RÉPUBLIQUE FRANÇAISE
*Liberté*
*Égalité*
*Fraternité*

INERIS
maîtriser le risque
pour un développement durable

# DATA FUSION FOR AIR QUALITY MAPPING USING LOW-COST SENSOR OBSERVATIONS

Alicia Gressent (alicia.gressent@ineris.fr)

Air Sensor International Conference – May 11-13, 2022.

RÉPUBLIQUE FRANÇAISE
*Liberté Égalité Fraternité*

INERIS
maîtriser le risque pour un développement durable

# Air quality sensors

## Usage

Monitoring Air Quality (fixed sensors)

*AirSensEUR*

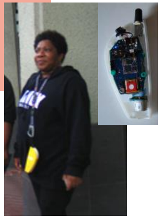
*Unitec srl, ETL3000 multi sensor station*

*Aeroqual, AQM 60 Air Quality station*

*AQMesh*

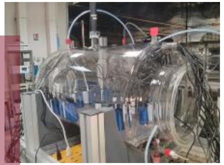
Public awareness

Individual exposure assessment

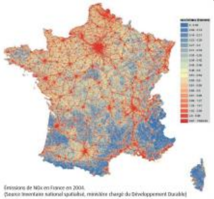



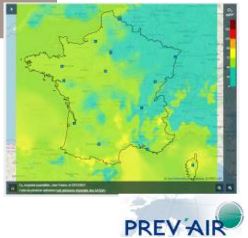
*Common sense, INTEL Lab, Berkley - USA*

Improvement of sensor performance




Improvement of emissions inventories and AQ modelling






Redesign the regulatory AQ monitoring network

# Air quality sensors

## Usage

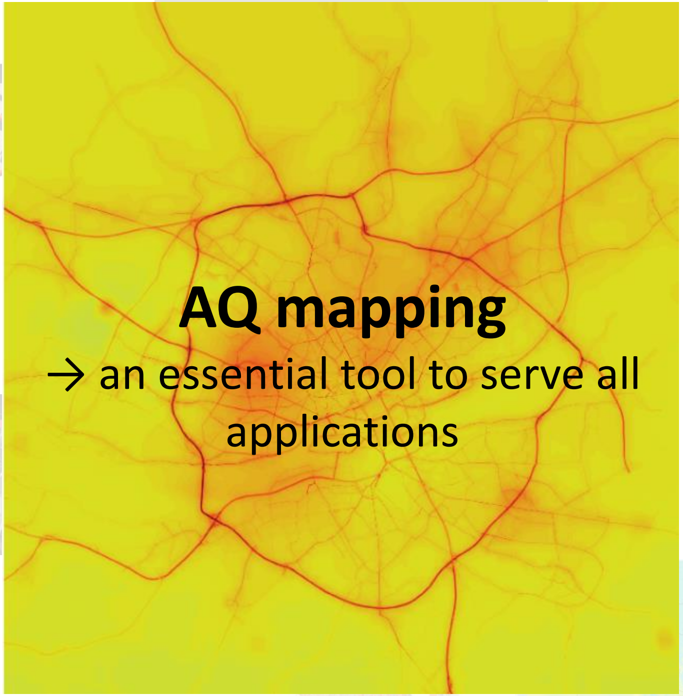
Monitoring Air Quality (fixed sensors)


*AirSensEUR*


*Unitec srl, ETL3000 multi sensor station*


*Aeroqual, AQM 60 Air Quality station*


*AQMesh*



**AQ mapping**

→ an essential tool to serve all applications

Individual exposure assessment




*Common sense, INTEL Lab, Berkley - USA*

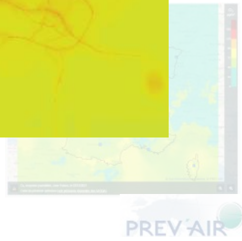
Improvement of sensor performance






Re-design the regulatory AQ monitoring network



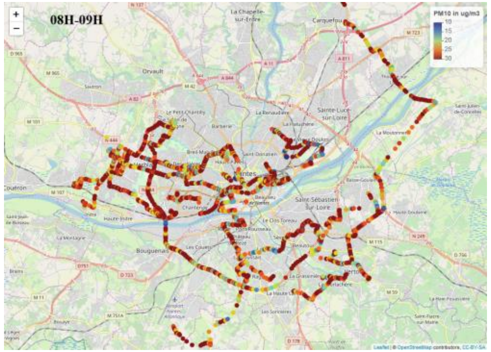


RÉPUBLIQUE FRANÇAISE
*Liberté Égalité Fraternité*

INERIS
maîtriser le risque pour un développement durable

# AQ mapping at the urban scale

## Challenges



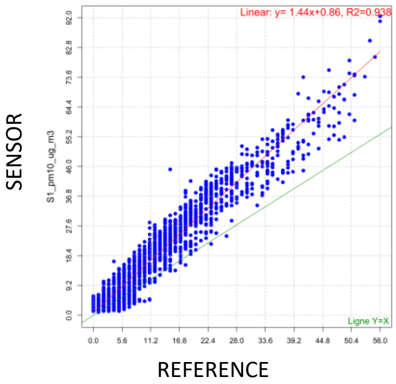

**Representativeness**
(Support/sampling plan)

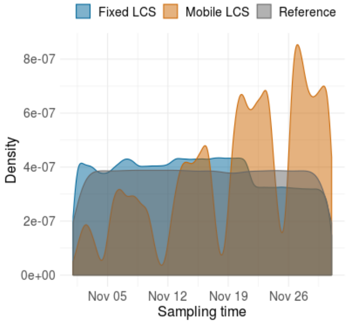
**Sensor observations**

**Uncertainties**
(Regulatory framework)

**Heterogeneities**
(Time and space)


PM2.5



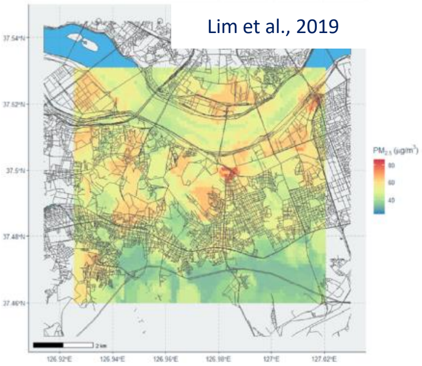



# AQ mapping at the urban scale

## Methods

### Land use regression models

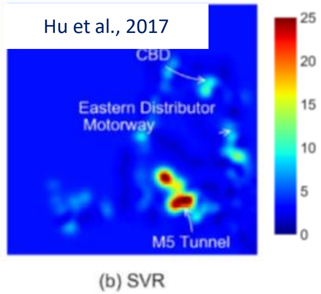
**Multilinear regressions** between pollutant concentrations measured by sensors and predictive variables (+ machine learning ⇔ random forest, stacked ensemble)




Estimate of $PM_{2.5}$ concentrations in Seoul, South Korea.

### Other statistical approaches

**Machine learning:** SVR : Support Vector Regression, DTR : Decision Tree Regression, RFR : Random Forest Regression, XGB : Extreme Gradient Boosting, MLP : Multi-Layer Perceptrons, LR : Linear Regression, ABR : Adaptive Boosting Regression.




Estimate of CO concentrations, Sydney, Australia.

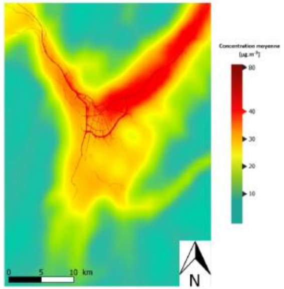
### Bayesian approach for data fusion

Consider uncertainties related to model and measurements

Update of concentrations values and uncertainties in fused map

**BLUE approach (Best Linear Unbiased Estimator) – Kalman Filter**

MOBICIT'AIR (Atmo Auvergne Rhône Alpes)



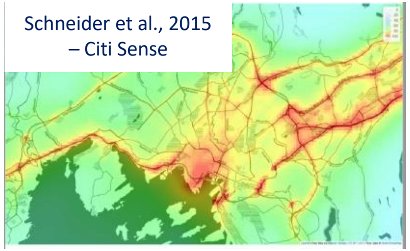

Estimate of $NO_2$ concentrations on 01/09/2017 at 9 am, Grenoble, France.

### Geostatistical approach for data fusion

**Kriging:** estimate that consider observed values and the information on the position

Concept of spatial continuity

**Nonstationary case ⇔ data fusion based on kriging with an external drift (universal kriging)**




Estimate of $NO_2$ concentration in Oslo, Norway.

# AQ mapping at the urban scale

## SESAM (data fusion with SEnSor for Air quality Mapping)

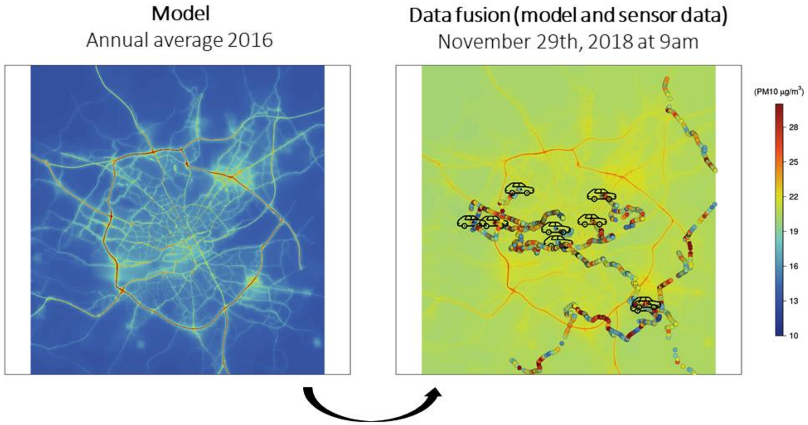
- **Data fusion** ⇔ combination of sensor observations and modelling estimates at the urban scale
- **Method:** kriging with an external drift with a weighting of the sensor observations depending on data dispersion and measurement uncertainty (⇔ **V**ariance of **M**easurement **E**rror)
- **Application:** Nantes (modelling data provided by Air Pays de la Loire – a regional AQ monitoring association / PM sensor data provided by AtmoTrack)


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]

Concentrations of $PM_{10}$ in Nantes



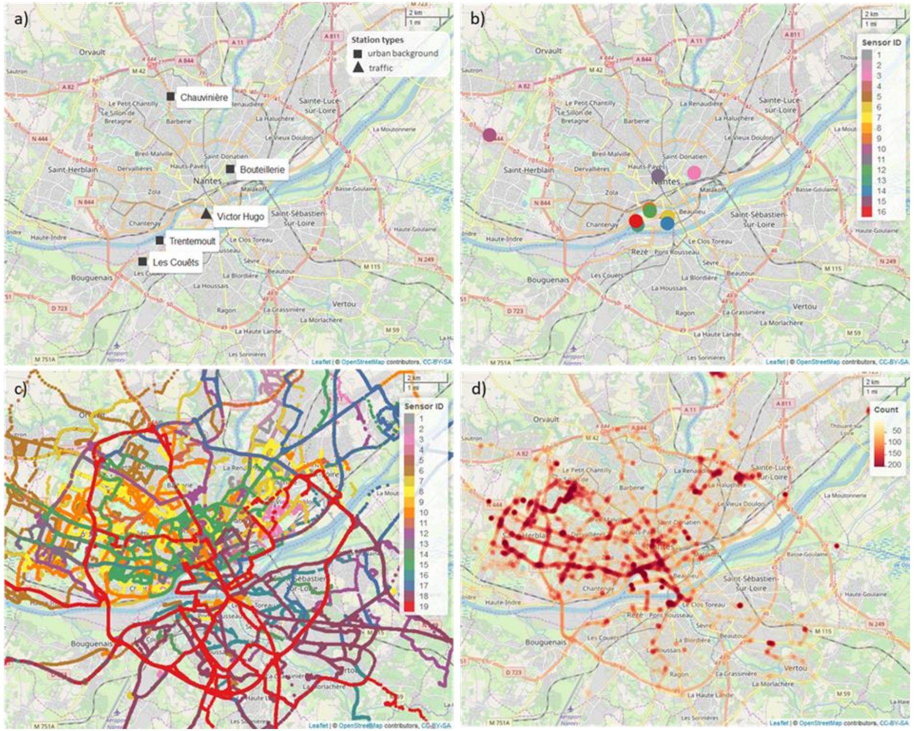

SESAM
(data fusion with SEnSors for Air quality Mapping)

https://github.com/AliciaGressent/SESAM

# AQ mapping at the urban scale

## SESAM (data fusion with SEnSor for Air quality Mapping)

### Application in Nantes




air pays de la loire
www.airpl.org

ATMOTRACK

AQ monitoring network stations (Air Pays de la Loire) (a), AtmoTrack fixed sensors positions (b), sampling trajectories of AtmoTrack mobile sensors (c) density of fixed and mobile sensor observations (d), are presented for November 2018 in Nantes.

1. Analysis and preprocessing of sensor data

2. Estimate of the variance of measurement error

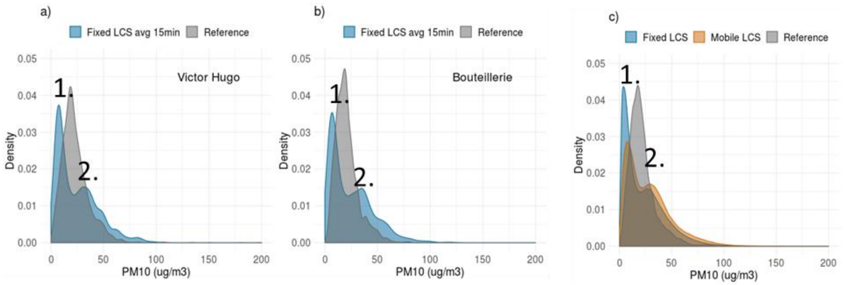
3. Kriging with an external drift ⇔ fixed and mobile $PM_{10}$ sensor data and ADMS-Urban calculations

# AQ mapping at the urban scale

## SESAM (data fusion with SEnSor for Air quality Mapping)

### Preprocessing of sensor data

Sensor obs. vs. reference




1. Ultrafine particles, algorithm effect
2. Real concentrations, particle size effect

Preprocessing of data set in **two steps:**

- Identification of pollution free periods and selection of the corresponding sensor observations
- Standard deviation:

$$Sr = \sqrt{\frac{\sum(R_i-\bar{R})^2}{N-1}}$$

- Repeatability: $\boldsymbol{R} = \boldsymbol{2}\sqrt{\boldsymbol{2S_r}}$

Elimination of values below the repeatability threshold

(Spinelle et al., 2013)

Correction of the daily variation of the background

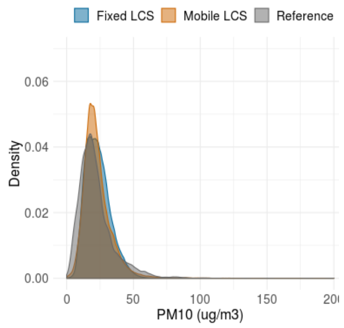
**Cr = Ci - [Bg_data - Bg_ref]**

**Corrected conc.** **Initial conc.** **Reference stations (daily mean)**

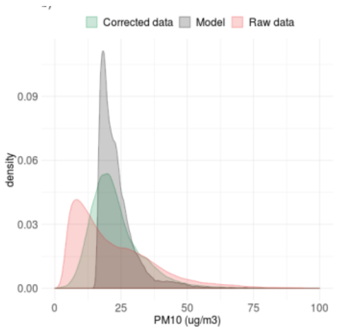
**15-min moving median on a continuous run of measurements**

(Hankey et al., 2015)

Corrected sensor obs. vs. reference




Corrected sensor obs. vs. model



# AQ mapping at the urban scale

## SESAM (data fusion with SEnSor for Air quality Mapping)

### Kriging parameters

- Variance of measurement error:

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

Where $\sigma$ is the standard deviation of pollutant observations at position i, N is the number of observations at position i, $\nu_r$ is the constant relative uncertainty of type (that depends on the sensor type), and $C_j$ is the j$^{ème}$ pollutant concentration at position i.

$\nu_r$ ⇔ constant relative uncertainty of type:

(25% reference observations)
50% fixed sensor observations
75% mobile sensor observations

→ That definition relies on the **European Directive** (Directive 2008/50/CE) and a sensor data analysis

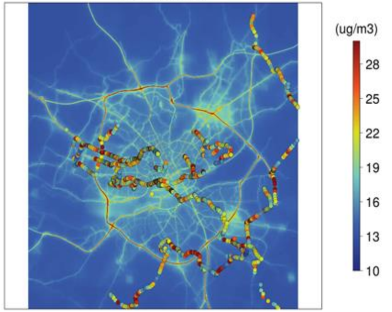
- Universal kriging with the 2016 annual mean of the pollutant concentrations estimated by ADMS-Urban as the drift
- Hourly estimate between 7am and 7pm using sensor observations only
- Spatial resolution ⇔7m

# AQ mapping at the urban scale

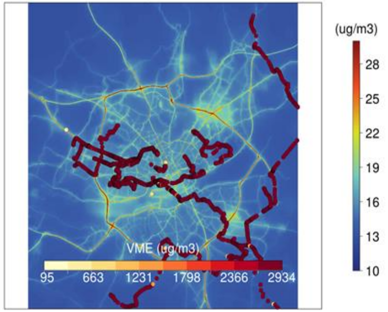
## SESAM (data fusion with SEnSor for Air quality Mapping)

### Kriging

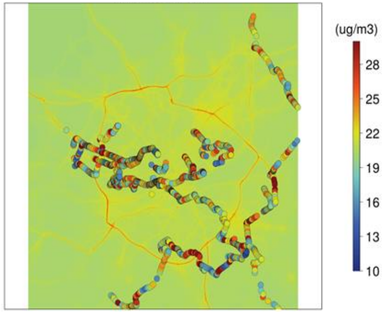
Model + sensor obs.



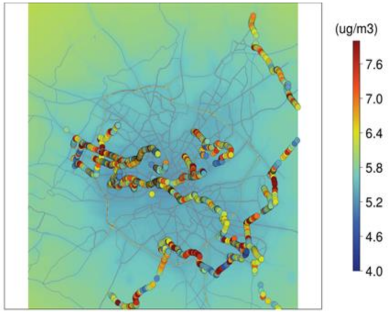

Model + VME




Fused map + sensor obs.

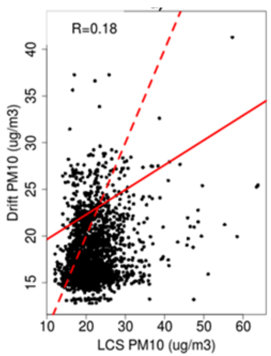



Kriging standard deviation + sensor obs.

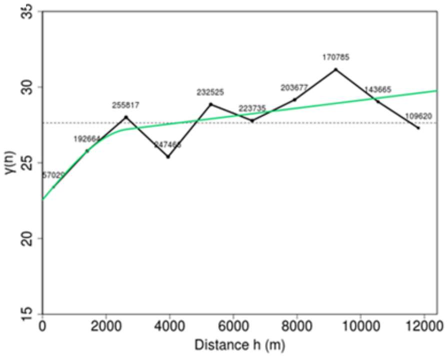



kriging

Correlation obs. and model




Variogram of the residuals




Low correlation between sensor observations and model estimate that can be explained by:

- The definition of the drift (annual mean)
- Sensor data quality ⇔ spatial and temporal representativeness (impact of mobility)

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INERIS
maîtriser le risque pour un développement durable

# AQ mapping at the urban scale

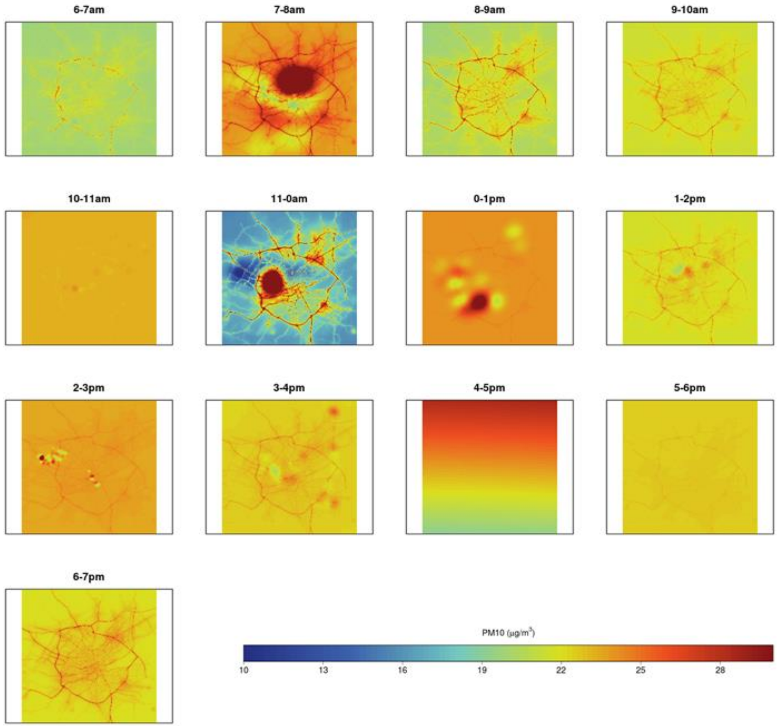
## SESAM (data fusion with SEnSor for Air quality Mapping)

### Kriging results

- **At 8am, 11am, 0pm, 1pm, 2pm, 3pm and 4pm**: local hotspots
  → few LCS data points of high or low $PM_{10}$ concentrations
  → high influence in the data fusion / low VME in kriging
  Correlation between the LCS data and the drift ⇔ [0.01 to 0.18] & high nugget effect (20 µg/m³) in the variogram
  ⇔ high kriging standard deviation, up to 12 µg/m³.

- **At 7am, 9am, and 10am**: no hotspots
  → more data points associated with low VME
  Correlation between the LCS data points and the drift ⇔ [0.05 to 0.28] & high nugget effect (20 µg/m³) in the variogram but it is better structured and a variogram model can be more easily fitted
  ⇔ lower kriging standard deviation, up to 6 µg/m³.

- **At 5pm, 6pm and 7pm**: the correlation between the data and the drift <0
  ⇔ estimations are admitted as no relevant



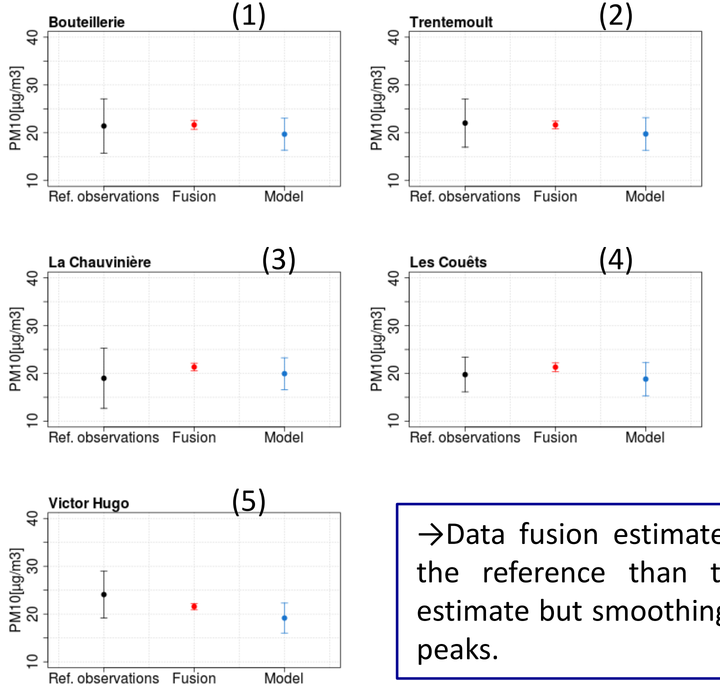

The fused maps for $PM_{10}$ derived from the external drift kriging approach are shown from 7am to 7pm on 11/29/2018 in Nantes.

# AQ mapping at the urban scale

## SESAM (data fusion with SEnSor for Air quality Mapping)

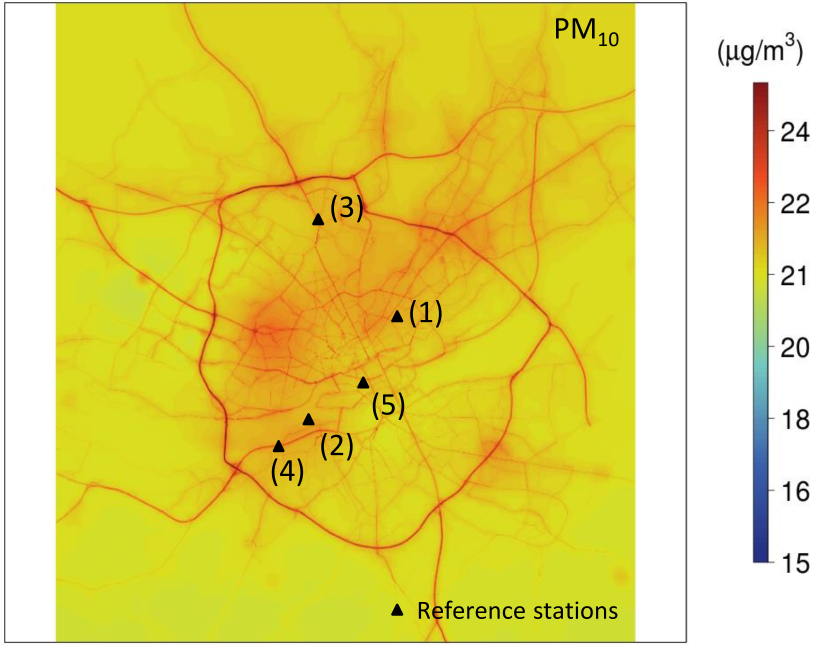
### Kriging results

Comparison of the daily mean at the stations




→Data fusion estimate is closer to the reference than the modeled estimate but smoothing of pollution peaks.

Fused map of the daily mean (11/29/2018)



# AQ mapping at the urban scale

## SESAM (data fusion with SEnSor for Air quality Mapping)

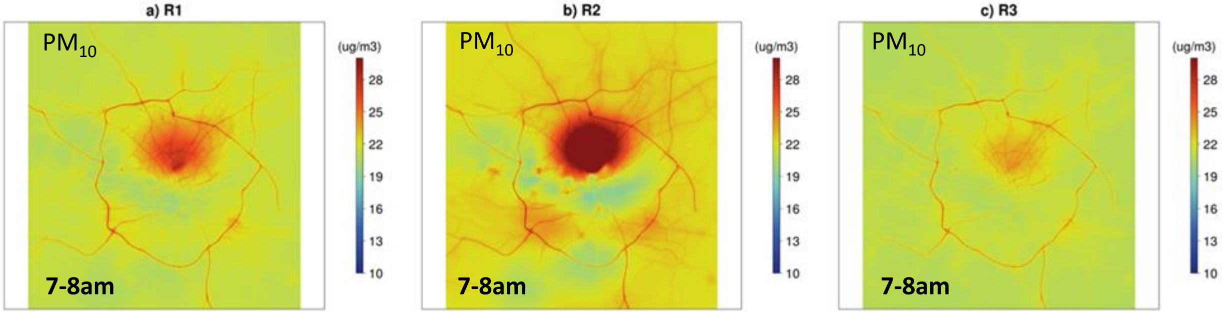
### Kriging results

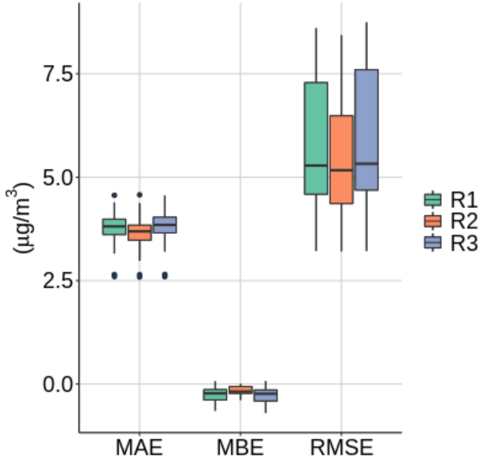
Impact of the measurement uncertainty on estimation

Uncertainties: 25% REF, 50% FS, 75% MS

Uncertainties: 25% REF, 25% FS, 25% MS

Uncertainties: 25% REF, 100% FS, 150% MS




→ Efforts needed to quantify measurement uncertainty associated with sensors / impact of mobility

# Conclusions

## Main findings and ongoing work

### Kriging with sensor data

- Data fusion reduces the bias from 8% to 2% when considering sensor observations instead of the model alone
- Data fusion smooths the $PM_{10}$ concentration peaks but presents better estimate than model of the pollutant levels in average
- Data fusion performance is increasing by reducing the sensor data uncertainty + spatial impact on the $PM_{10}$ concentration fields

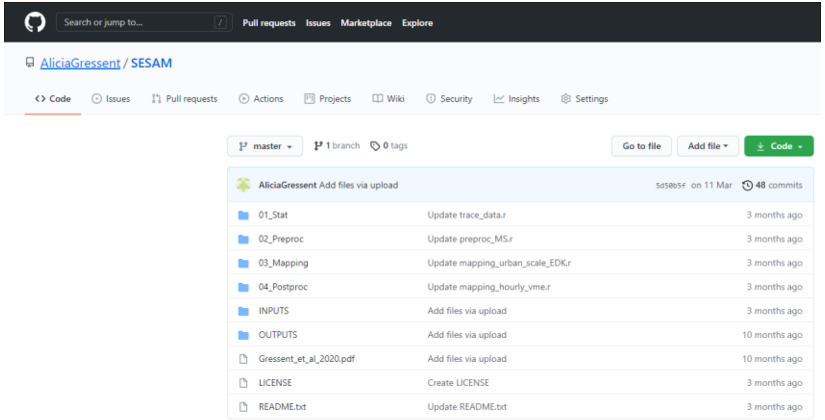
### Improve sensor data characterization and data fusion approaches

- Outliers' detection, drift of the sensors…
- Fixed and mobile sensor network: sensor network recalibration, rendez-vous approach (Rollin et al., in prep., Ineris)
- Qualification/quantification of measurement uncertainty to be better considered in SESAM
- Validity of sensor data in mobility
- Development of new methods of data fusion: numerical variogram, spatio-temporal kriging, SPDE (Stochastic Partial Derivative Equations), Machine Learning / deep learning

RÉPUBLIQUE FRANÇAISE
*Liberté Égalité Fraternité*

INERIS
*maîtriser le risque pour un développement durable*

# Thank you for your attention!




Contact: alicia.gressent@ineris.fr

https://github.com/AliciaGressent/SESAM

LCSQA
Laboratoire Central de Surveillance de la Qualité de l'Air