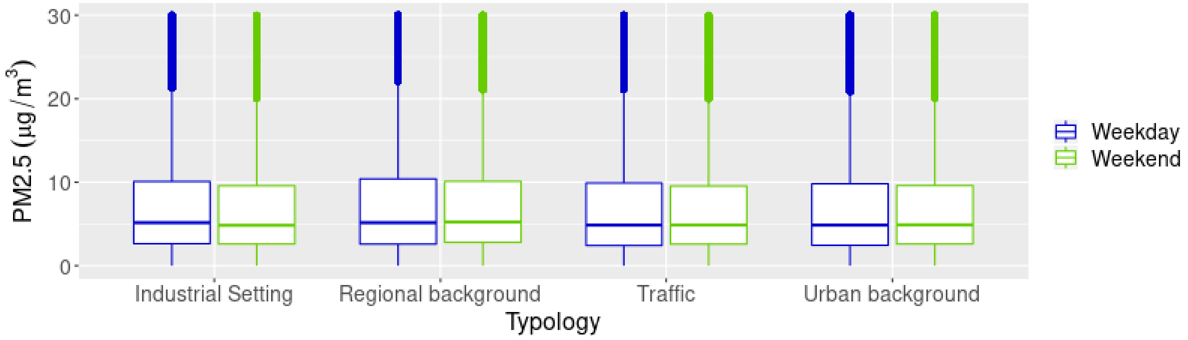


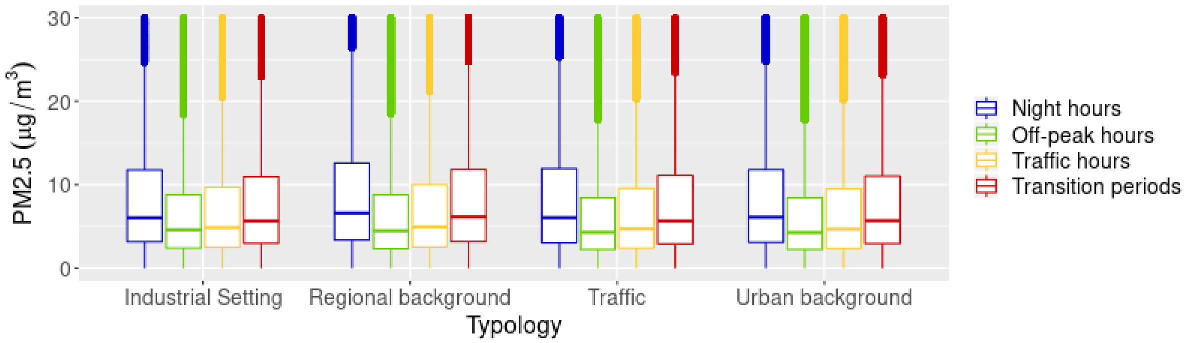




Sensors







- $PM_{2.5}$ levels mainly influenced by season

Fairmode Technical meeting, October 6-8, 2021.

INERIS
maîtriser le risque
pour un développement durable