

Couplage entre modèle et observation pour la simulation de la qualité de l'air

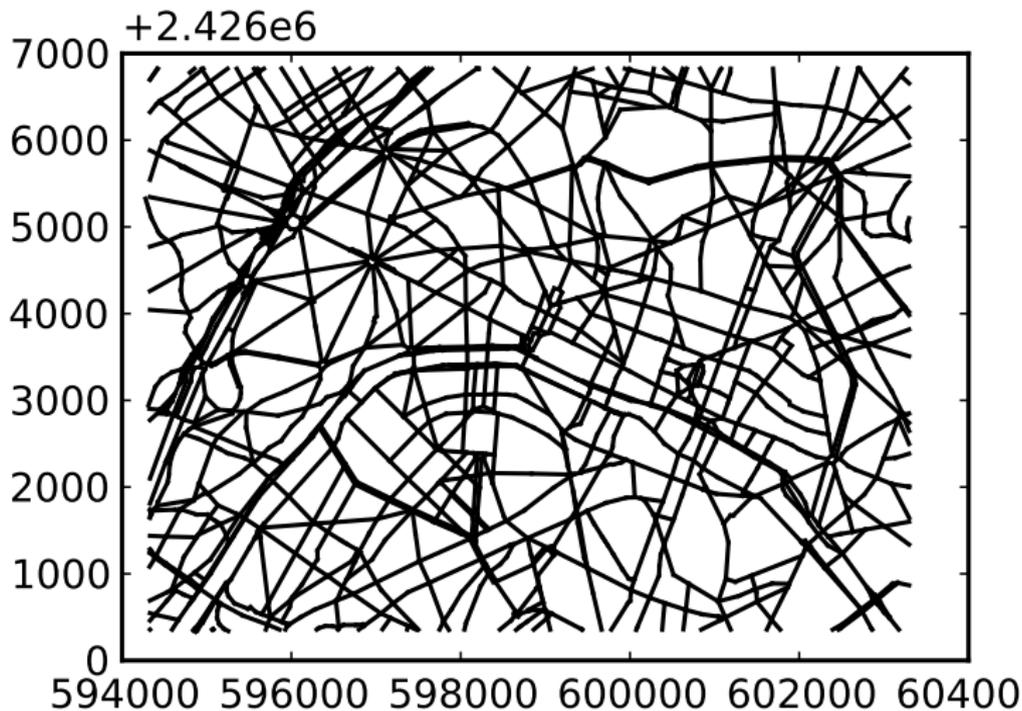
Vivien Mallet^{1,2}

¹Inria

Séminaire Aristote, mars 2015

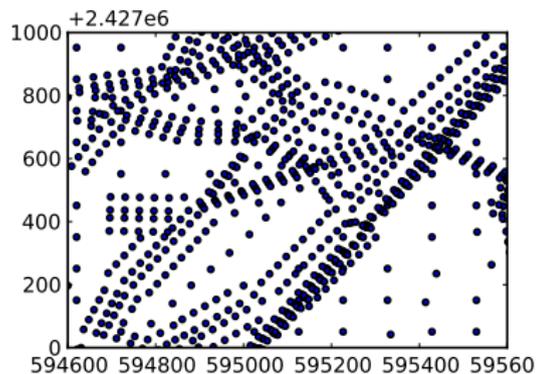
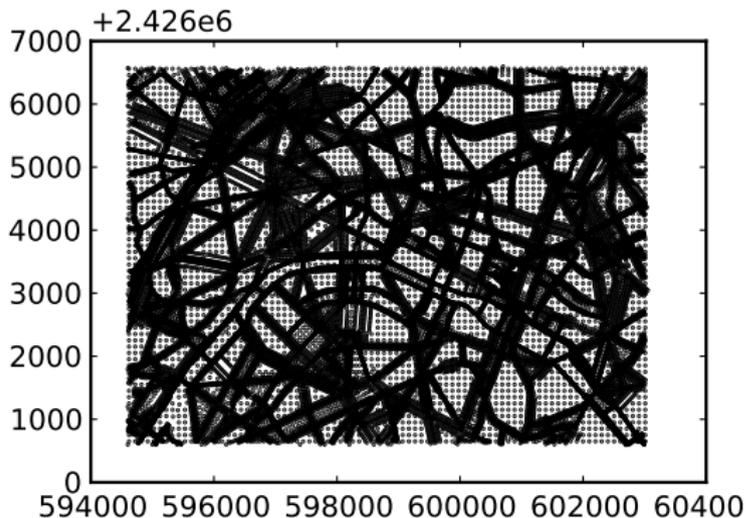
Réseau routier pour la modélisation du trafic

Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech



Points de sortie du modèle de qualité de l'air

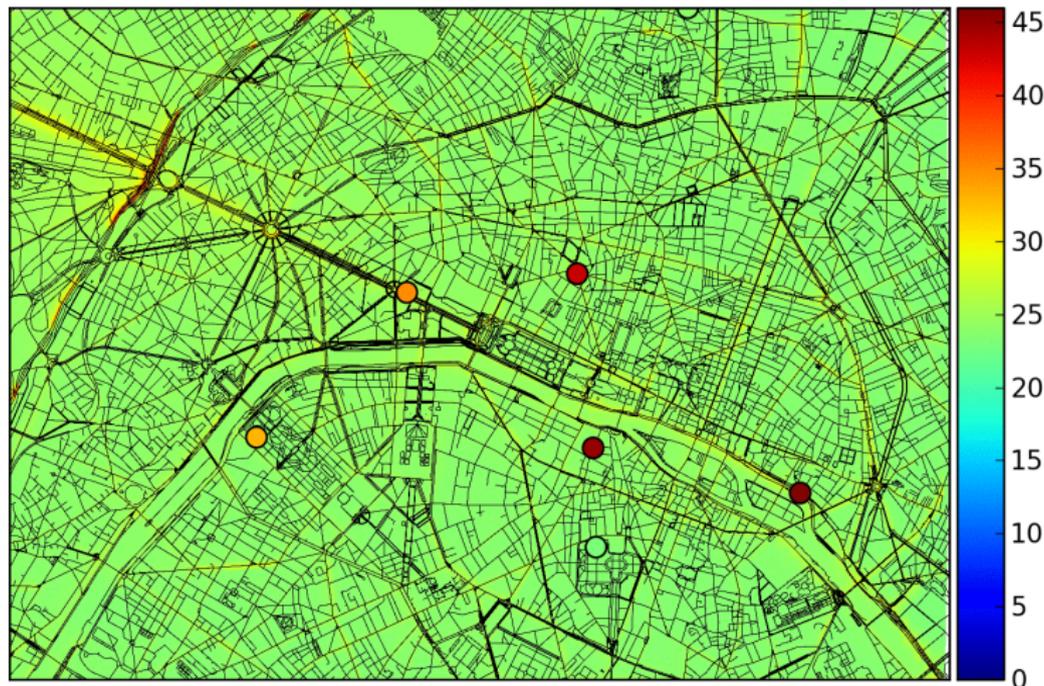
Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech



Qualité de l'air simulée

Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech

2011-06-26 07:00:00

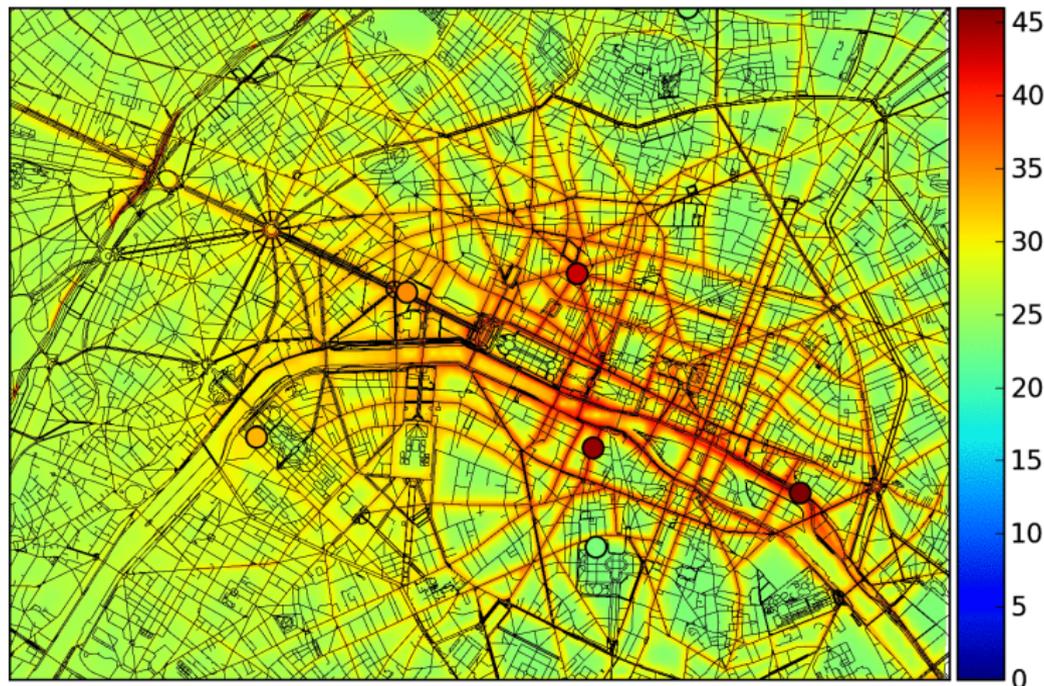


Carte de $[\text{NO}_2]$ ($\mu\text{g m}^{-3}$) simulée pour le 26 juin 2011 à 7 heures

Qualité de l'air après assimilation de données

Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech

2011-06-26 07:00:00



Carte de [NO₂] (µg m⁻³) analysée pour le 26 juin 2011 à 7 heures

Assimilation de données sur Paris

Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech

2011-06-26 07:00:00



Carte de [NO₂] (μg m⁻³) simulée

2011-06-26 07:00:00



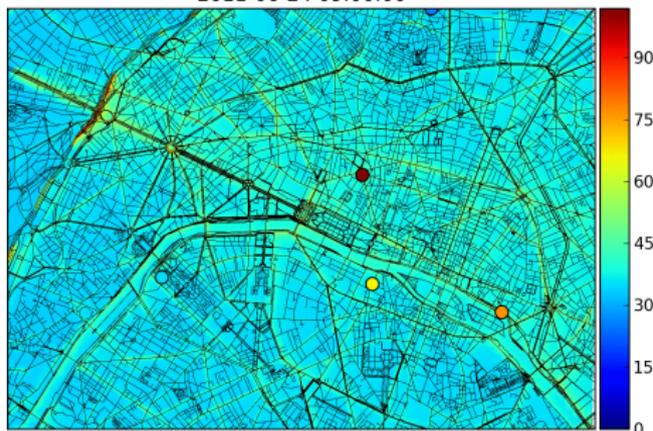
Carte de [NO₂] (μg m⁻³) analysée

26 juin 2011 à 7 heures

Assimilation de données sur Paris

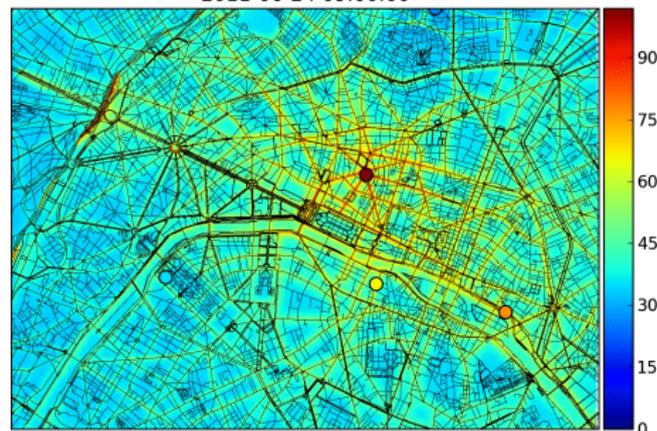
Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech

2011-06-24 09:00:00



Carte de [NO₂] (µg m⁻³) simulée

2011-06-24 09:00:00



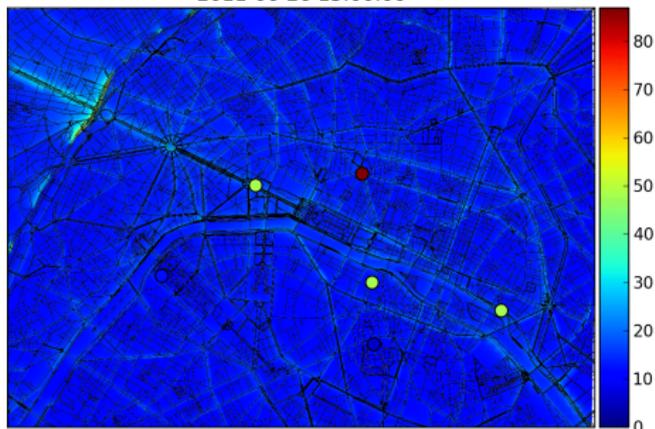
Carte de [NO₂] (µg m⁻³) analysée

24 juin 2011 à 9 heures

Assimilation de données sur Paris

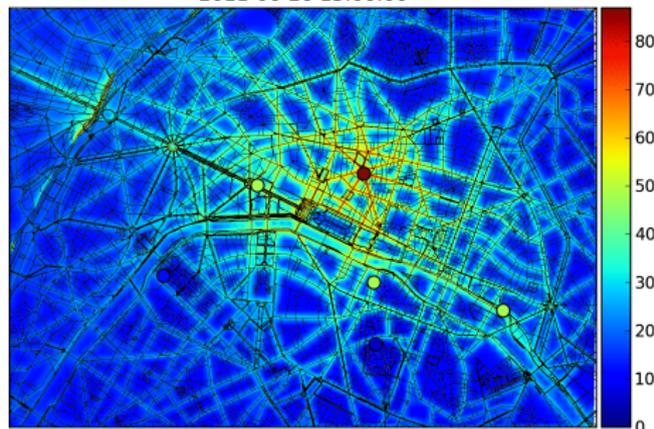
Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech

2011-06-26 15:00:00



Carte de [NO₂] (µg m⁻³) simulée

2011-06-26 15:00:00



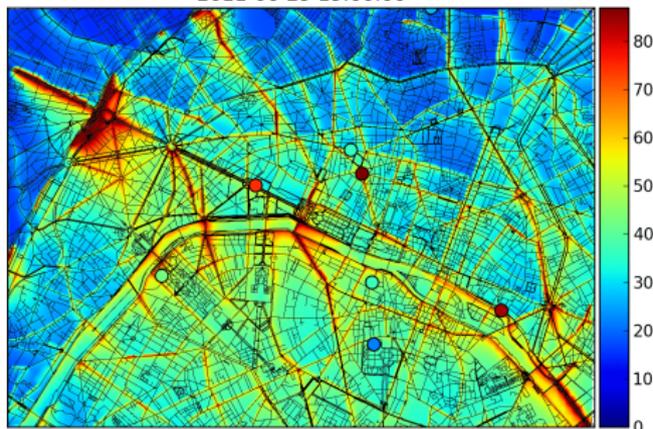
Carte de [NO₂] (µg m⁻³) analysée

26 juin 2011 à 15 heures

Assimilation de données sur Paris

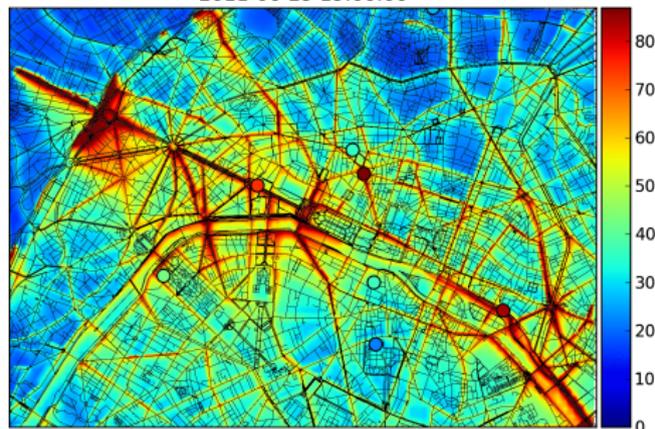
Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech

2011-06-29 19:00:00



Carte de [NO₂] (µg m⁻³) simulée

2011-06-29 19:00:00

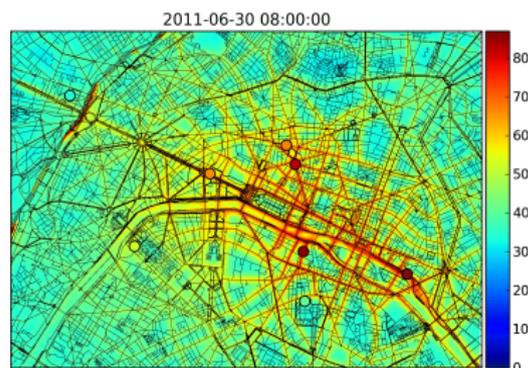
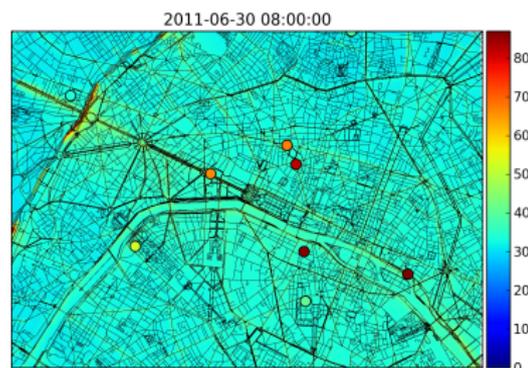


Carte de [NO₂] (µg m⁻³) analysée

29 juin 2011 à 19 heures

Assimilation de données sur Paris

Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech



Carte de $[\text{NO}_2]$ ($\mu\text{g m}^{-3}$) simulée

Carte de $[\text{NO}_2]$ ($\mu\text{g m}^{-3}$) analysée

30 juin 2011 à 8 heures

Logiciel transféré dans Urban Air System, commercialisé par Numtech auprès des collectivités ou associations de surveillance de la qualité de l'air. En cours de déploiement sur 12 villes.

Application mobile : indice de qualité de l'air

Exemple de « Votre Air » (partie de Paris), en collaboration avec Airparif et Numtech



Calculs sous-jacents

Meilleur estimateur linéaire sans biais

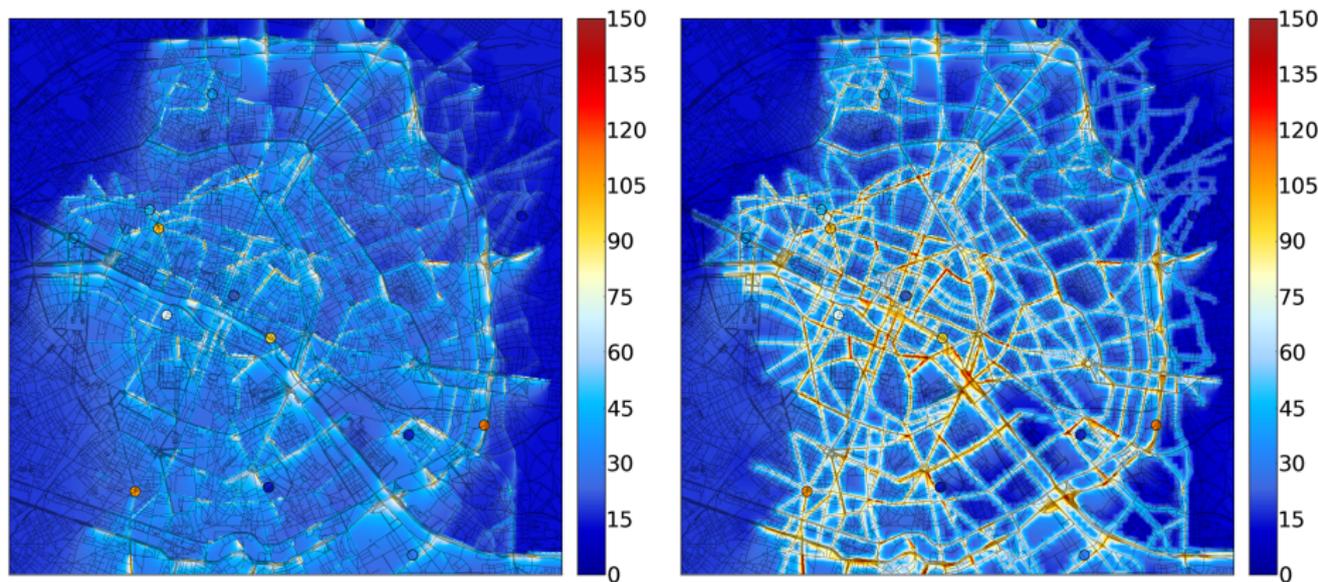
- Le modèle simule le vecteur \mathbf{c} avec une erreur supposée
 - D'espérance nulle
 - De variance \mathbf{B}
- Le vecteur des observations \mathbf{o} a une erreur
 - D'espérance nulle
 - De variance \mathbf{R}
 - Indépendante de l'erreur du modèle
- L'opérateur d'observation \mathbf{H} permet de comparer \mathbf{o} à \mathbf{Hc}
- L'estimateur linéaire, sans biais et d'erreur de variance minimale (trace), est

$$\mathbf{c}^* = \mathbf{c} + \mathbf{K}(\mathbf{o} - \mathbf{Hc})$$
$$\mathbf{K} = \mathbf{BH}^\top (\mathbf{HBH}^\top + \mathbf{R})^{-1}$$

Note : c'est l'étape de correction d'un filtre de Kalman.

- Cf. la bibliothèque d'assimilation de données Verdandi,
<http://verdandi.gforge.inria.fr/>

Application dans le projet PREQUALIF (porté par le LSCE)



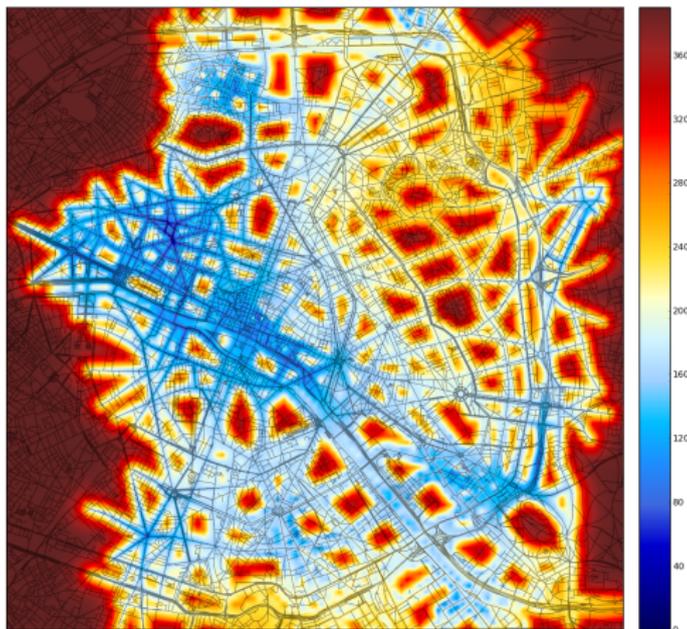
Carte de $[\text{NO}_2]$ ($\mu\text{g m}^{-3}$) simulée

Carte de $[\text{NO}_2]$ ($\mu\text{g m}^{-3}$) analysée

1er septembre 2012 à 18h

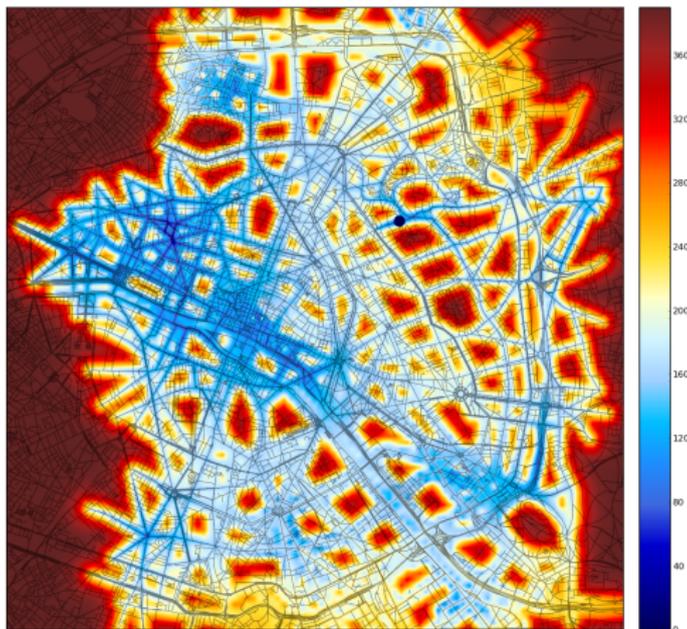
Ces travaux ont été appliqués sur la partie est de Paris et sur l'ensemble de l'Île-de-France, pour NO_2 , $\text{PM}_{2,5}$, PM_{10} et BC

Quantification de l'apport des observations



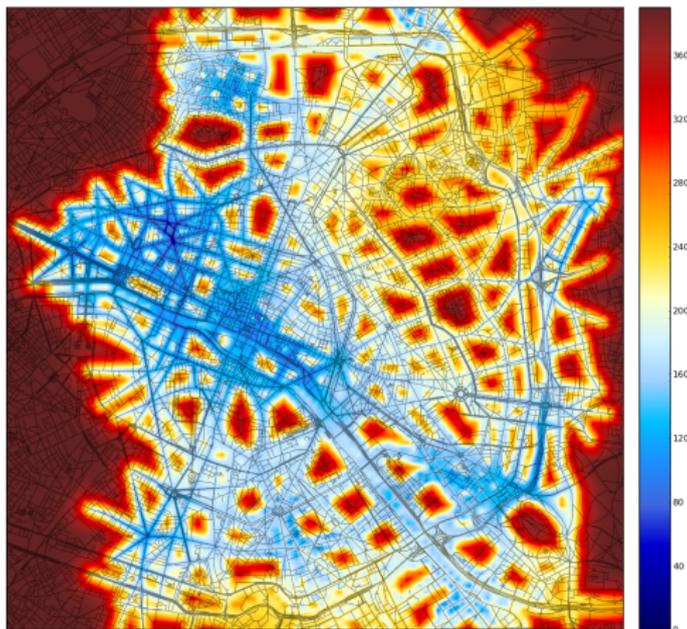
Variance ($\mu\text{g}^2\text{m}^{-6}$) de l'erreur a posteriori sur $[\text{NO}_2]$

Quantification de l'apport des observations



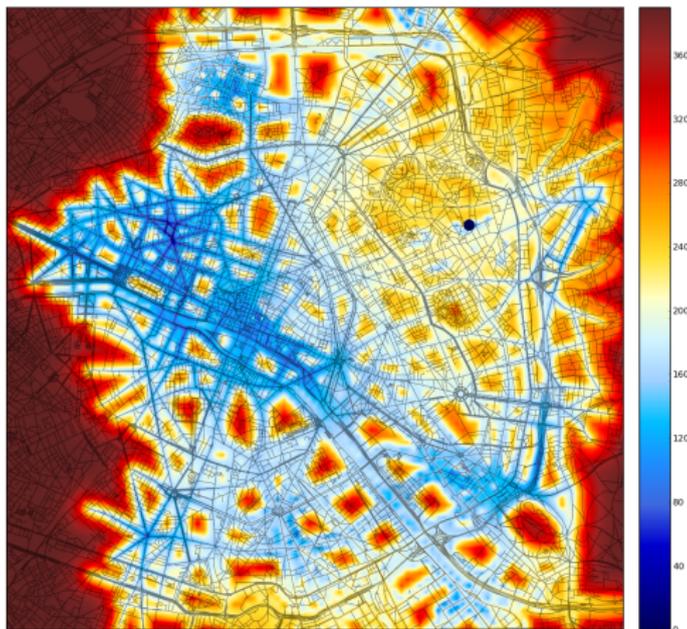
Variance ($\mu\text{g}^2\text{m}^{-6}$) de l'erreur a posteriori sur $[\text{NO}_2]$

Quantification de l'apport des observations



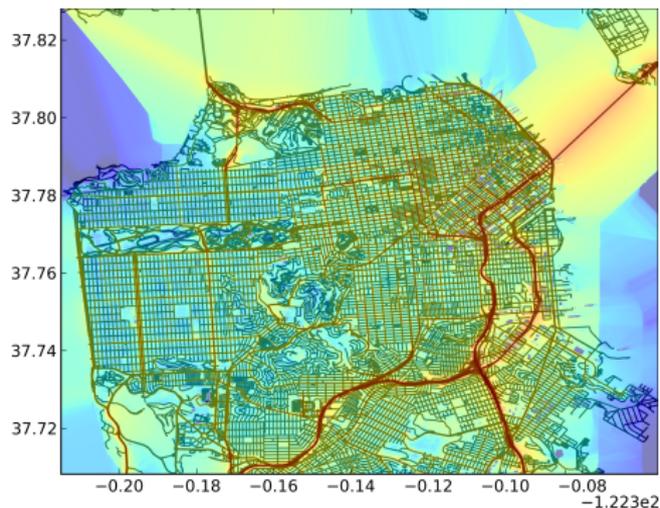
Variance ($\mu\text{g}^2\text{m}^{-6}$) de l'erreur a posteriori sur $[\text{NO}_2]$

Quantification de l'apport des observations



Variance ($\mu\text{g}^2\text{m}^{-6}$) de l'erreur a posteriori sur $[\text{NO}_2]$

Vers une application à la pollution sonore



Carte du bruit extérieur simulé à San Francisco

- Application à la génération de cartes de bruit en cours de réalisation
- Observation fournie par les téléphones portables
- Soutien de la ville de Paris
- En collaboration avec l'université de Berkeley, et déploiement à venir sur une ville californienne

Vers une plate-forme pour la modélisation, la simulation et l'observation de la ville

Constitution du groupe CityLab

- Regroupe 9 équipes-projets Inria
- Traite des réseaux de communication, de gouvernance, d'observation, l'analyse de masses de données, etc.
- <http://citylab.inria.fr/>

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

Formulation du modèle à réduire

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

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{c}

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

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

Réduction de dimension

- Modèle complet : $\mathbf{c} = \mathcal{M}(\mathbf{p})$
- Modèle réduit : $\boldsymbol{\alpha} = \boldsymbol{\Psi}^\top \mathcal{M}(\mathbf{p})$; d'où $\mathbf{c} \simeq \boldsymbol{\Psi} \boldsymbol{\Psi}^\top \mathcal{M}(\mathbf{p})$
- En pratique, $\mathbf{p} \in \mathbb{R}^{11}$ and $\boldsymbol{\alpha} \in \mathbb{R}^8$

Émulation

- Les dépendances entre les composantes de $\boldsymbol{\alpha}$ et \mathbf{p} sont lisses
- L'émulation consiste à construire un substitut m à $\boldsymbol{\Psi}^\top \mathcal{M}$
- Alors, $\mathbf{c} \simeq \boldsymbol{\Psi} m(\mathbf{p})$
- On interpole entre M tirages $\boldsymbol{\Psi}^\top \mathcal{M}(\mathbf{p}^{(i)})$

$$m_j(\mathbf{p}) = \sum_{i=1}^M w_{i,j}(\mathbf{p}, \mathbf{p}^{(1)}, \dots, \mathbf{p}^{(M)}) \boldsymbol{\Psi}_j^\top \mathcal{M}(\mathbf{p}^{(i)})$$

- Note : l'interpolation peut se faire par krigeage

Suite de la réduction/émulation

- ① Projet ANR ESTIMAIR (avec Numtech, LVMT, LMFA) : propagation des incertitudes dans un modèle de trafic et dans un modèle de qualité de l'air
- ② Analyse de sensibilité : quelles incertitudes expliquent celle sur c ?

Suite de l'assimilation de données

- ① Comment utiliser l'émulation dans le contexte de l'assimilation de données ?
- ② Quantification de l'incertitude a posteriori, par exemple pour mieux évaluer l'exposition des populations à la pollution
- ③ Modélisation inverse des émissions de CO₂ sur Paris et San Francisco
 - EcoLogicSense, Air Sûr, Numtech, CEA LETI, Airparif

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

Formulation du modèle à réduire

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

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).

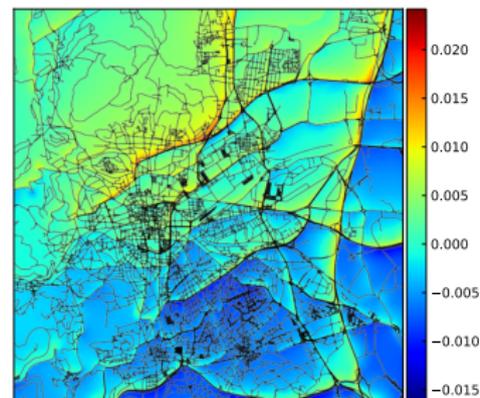
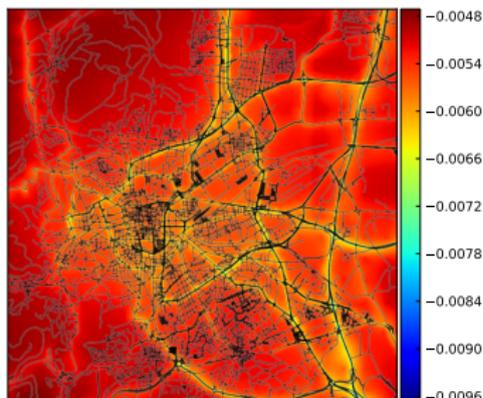
Sorties \mathbf{c}

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

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{c} .



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{c} .
- On obtient

$$c \simeq \bar{c} + \Psi\Psi^\top (\mathcal{M}(p) - \bar{c}) .$$

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

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

- 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 $f_j(\mathbf{p}) = \Psi_j^\top \mathcal{M}(\mathbf{p})$.
- 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

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

On note

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

Alors, les poids du krigeage sont

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

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 : krigage 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>

Troisième étape : reconstruction

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

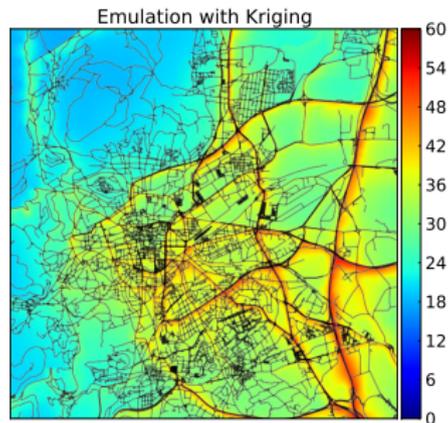
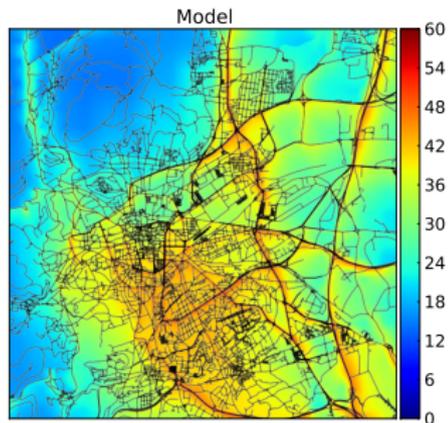
Avec l'émulation, on obtient finalement

$$\mathbf{c} \simeq \bar{\mathbf{c}} - \Psi\Psi^\top\bar{\mathbf{c}} + \Psi \begin{bmatrix} \hat{f}_1(\mathbf{p}) \\ \vdots \\ \hat{f}_N(\mathbf{p}) \end{bmatrix}.$$

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

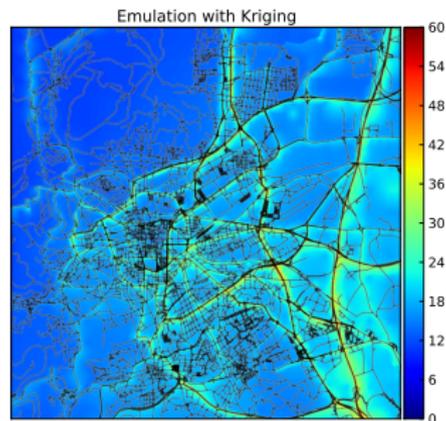
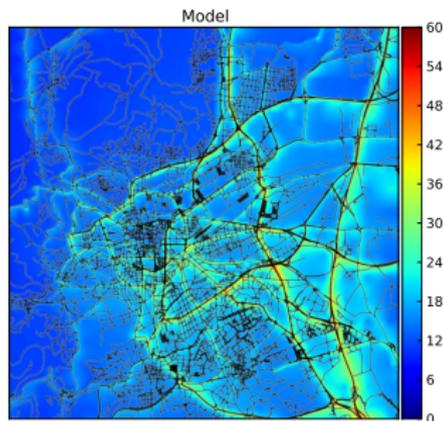
Modèle complet versus modèle réduit et émulé

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



Modèle complet versus modèle réduit et émulé

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



Modèle complet, modèle réduit et émulé, et observations

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

