Optimization of environmental sensor placement in industrialized areas

Anton Sokolov[(1)*], Khaoula Karroum[(2)], Hervé Delbarre[(1)], Yann Ben Maissa[(3)] Mohamed Elhaziti [(2)]

[(1)] **Laboratory of Physics And Chemistry of Atmosphere, University of Littoral Côte d'Opale & University Lille Nord, Dunkirk, France**

[*] ***Anton.Sokolov@univ-littoral.fr***

[(2)] **LRIT-CNRST, URAC 29, Faculty of Sciences Mohammed V University in Rabat, Rabat, Morocco**

[(3)] **Telecommunication Systems, Networks and Services Lab, National Institute of Posts and Telecommunications, Rabat, Morocco**

Scientific conference EcoAirSolutions 2021, Dnipro, December 9 - 10

Plan of the talk

1. Introduction
2. Methods and algorithms
   - Interpolation of measurements
   - Optimization algorithms
   - Error and constraints
3. Area of study and the dataset
4. Results of optimization
5. Error sensitivity to N of sensors
6. Computational aspects
7. Conclusions and perspectives

1. Introduction

The optimization of positions of sensors could be done on the basis of :

1. **some a-priori knowledge of distribution and sources of a physical field,**
2. **previous measurements,**
3. **the modeling of the field,**

   - Background "True" state

We can have a particular interest for the field value in a specific region

In our case the physical field is an atmospheric PM10 pollution, and the result of the modelling considered as the 'true' state

2. Methods and algorithms

The quality of a measuring network can be characterized by the difference $\overrightarrow{\delta} = \{\delta_i\}$ between interpolated and "true" values in control points:
$\boldsymbol{\delta_i}(\overrightarrow{\boldsymbol{x^s}}, \overrightarrow{\boldsymbol{y^s}}) = \boldsymbol{I_i}(\overrightarrow{\boldsymbol{x^s}}, \overrightarrow{\boldsymbol{y^s}}) - \boldsymbol{P^b}(\boldsymbol{x_i}, \boldsymbol{y_i})$, where (x_i, y_i) are coordinates of the i^{th} control point, I_i is an interpolated pollution value in this point, $\boldsymbol{P^b}$ is a background true pollution value and, $(\overrightarrow{\boldsymbol{x^s}}, \overrightarrow{\boldsymbol{y^s}})$ - are coordinates of sensors.

Interpolation techniques in 2D:

1. The inverse distance weighting (*)
2. Linear interpolation with nearest neighbour extrapolation
3. Natural neighbour with nearest neighbour extrapolation
4. Gaussian Process interpolation

To find a global minimum of the error, the following algorithms were applied:

a) GlobalSearch Algorithm
b) Genetic Algorithm (GA, *)
c) Particle Swarm Optimization (PSO, *)

(*) PhD thesis of ***K. Karoum***, December 2020

2.1 Interpolation techniques in 2D

Inverse Distance Weighting

$$y(x) = \frac{\sum_{i=0}^{N} w_i(x) y_i}{\sum_{i=0}^{N} w_i(x)} \qquad w_i(x) = \frac{1}{\mathrm{d}(x, x_i)^p}$$

y : estimated PM_{10} pollution value at x position,

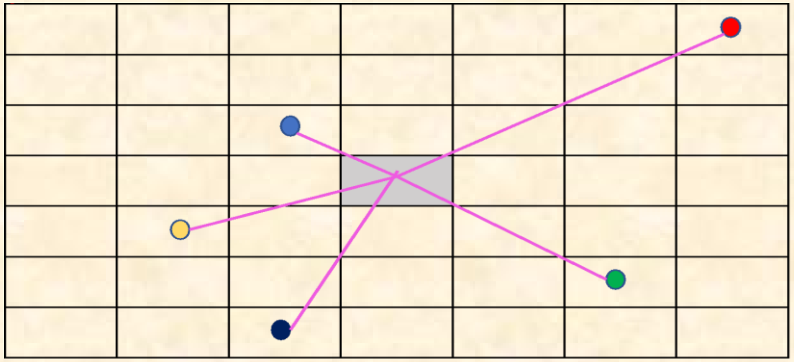
y_i: known values of the pollution

x_i: position of sensors,

$\mathrm{d}(x, x_i)$: the distance between x and x_i,

n : number of measurements,

p – some hyperparameter

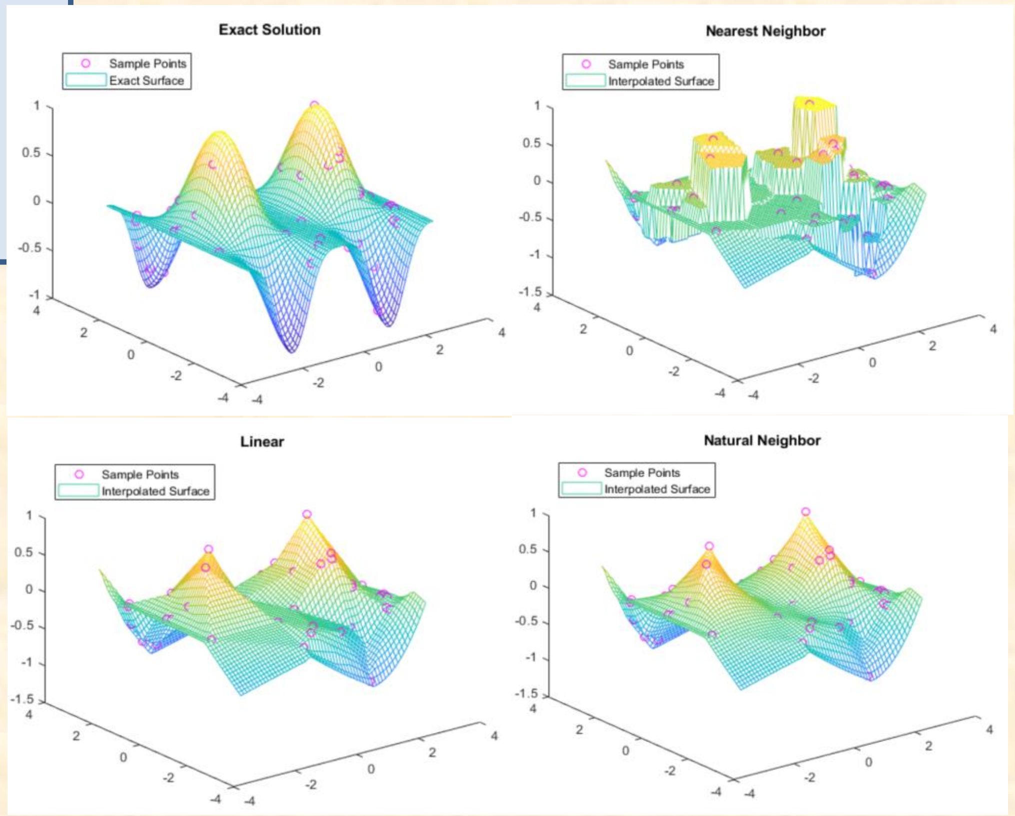


Triangulation-based methods:

- Nearest neighbor (NN)
- Linier interpolation*
- Natural neighbor *

* NN method was used for extrapolation

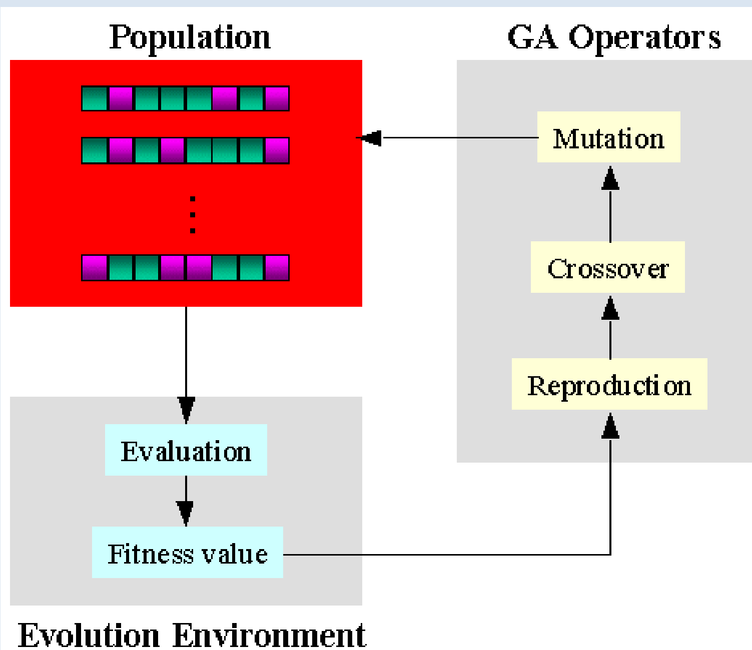
Examples are taken from Matlab software help for function *v(x,y)=sin⁴(x)·cos(y)*



2.2. Minimization algorithms

$$\varepsilon\left(\overrightarrow{x^s},\overrightarrow{y^s}\right)=\left\|\left\{I_i\left(\overrightarrow{x^s},\overrightarrow{y^s}\right)-P^b(x_i,y_i)\right\}\right\|_p$$

GA is a method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution



https://www.ewh.ieee.org/soc/es/May2001/14/Begin.htm

PSO is a computational method that optimizes a problem by iteratively trying to improve a candidate solution. It solves a problem by having a population of candidate solutions(particles)

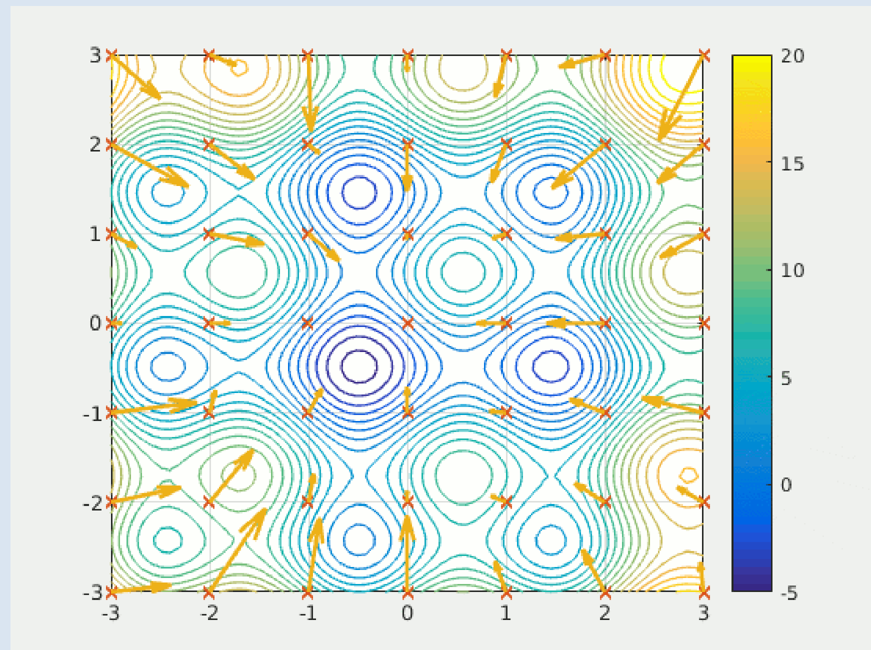


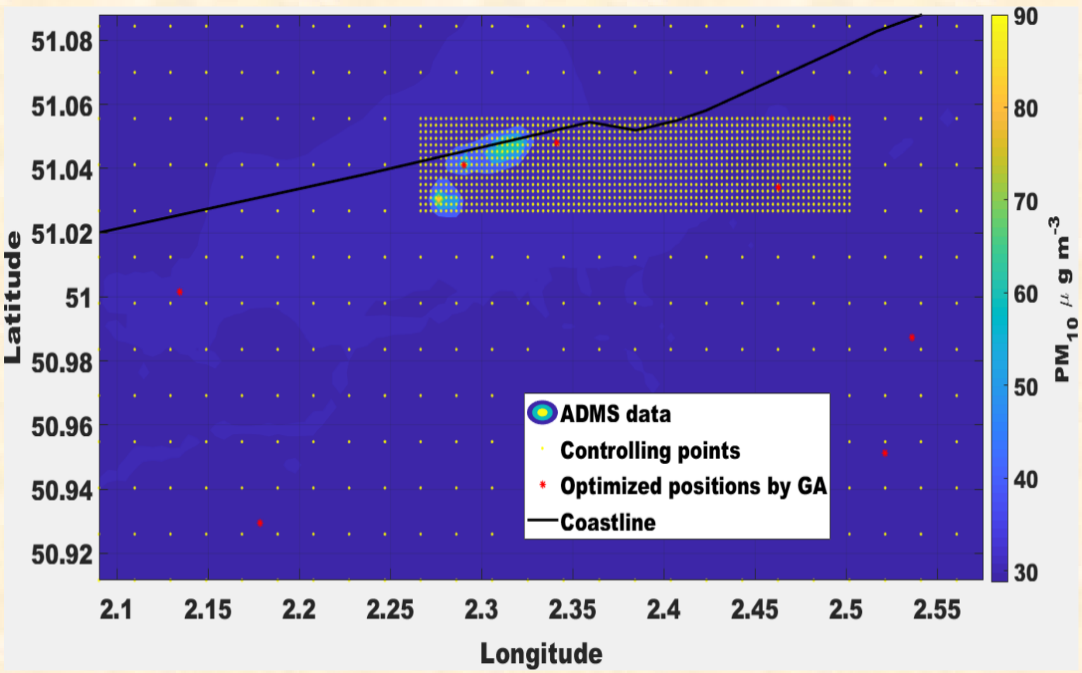
Image from Ephramac - Own work, CC BY-SA 4.0,https://commons.wikimedia.org/w/index.php?curid=54975083

2.3. Error to optimise and constraints

- $(\vec{x}, \vec{y}) = \arg\min\limits_{\overrightarrow{x^s}, \overrightarrow{y^s}} \varepsilon\left(\overrightarrow{x^s}, \overrightarrow{y^s}\right),\ \varepsilon\left(\overrightarrow{x^s}, \overrightarrow{y^s}\right) = \left\|\vec{\delta}\left(\overrightarrow{x^s}, \overrightarrow{y^s}\right)\right\|_p,$

$$\delta_i\left(\overrightarrow{x^s}, \overrightarrow{y^s}\right) = I_i\left(\overrightarrow{x^s}, \overrightarrow{y^s}\right) - P^b(x_i, y_i)$$

- Using different p-norms allows to vary the importance of areas with big error values
- The control points positions (white dots on the fig. below *) could be uniformly distributed or concentrated in a domain of interest



- Optimizations could be constrained by:
  - Putting stations in available/prohibited domain
  - Limiting the distance between sensors
  - …

(*) From Karroum et al. 2021, in preparation

3. Area of study and the dataset

-

Dunkirk region North of France

Sources of PM10 pollution:

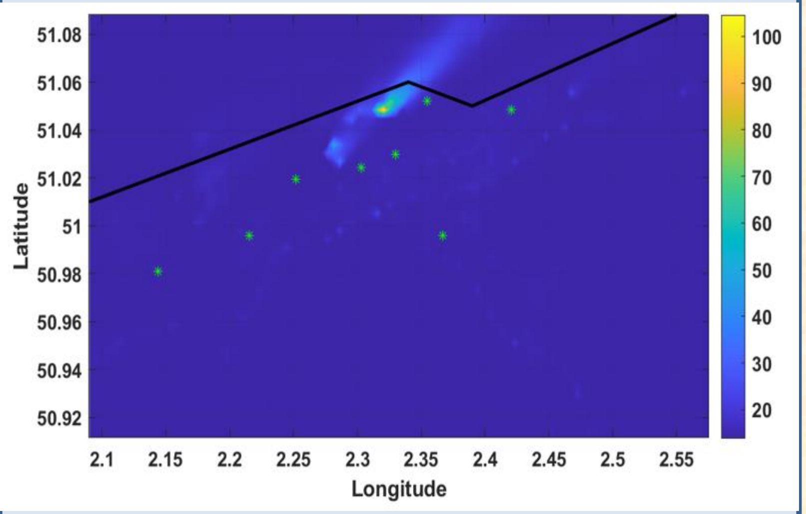
- Local industry
- Sea port and English channel boat circulation
- Local transport
- Far away sources

Data on pollution: air quality agency ATMO-Nord (https://www.atmo-hdf.fr/)

Dunkirk region corresponding to ADMS model domain with superposed population density.

ADMS is a pollution dispersion model used to model the air quality impact of existing and proposed industrial installations

Background true values – one year of PM10 modeling, data at surface level, resolution ~1 hour, ~500 m

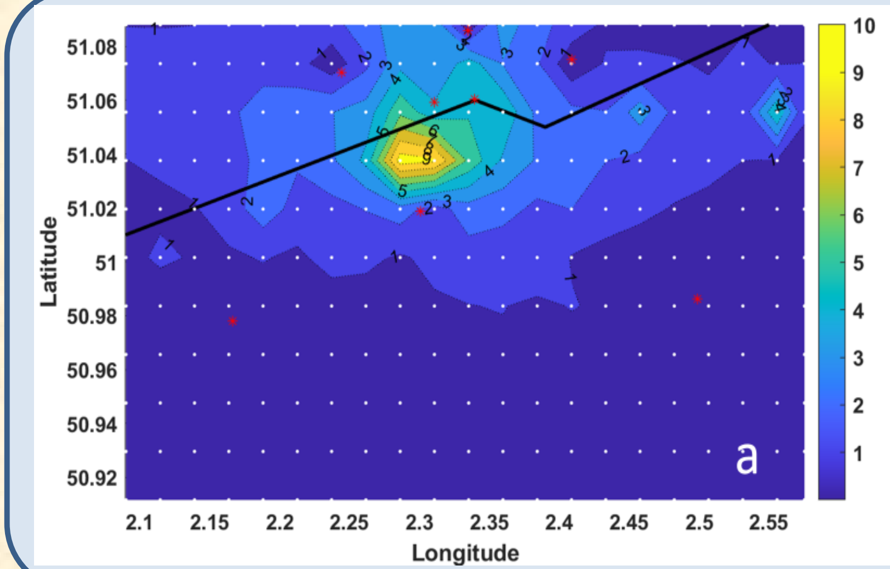


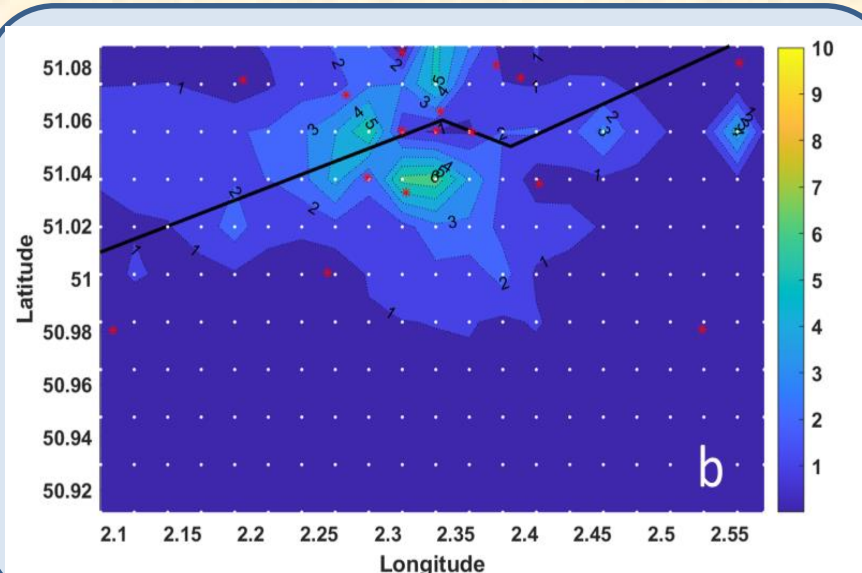
Mean pollution value in µg/m3 and positions of existing ATMO measuring stations (green stars). Error of interpolation is 2.45 µg/m3

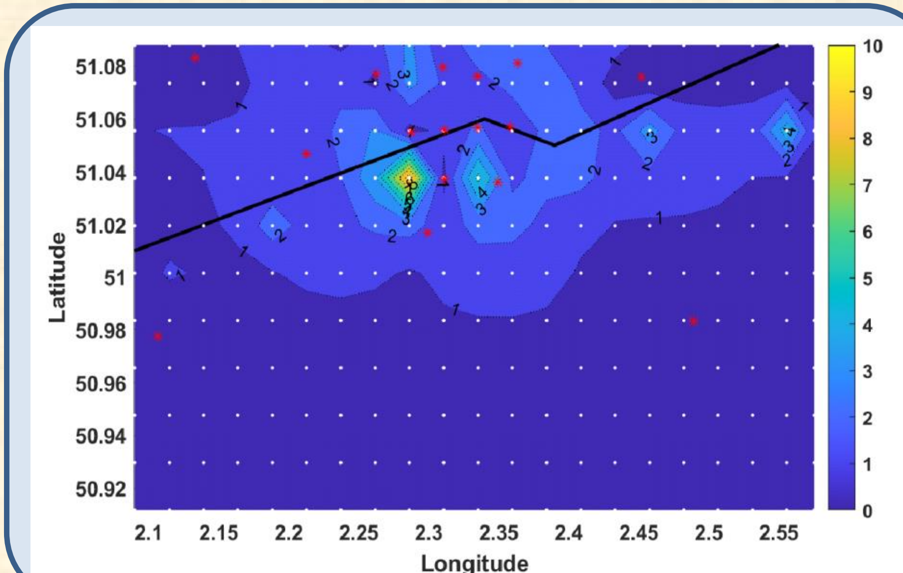
4. Results of optimization

Optimal positions of
(a) 8 sensors + linear scatter interp.,
(b) 16 sensors + linear scatter interp.,
(c) 16 sensors + **natural neighbor** interp.,

and the difference between the estimated (interpolated) value ant the ‘true’ background pollution in μg/m^3
Corresponding $\boldsymbol{\varepsilon}$ values are **1.80**, **1.48** and **1.39**



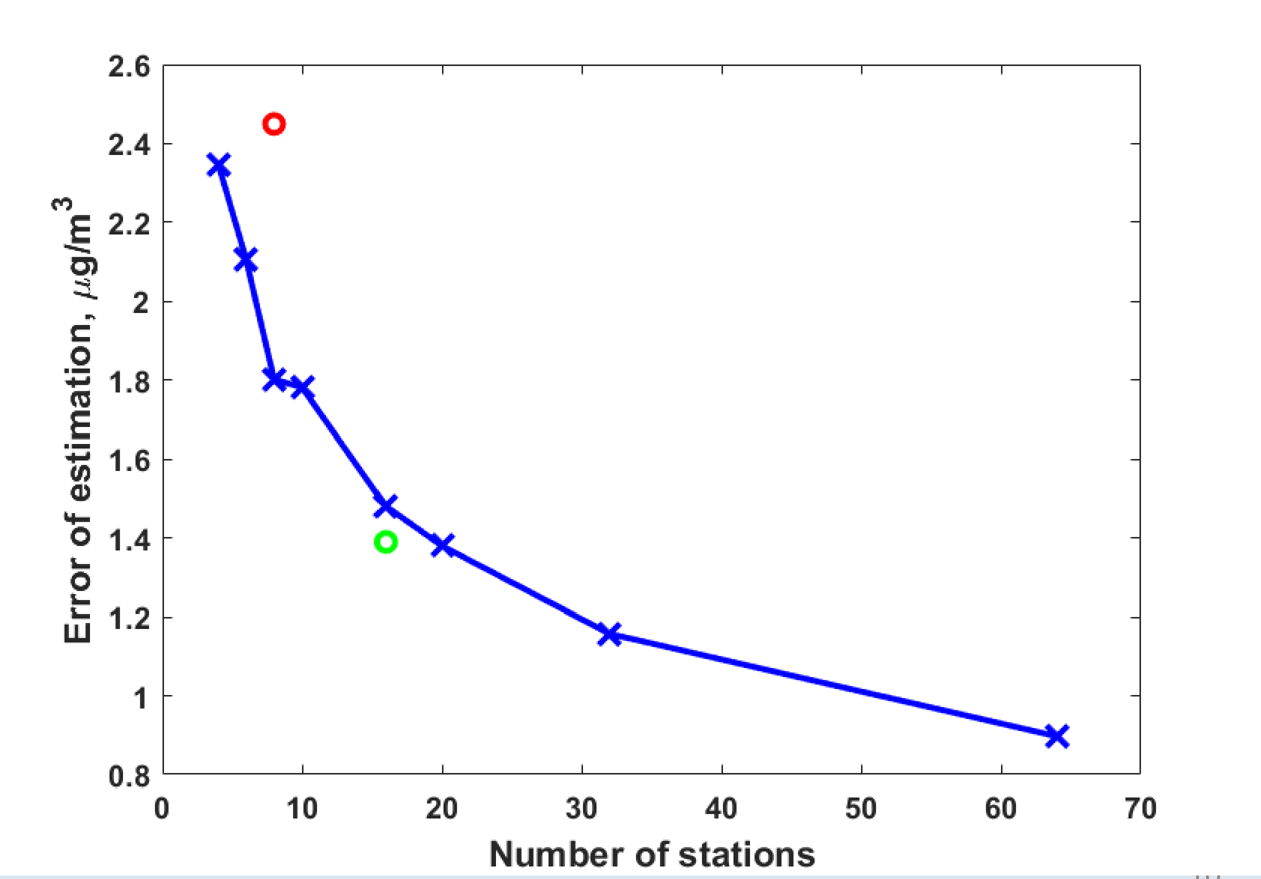




5 Error as a function of N stations

- Allows to find the optimal compromise between the precision of the pollution estimation and the cost of the deployment of the measuring network

Fig. Estimated optimal error as a function of number of stations. Red circle corresponds to the error value calculated by the actual ATMO measuring station configuration. Green circle corresponds to natural neighbor scattered interpolation technique.



6. Computational aspects.

- Minimization efficiency highly depends on dimensions of data and the number of stations
- Both the **Interpolation** and the **optimization** algorithms applied, could be used efficiently on parallel systems
- One 16 stations optimization took about 30 minutes using 72 cores of the local University multiprocessor cluster*
- Genetic Algorithms and Particle Swarm Optimization converges ~10 times more rapidly than the Global Search algorithm

(*) https://www-calculco.univ-littoral.fr/

Conclusions

- Diminishing of estimation error of PM10 pollution ε from 2.45 μg/m3 to 1.8 μg/m3 (~25%) for 8 stations
- The calculated dependence of ε on stations number N, allowing estimating the tradeoff between the cost of deployment and the precision of pollution estimation
- Application of more advanced natural neighbor interpolation increases the precision of pollution estimation by ~6%
- Genetic Algorithms and Particle Swarm Optimization has a similar performance. They outperforms the Global Search algorithm.

Perspectives

- Optimization of sensor position in Kryvyi Rih region using a few years simulation of CalPuff dispersion model.
- Application of the Gaussian process regression interpolation which take into account the dispersion proprieties of the atmosphere (the wind)

Thank you for your attention…