Application and intercomparison of three data assimilation methods for air quality evaluation on the Île-de-France area

[a]Chi Vuong NGUYEN, [a]Lionel SOULHAC, [b]Cyril Joly and [b]Olivier SANCHEZ

3rd June 2019

19th HARMO conference, Bruges

[a] Univ Lyon, Ecole Centrale de Lyon, INSA Lyon, Université Claude Bernard Lyon I, CNRS, Laboratoire de Mécanique des Fluides et d'Acoustique, UMR 5509, 36 Avenue Guy de Collongue, F-69134, ECULLY, France

[b] Airparif

ÉCOLE CENTRALE LYON

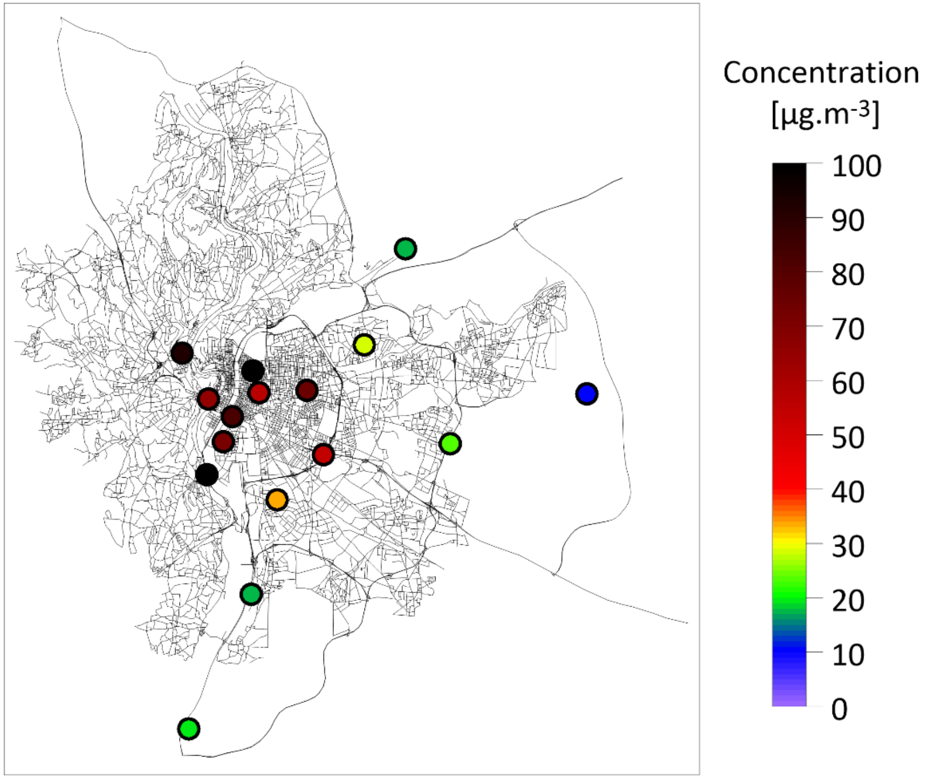
INTRODUCTION

Introduction

Evaluation of urban air quality: measurements

✓ Accurate data

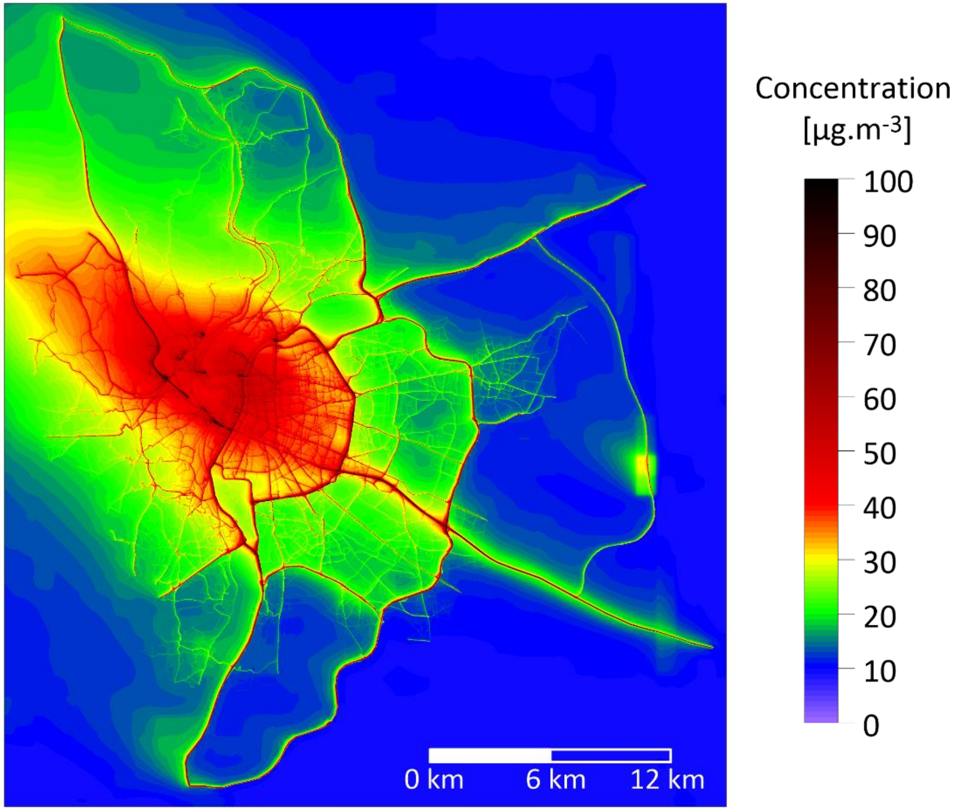
X Heterogeneous spatial distribution



Introduction

Evaluation of urban air quality: modelling

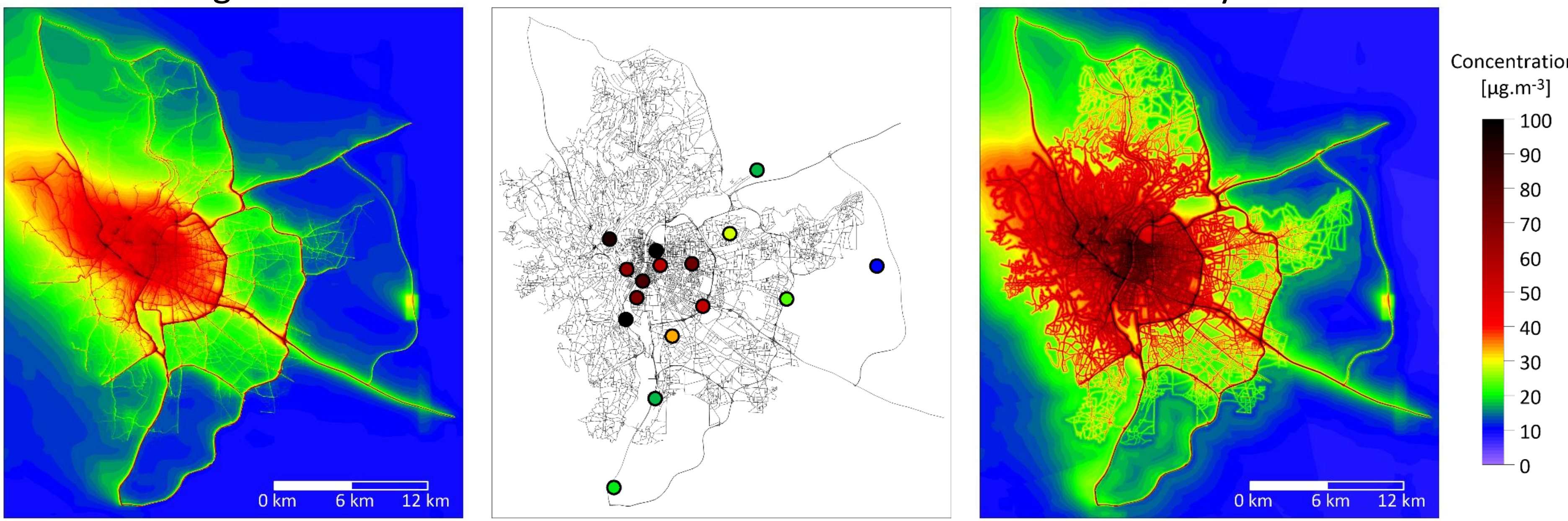
- ✓ Fine spatial resolution
- ✓ Forecast
- ✓ Scenario studies
- ✓ High number of species
- X High uncertainties



Introduction

Evaluation of urban air quality: data assimilation

Background | Measurements | Analysis



Data assimilation (DA): combination of measurements and modelled data to determine the best estimate of the system state

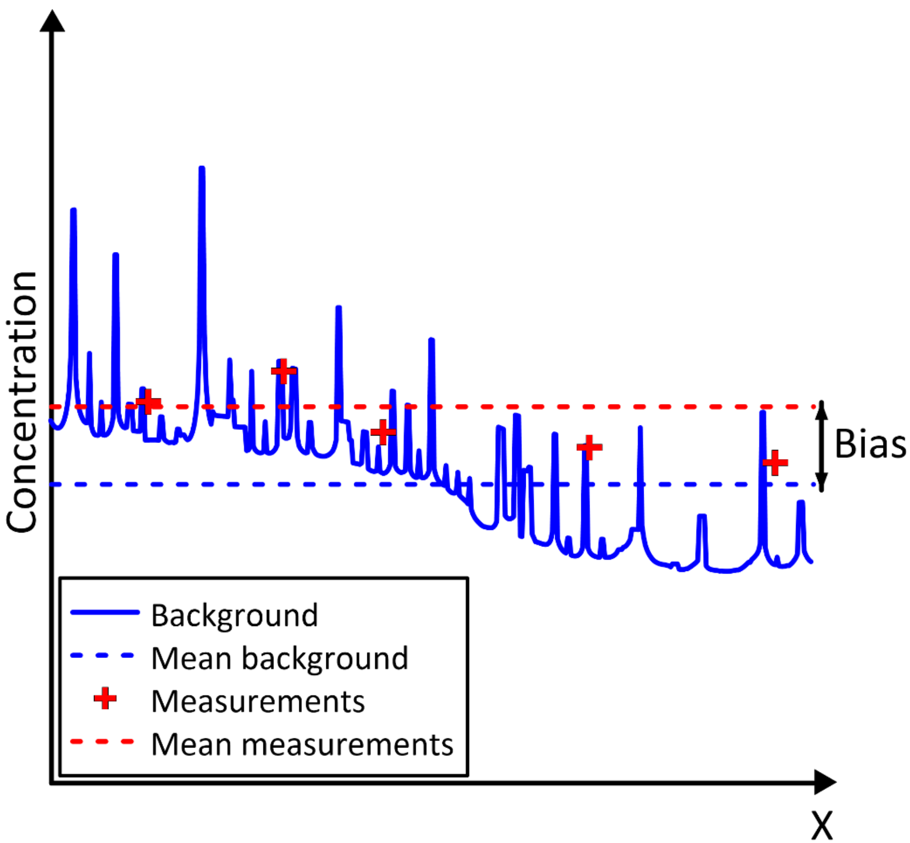
- $\mathbf{x}^{\mathbf{b}}$: background (n)
- $\mathbf{y}$: observations (m)
- $\mathbf{x}^{\mathbf{a}}$: analysis (n)
- $\mathbf{H}$: observation operator (m x n)

DATA ASSIMILATION

Data assimilation

Bias Adjustment Technique (BAT)

a) Before assimilation
b) After assimilation



- Analysis: $\mathbf{x^a} = \alpha \mathbf{x^b}$
- Correction coefficient: $\alpha = \dfrac{\sum_i^m y_i}{\sum_i^m x_i^b}$

with:

- x_i^b: background at point p_i
- y_i: measurement at point p_i
- m: number of observations

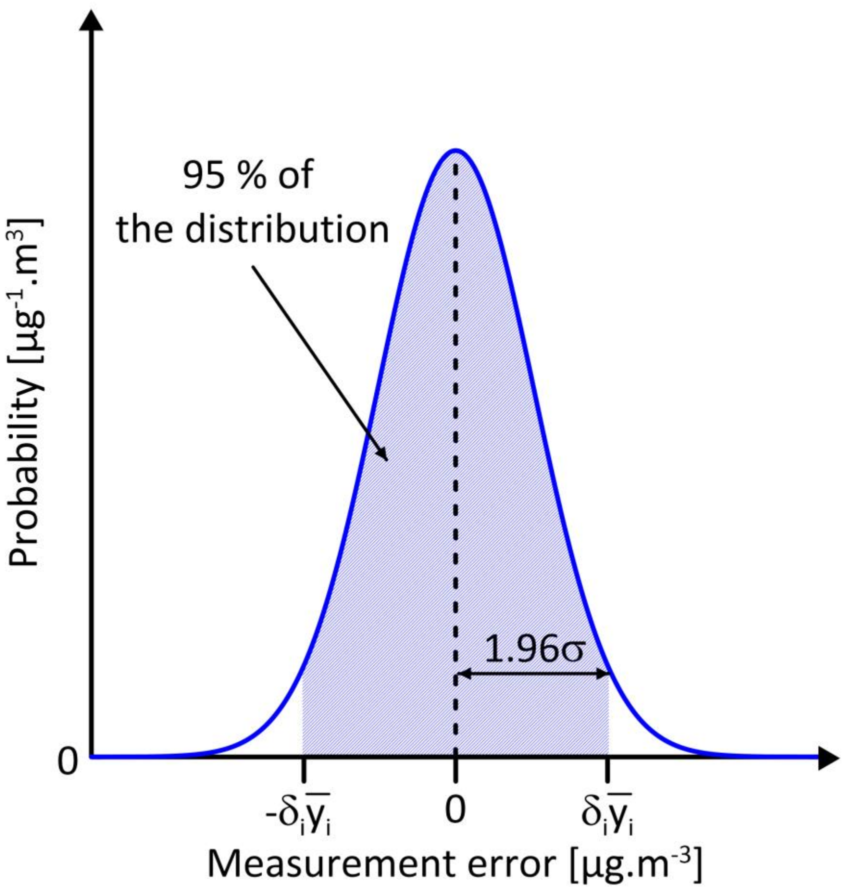
Data assimilation

Best Linear Unbiased Estimator (BLUE)

- Analysis: $\mathbf{x^a} = \mathbf{x^b} + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x^b}\right)$
- Kalman gain: $\mathbf{K} = \mathbf{B}\mathbf{H}^{T}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{T} + \mathbf{R}\right)^{-1}$

with:

- $\mathbf{K}$: Kalman gain
- $\mathbf{R}$: observation error covariance matrix
- $\mathbf{B}$: background error covariance matrix



- Modelling of matrix $\mathbf{R}$:
  - $\mathbf{R} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2)$
  - $1{,}96\sigma_i = \delta_i \overline{y}_i$

  with:
  - $\overline{y}_i$: mean measurement at point p_i
  - δ_i: uncertainty at the point p_i

Data assimilation

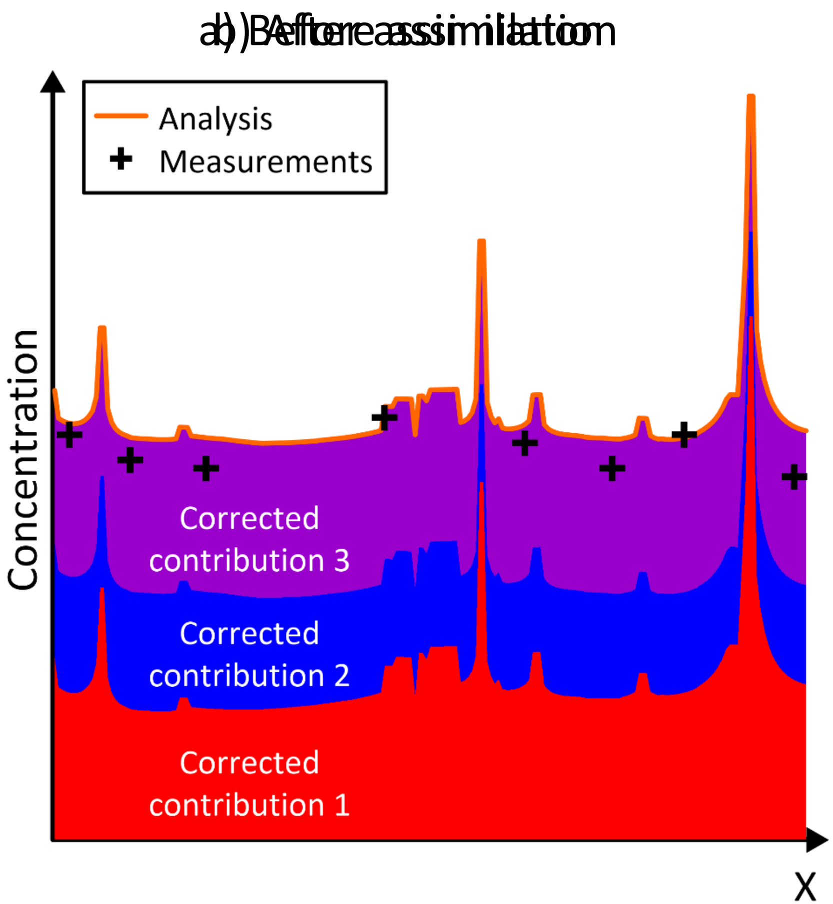
Best Linear Unbiased Estimator (BLUE)

- Modelling of matrix $\mathbf{B}$:
  - **Assumption**: background errors at points $\mathrm{p_i}$ and $\mathrm{p_j}$ are more correlated when these points are impacted by the same events
  - $\mathrm{B_{ij}} = \gamma\sqrt{\sigma_i^{2,b}\sigma_j^{2,b}}\rho_0 \exp\left(\frac{\rho_{ij}^b - 1}{L_\rho}\right)$

  with:
    - $\sigma_i^{2,b}$: background variance at point $\mathrm{p_i}$
    - ρ_{ij}^b: correlation coefficient of the background at points $\mathrm{p_i}$ and $\mathrm{p_j}$
    - γ: adjustment coefficient
    - ρ_0: characteristic correlation coefficient
    - L_ρ: characteristic correlation distance
  - γ, ρ_0 and L_ρ are estimated with the χ^2 diagnosis and by minimising the RMSE after cross-validation

Data assimilation

Source Apportionment Least Square (SALS)



- **Assumption**: modelling errors are mainly due to errors on emissions estimates

- Analysis: $\mathbf{x^a} = \sum_{\mathrm{g}}^{\mathrm{G}} \beta_g \, \mathbf{x_g^b}$

with:

  - $\mathbf{x_g^b}$: background of the source group g
  - β_g: modulation coefficient of the source group g
  - G: number of source groups

- The β_g coefficients are estimated by minimising the cost function J:

$$J(\beta_1, \beta_2, \dots, \beta_{\mathrm{G}}) = (\mathbf{y} - \mathbf{x^a})^T (\mathbf{y} - \mathbf{x^a})$$

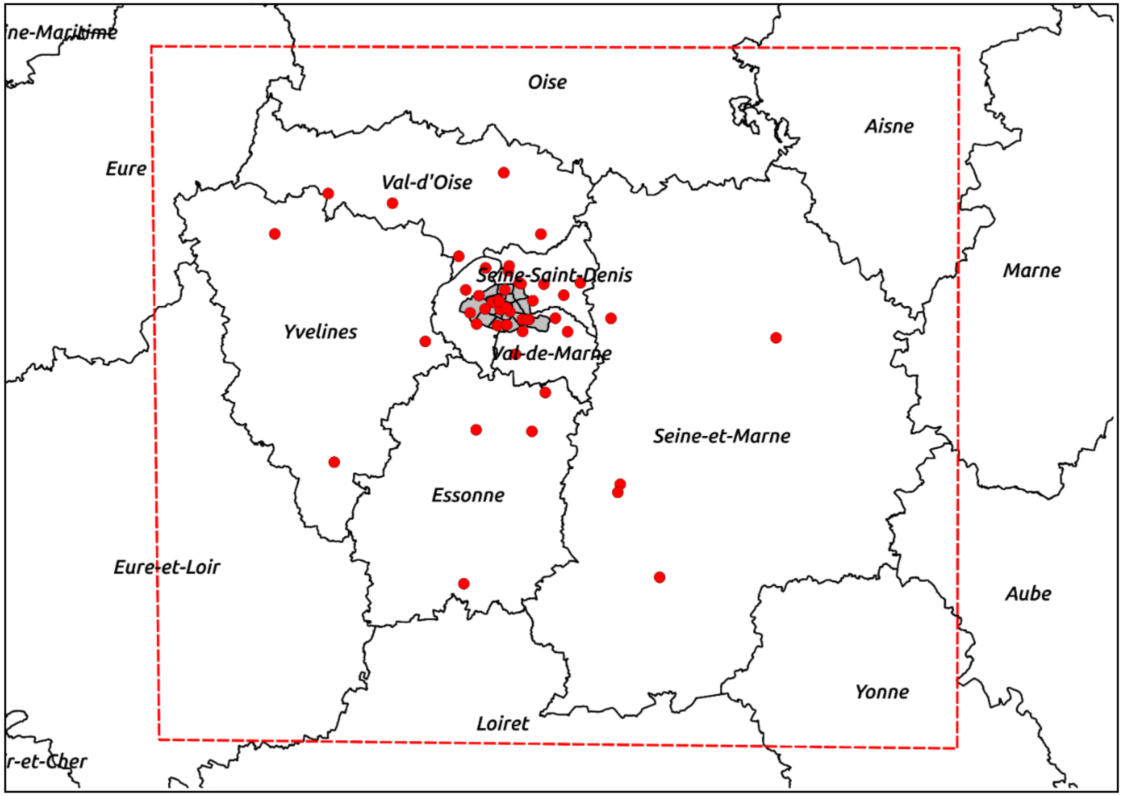
CASE STUDY

Case study
Description

- **Goal: air quality evaluation on the Île-de-France area**

- Scenario:
  - From 01/12/16 to 30/06/17
  - Pollutant: NO_2
  - 35 monitoring stations

- 3 groups for the SALS:
  - Traffic
  - Other emissions
  - Background concentration



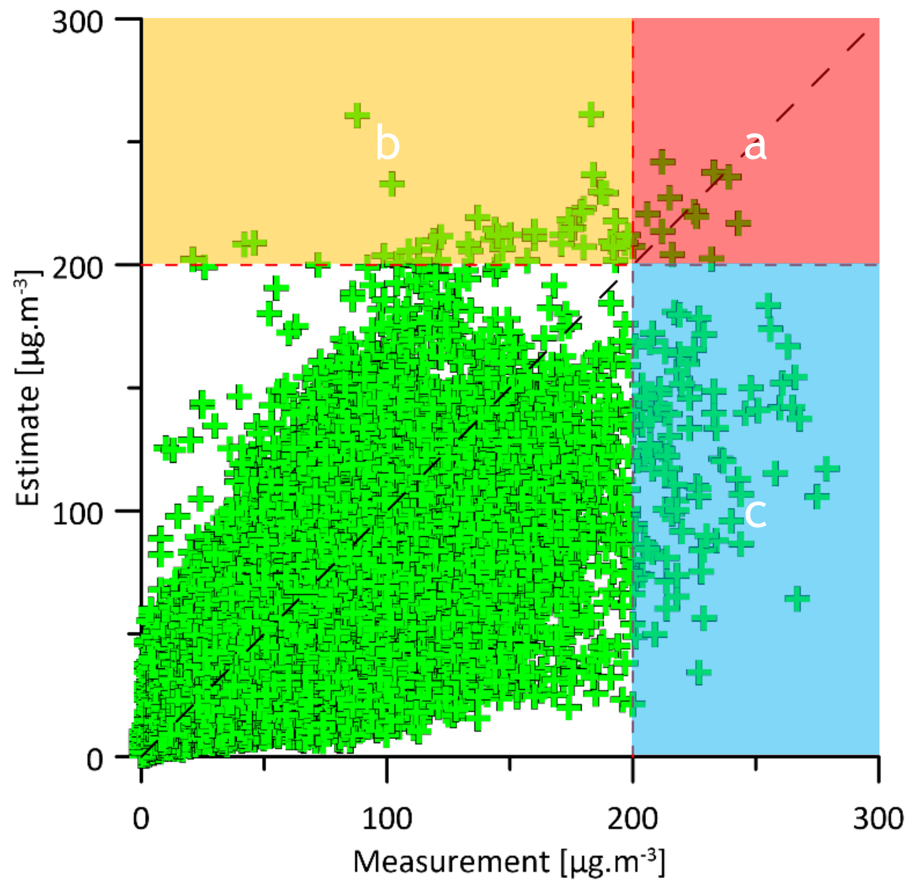
Domain of the case study

Case study

Statistical indices

Statistical indices	Expression	Optimal value
Bias	$\overline{\mathbf{x^m}} - \bar{\mathbf{y}}$	0
RMSE	$\sqrt{\overline{(\mathbf{x^m} - \mathbf{y})^2}}$	0
Corr	$\dfrac{\overline{(\mathbf{x^m} - \overline{\mathbf{x^m}})(\mathbf{y} - \bar{\mathbf{y}})}}{\sqrt{\overline{(\mathbf{x^m} - \overline{\mathbf{x^m}})^2}\ \overline{(\mathbf{y} - \bar{\mathbf{y}})^2}}}$	1
POD	$\dfrac{\mathrm{a}}{\mathrm{a + c}}$	1
FAR	$\dfrac{\mathrm{b}}{\mathrm{a + b}}$	0

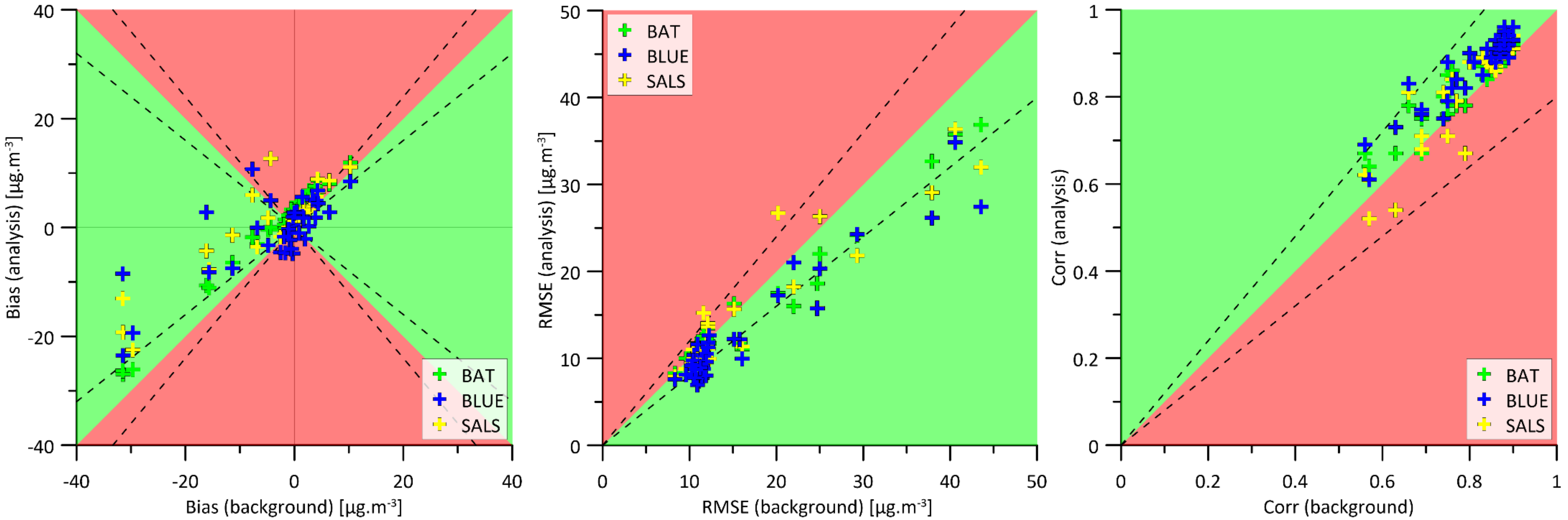
$\mathbf{x^m}$: estimates; $\mathbf{y}$: observations



The analysis is estimated with the leave-one-out cross-validation

Case study

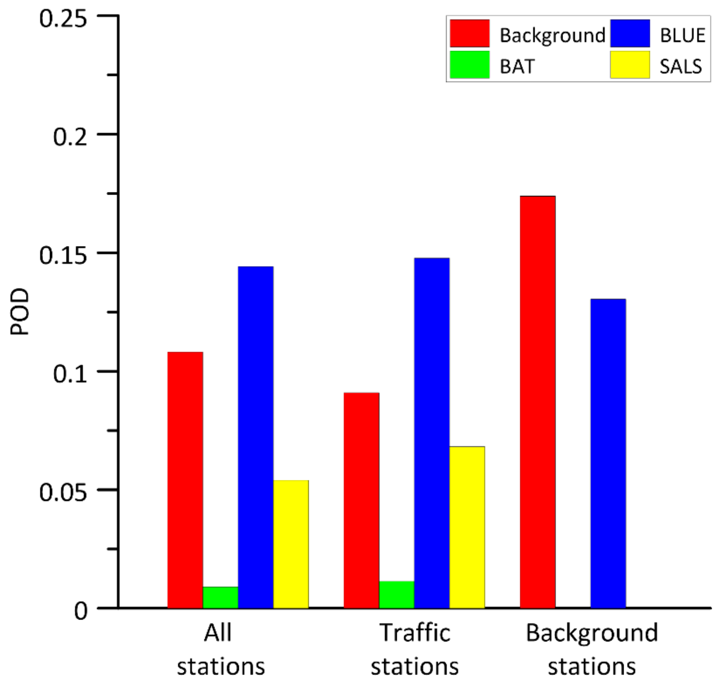
Results: Bias, RMSE and Corr

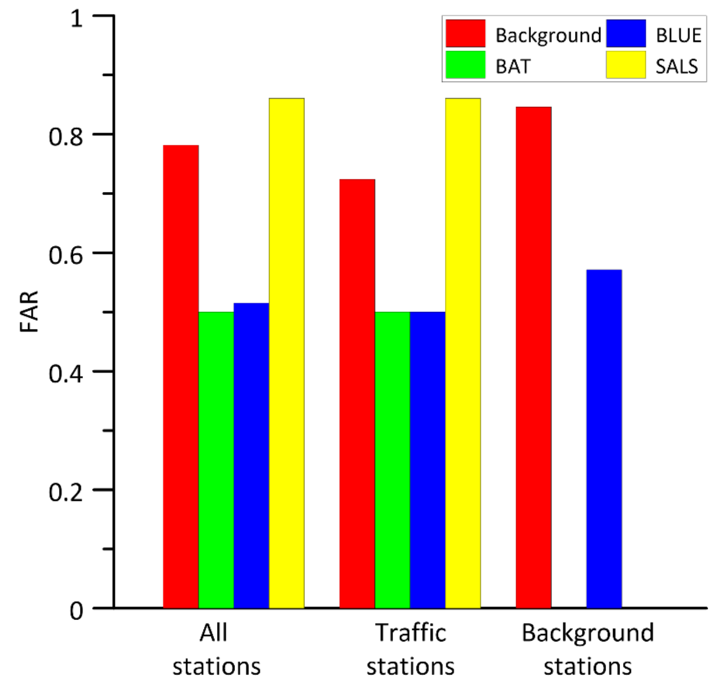


- Mixed results for the Bias
- Improvement of the RMSE (≈ 20 %) and Corr (≈ 10 %)
- The BLUE method leads to the best results

Case study

Results: POD and FAR

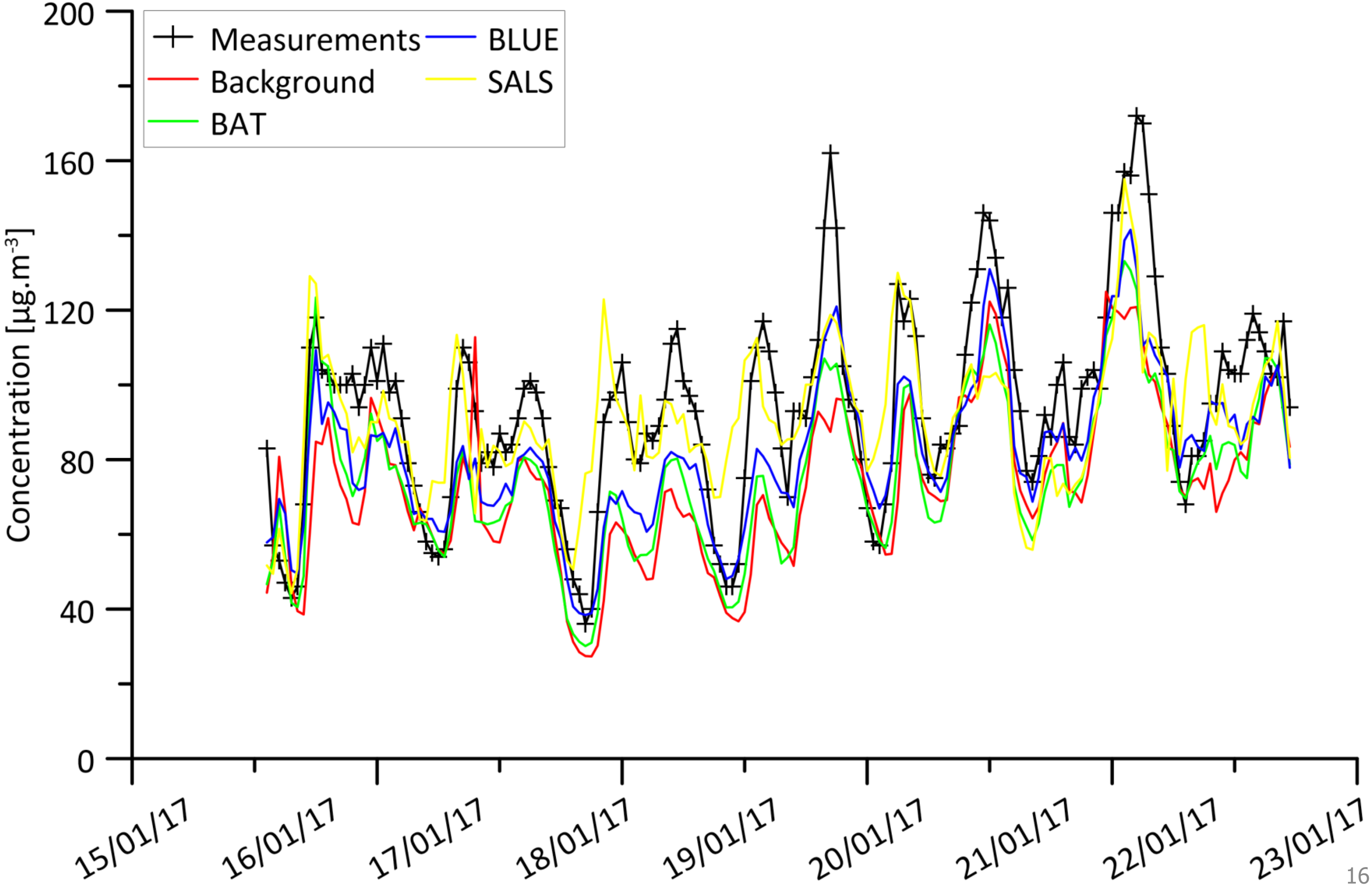




- Improvement of high concentration (> 200 µg.m^{-3}) detection with the BLUE method
  - Increase of the POD from 36 % to 67 % (except for the background stations)
  - Decrease of the FAR from 30 % to 36 %
- However, a significant number of high concentrations remain undetected

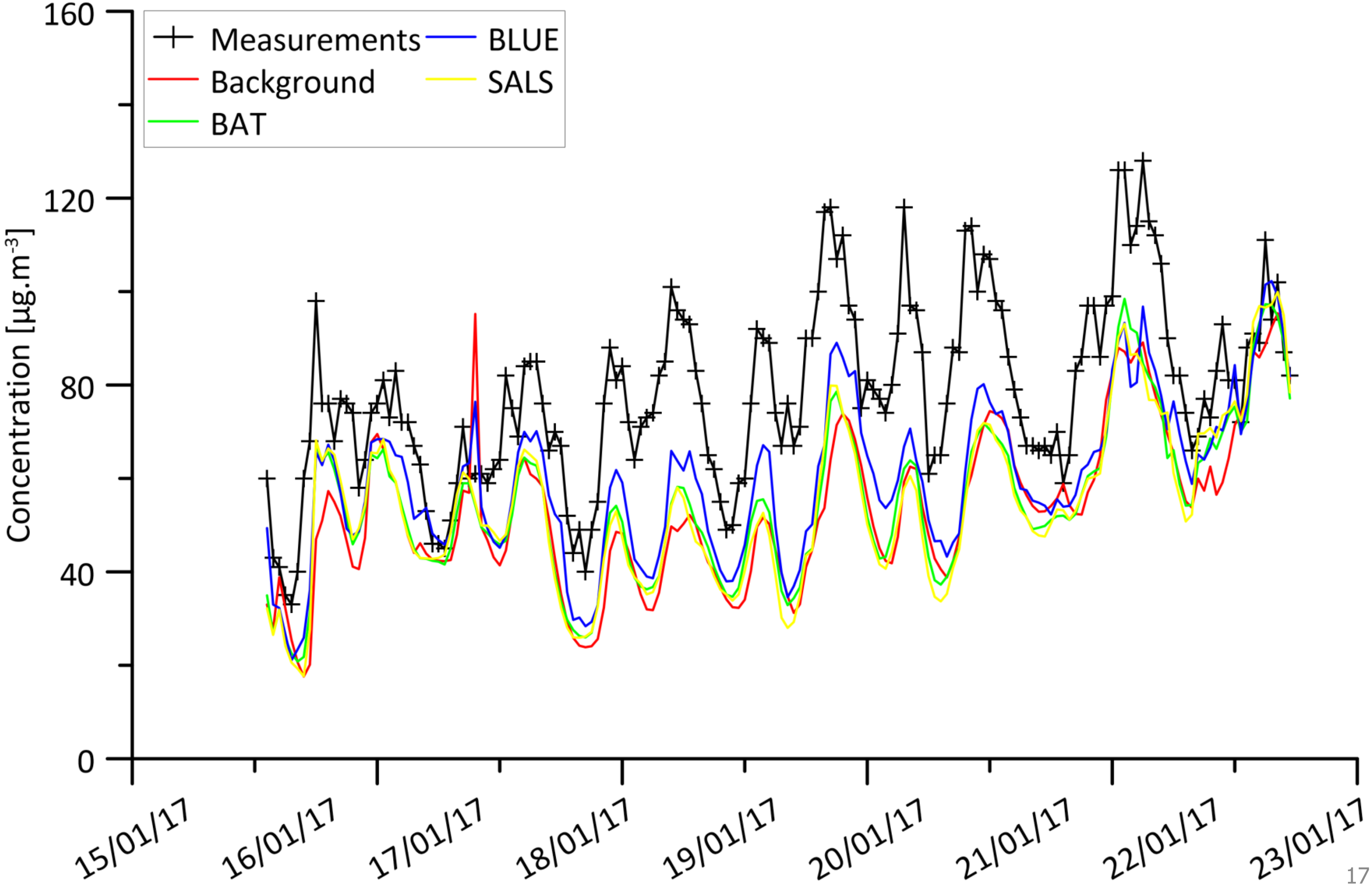
Case study

Results: hourly concentrations (A1)



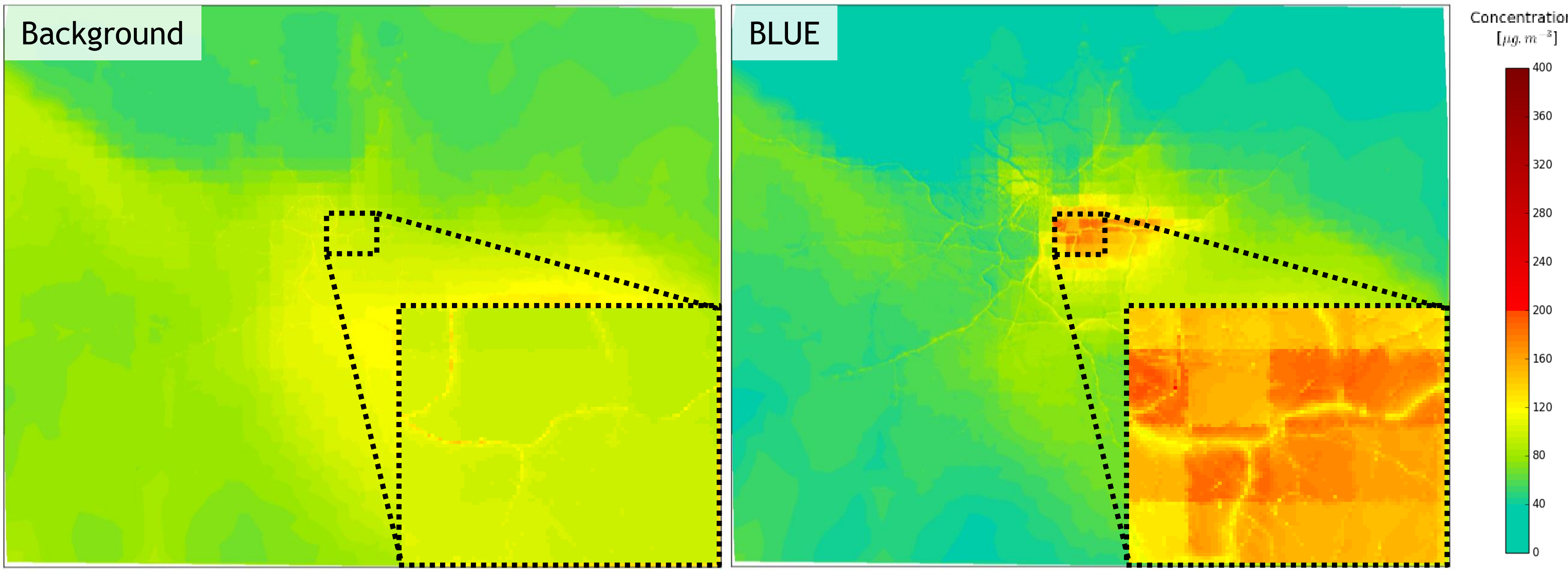
Case study

Results: hourly concentrations (RN2)



Case study

Results: concentration fields (0h 02/12/2016)



- **The BLUE method can lead to concentration fields which are not physically consistent because:**
  - This method is a statistical method which is not governed by physical laws
  - This method is an interpolation of the innovation
  - The matrix $\mathbf{B}$ has a *monotonous behavior* regardless of the innovation

CONCLUSION

Conclusions

- Data assimilation:
  - Global improvement of the statistical indices
  - Sometimes an improvement of the *high* concentration detection
  - Occasionally the estimates are worse after DA

- Performances of the 3 DA methods:
  - Globally the BLUE method leads to the best results
  - The best estimates are not always associated to the same method temporally and spatially

- The BLUE method can lead to concentration fields which are not physically consistent

ÉCOLE CENTRALE LYON

Acknowledgements

This work was done in collaboration with Airparif in the framework of the FAIRCITY project funded by the FUI and the Région Auvergne-Rhône-Alpes

Thank you for your attention ☺

Questions ?