

Émulation statistique d'un modèle statique de qualité de l'air à l'échelle urbaine

Vivien Mallet^{1,2} Anne Tilloy^{1,2} David Poulet³
Fabien Brocheton³

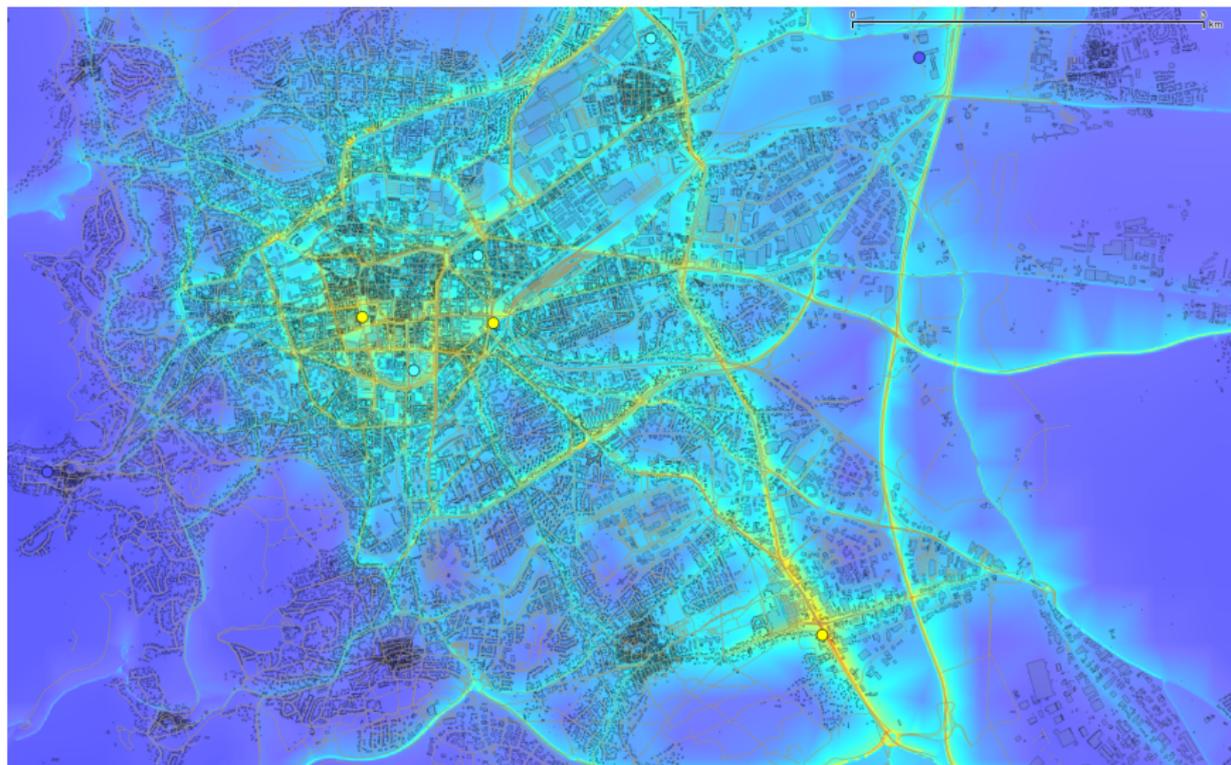
¹INRIA

²CEREA, laboratoire commun ENPC - EDF R&D, université Paris-Est

³NUMTECH

Rencontre SAMA, École normale supérieure, Paris, février 2014

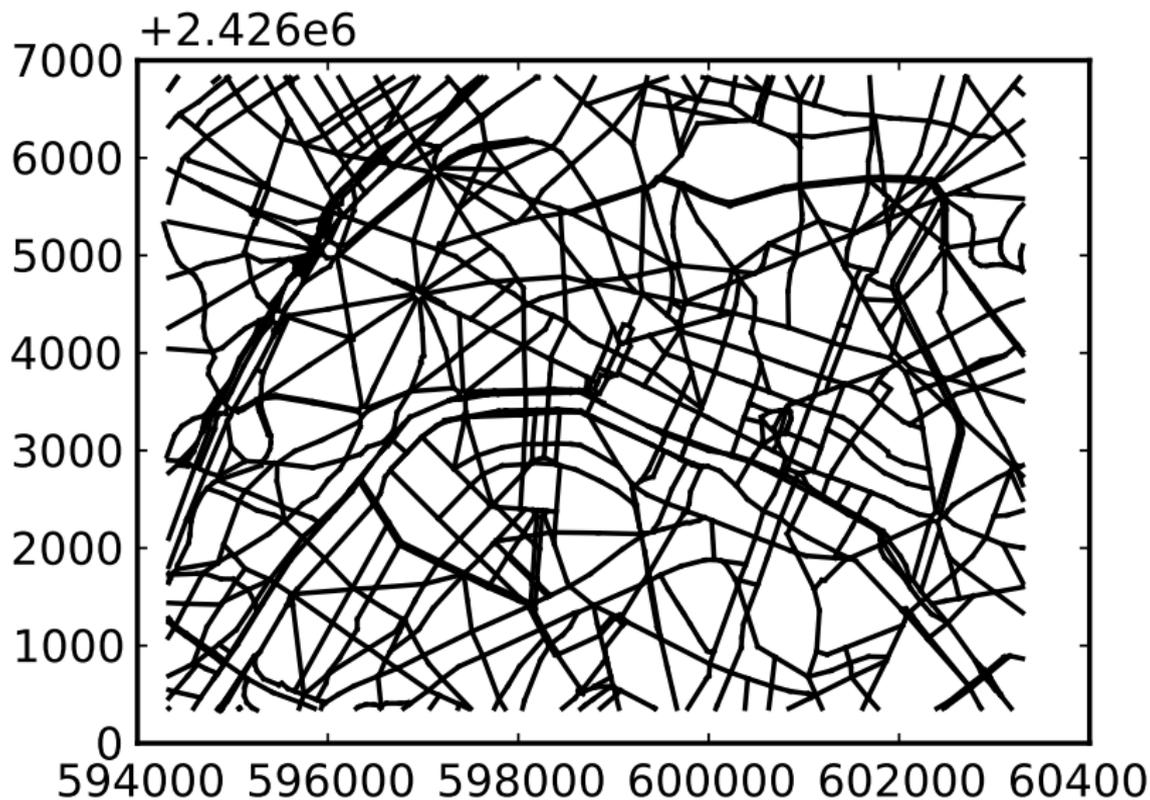
Simulation de la qualité de l'air à l'échelle urbaine



Concentration de dioxyde d'azote, NO_2 , dans l'air extérieur, à la résolution des rues sur Clermont-Ferrand à une heure et date données.

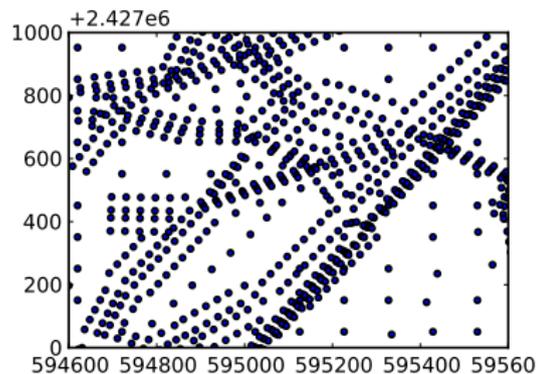
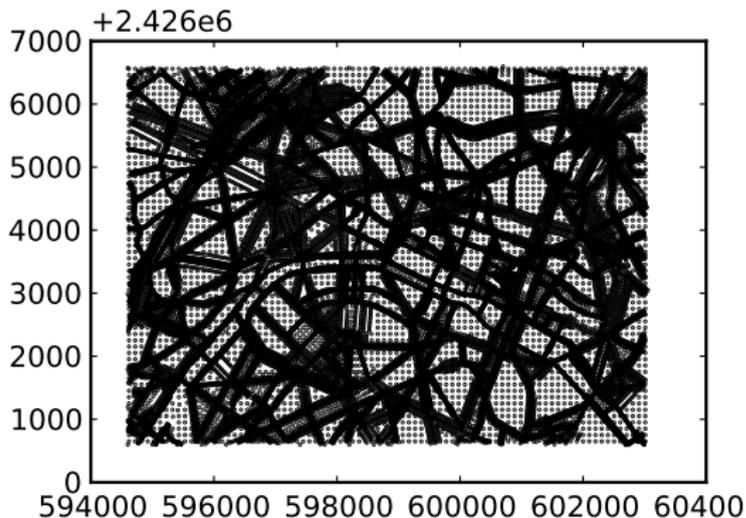
Simulation de qualité de l'air à l'échelle urbaine

Réseau routier émetteur, sur une partie de Paris (« Votre Air », avec Airparif)



Simulation de qualité de l'air à l'échelle urbaine

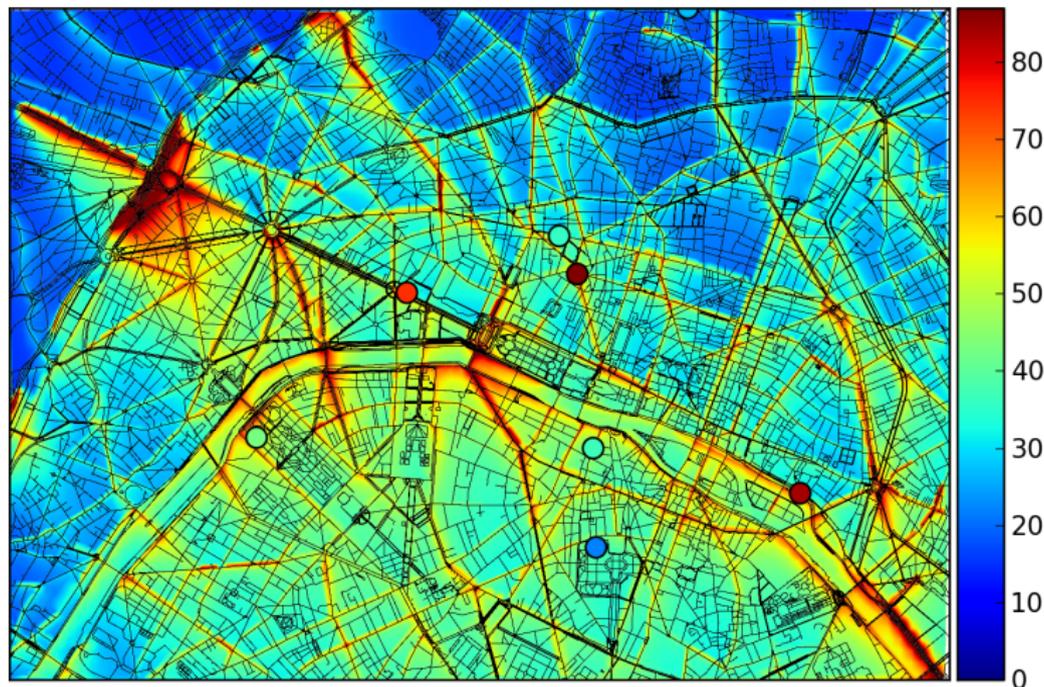
Sorties du modèle sur une partie de Paris (« Votre Air », avec Airparif)



Simulation de qualité de l'air à l'échelle urbaine

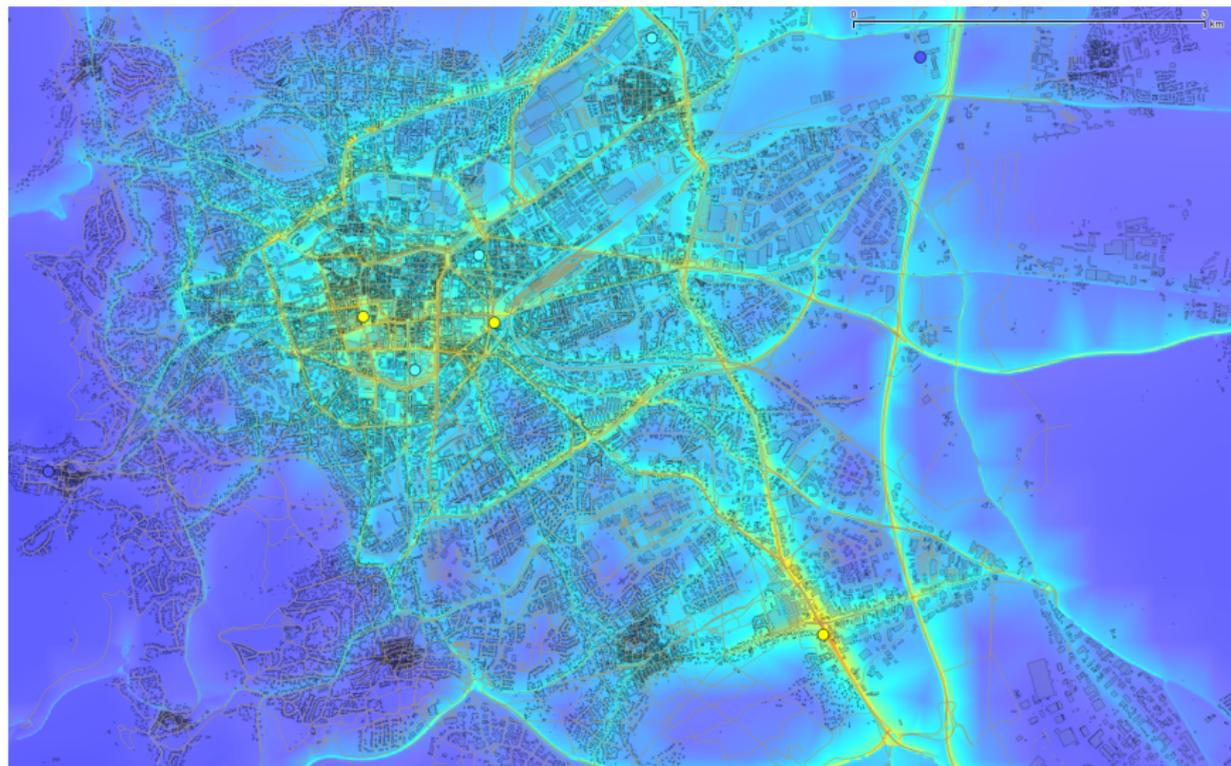
Concentration de NO_2 en $\mu\text{g}/\text{m}^3$ sur une partie de Paris (« Votre Air », avec Airparif)

2011-06-29 19:00:00



Simulation de qualité de l'air à l'échelle urbaine

Concentration de NO_2 en $\mu\text{g}/\text{m}^3$ sur Clermont-Ferrand



Simulation de qualité de l'air à l'échelle urbaine

Formulation du modèle

$$\mathbf{y} = \mathcal{M}(\mathbf{p}) .$$

Entrées \mathbf{p}

Dix scalaires : vitesse du vent, direction du vent, température, nébulosité, précipitations, jour de l'année, heure, $[\text{NO}_2]_{\text{fond}}$, $[\text{NO}_x]_{\text{fond}}$, $[\text{O}_3]_{\text{fond}}$.

Modèle \mathcal{M}

Temps de calcul : 4 minutes sur 16 cœurs pour une échéance (donc, une heure).

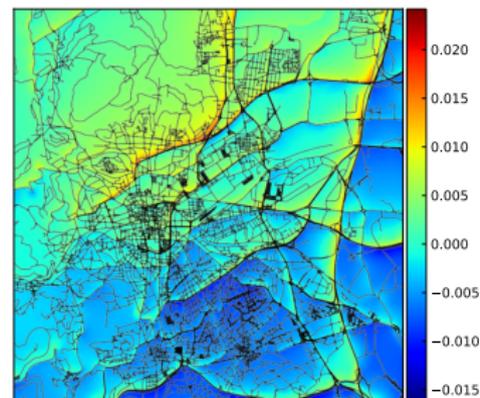
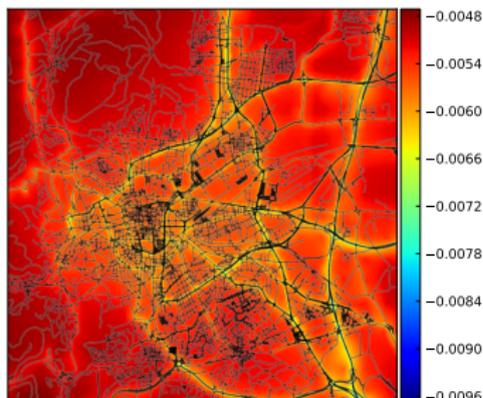
Sorties \mathbf{y}

$$\mathbf{y} \in \mathbb{R}^{3 \times 10^4} .$$

Objectif : construction d'un modèle $\widehat{\mathcal{M}}$ peu coûteux tel que $\widehat{\mathcal{M}}(\mathbf{p}) \simeq \mathcal{M}(\mathbf{p})$.

Première étape : réduction de dimension

- On projette les sorties du modèle dans un sous-espace engendré par les N colonnes Ψ_j de Ψ , avec N petit.
- La base peut être apprise par analyse en composante principale sur un premier échantillon d'apprentissage de moyenne $\bar{\mathbf{y}}$.



Première étape : réduction de dimension

- On projette les sorties du modèle dans un sous-espace engendré par les N colonnes Ψ_j de Ψ , avec N petit.
- La base peut être apprise par analyse en composante principale sur un premier échantillon d'apprentissage de moyenne $\bar{\mathbf{y}}$.
- On obtient

$$\mathbf{y} \simeq \bar{\mathbf{y}} + \Psi\Psi^T(\mathcal{M}(\mathbf{p}) - \bar{\mathbf{y}}).$$

- La fluctuation autour de la moyenne est donnée dans le sous-espace par le modèle réduit

$$f(\mathbf{p}) = \Psi^T \mathcal{M}(\mathbf{p}).$$

- Ce modèle réduit (en dimension) a un temps de calcul aussi élevé que le modèle complet.

Application : $N = 11$ modes.

Deuxième étape : émulation statistique du modèle réduit

- On cherche un substitut rapide $\hat{f}_j(\mathbf{p})$ à chaque $f_j(\mathbf{p}) = \Psi_j^T \mathcal{M}(\mathbf{p})$.
- On génère un ensemble de points d'apprentissage $\mathbf{p}^{(i)}$, $i = 1, \dots, M$, typiquement en échantillonnant par hypercube latin.
- On applique le modèle réduit $f_j(\mathbf{p}^{(i)})$.
- On construit un émulateur du modèle réduit :

$$\hat{f}_j(\mathbf{p}) = \underbrace{\sum_{k=1}^K \beta_{j,k} p_k}_{\text{régression}} + \underbrace{\sum_{i=1}^M w_{j,i}(\mathbf{p}, \mathbf{p}^{(1)}, \dots, \mathbf{p}^{(M)}) \left(f_j(\mathbf{p}^{(i)}) - \sum_{k=1}^K \beta_{j,k} p_k^{(i)} \right)}_{\text{interpolation des résidus}} .$$

Application : $M = 2000$ tirages.

Deuxième étape : émulation statistique du modèle réduit

Exemple : par krigeage

$$\widehat{f}_j(\mathbf{p}) = \underbrace{\sum_{k=1}^K \beta_{j,k} p_k}_{\text{régression}} + \underbrace{\sum_{i=1}^M w_{j,i}(\mathbf{p}, \mathbf{p}^{(1)}, \dots, \mathbf{p}^{(M)}) \left(f_j(\mathbf{p}^{(i)}) - \sum_{k=1}^K \beta_{j,k} p_k^{(i)} \right)}_{\text{interpolation des résidus}}.$$

On note

- $\mathbf{P} = [\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(M)}]$,
- $\mathbf{f}_j = [f_j(\mathbf{p}^{(1)}), \dots, f_j(\mathbf{p}^{(M)})]^T$,
- $R(\mathbf{p}, \mathbf{q})$ la covariance paramétrée du processus gaussien sous-jacent,
- $\mathbf{r}(\mathbf{p}) = [R(\mathbf{p}^{(1)}, \mathbf{p}), \dots, R(\mathbf{p}^{(M)}, \mathbf{p})]^T$, et
- $\mathbf{R} = [R(\mathbf{p}^{(i)}, \mathbf{p}^{(j)})]_{(i,j)}$.

Alors, les poids du krigeage sont

$$\begin{aligned} [\beta_{j,1}, \dots, \beta_{j,K}]^T &= (\mathbf{P}\mathbf{P}^T)^{-1} \mathbf{P}\mathbf{R}^{-1} \mathbf{f}_j, \\ [w_{j,1}(\mathbf{p}), \dots, w_{j,M}(\mathbf{p})]^T &= \mathbf{r}^T(\mathbf{p}) \mathbf{R}^{-1}. \end{aligned}$$

Autres méthodes d'interpolation

- Fonctions de base radiales (« radial basis function », RBF)
- « Inverse distance weighting » avec adaptation à la grande dimension



V. R. JOSEPH et L. KANG (2011). « Regression-based inverse distance weighting with applications to computer experiments ».

Dans : *Technometrics* 53.3, p. 254–265. DOI :

10.1198/TECH.2011.09154

Application : krigeage avec DiceKriging.



O. ROUSTANT, D. GINSBOURGER et Y. DEVILLE (2012).
« DiceKriging, DiceOptim : Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization ». Dans :

Journal of Statistical Software 51.1, p. 1–55. URL :

<http://www.jstatsoft.org/v51/i01>

$$\mathbf{y} \simeq \bar{\mathbf{y}} + \boldsymbol{\Psi}\boldsymbol{\Psi}^T(\mathcal{M}(\boldsymbol{\rho}) - \bar{\mathbf{y}}) .$$

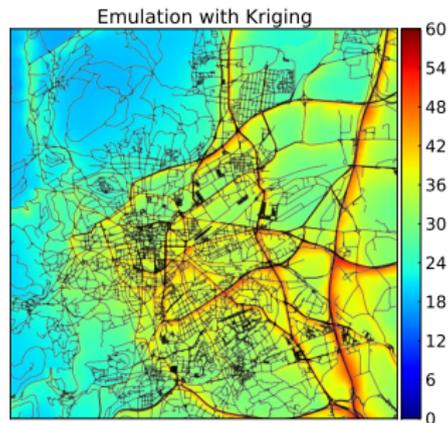
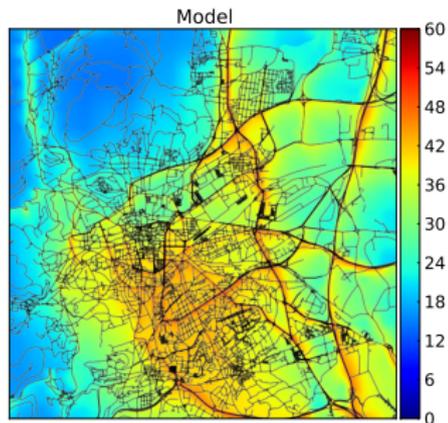
Avec l'émulation, on obtient finalement

$$\mathbf{y} \simeq \bar{\mathbf{y}} - \boldsymbol{\Psi}\boldsymbol{\Psi}^T\bar{\mathbf{y}} + \boldsymbol{\Psi} \begin{bmatrix} \hat{f}_1(\boldsymbol{\rho}) \\ \vdots \\ \hat{f}_N(\boldsymbol{\rho}) \end{bmatrix} .$$

Application : sur un cœur, coût de calcul de 50 millisecondes.

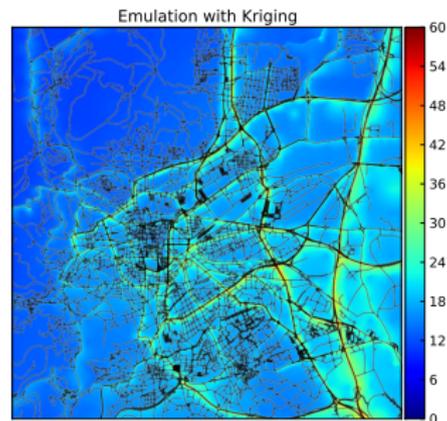
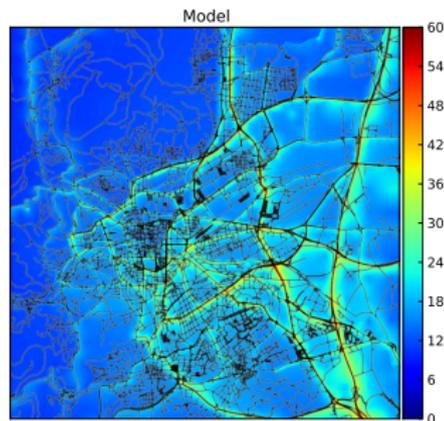
Modèle complet versus modèle réduit et émulé

Concentration de NO_2 en $\mu\text{g}/\text{m}^3$.



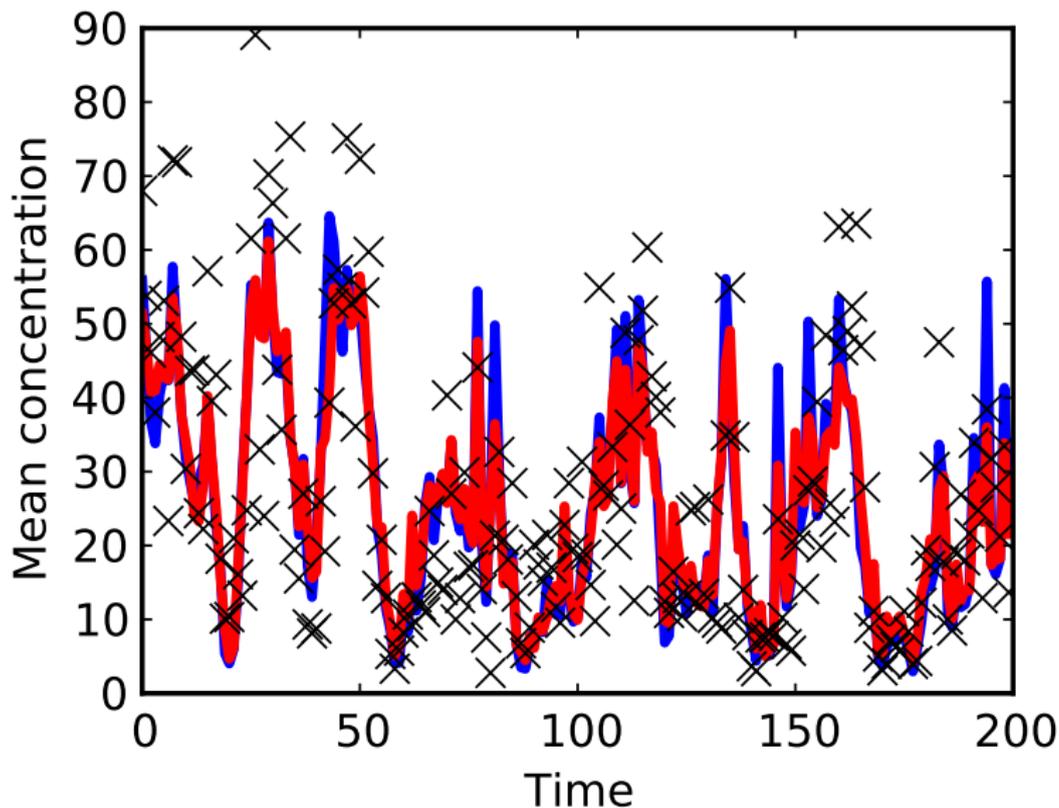
Modèle complet versus modèle réduit et émulé

Concentration de PM_{10} en $\mu\text{g}/\text{m}^3$.



Modèle complet, modèle réduit et émulé, et observations

Concentrations de NO_2 en $\mu\text{g}/\text{m}^3$, moyenne sur 9 stations.



Déclinaisons appliquées

- Insertion dans un outil d'aide à la décision
- Études d'impact

Quantification d'incertitude et assimilation de données

- Propagation fine des incertitudes
- Modélisation inverse
- Méthodes bayésiennes