

# Changements d'échelle en modélisation de la qualité de l'air, et estimation des incertitudes associées

*Soutenance de thèse – Doctorat de l'École Nationale des Ponts et Chaussées, Université Paris-Est*

Irène Bourdin-Korsakissok

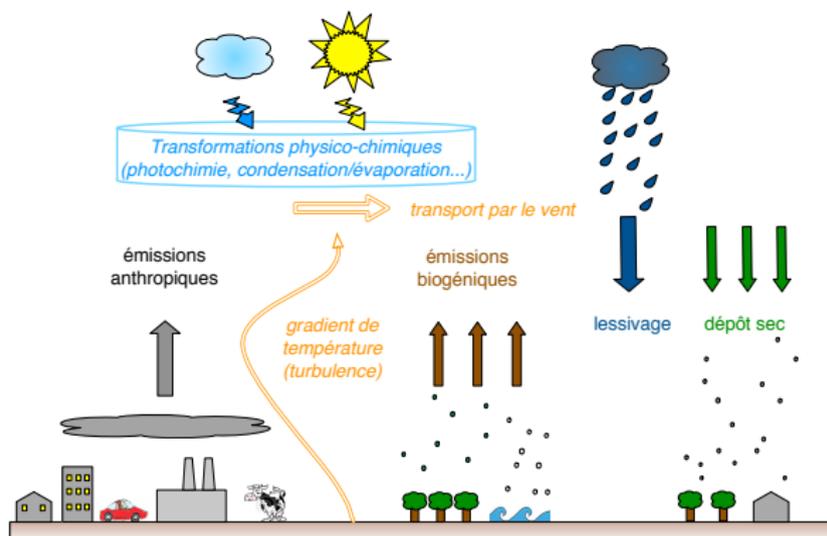


Co-direction Vivien Mallet et Bruno Sportisse



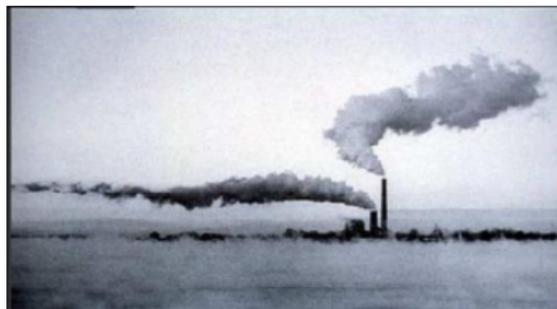
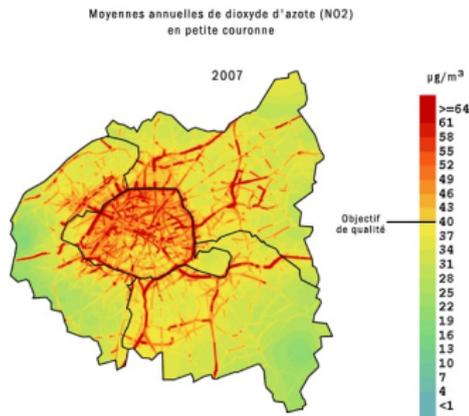
# Modélisation de la qualité de l'air

- Impact sanitaire, écosystèmes,
- Accidentel : risque industriel et radiologique...



Grande variété d'échelles spatiales et temporelles des processus physiques

# Variabilité spatiale des émissions

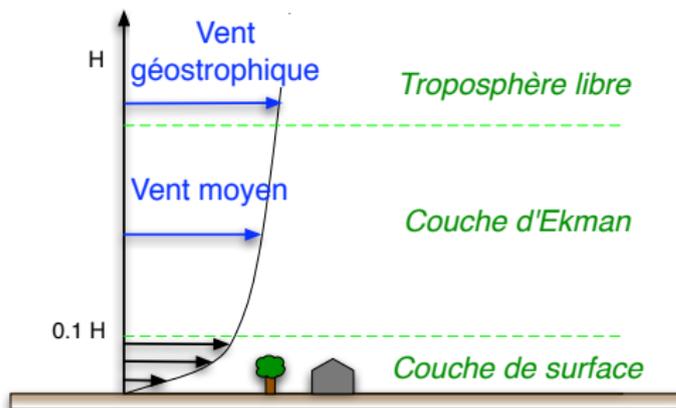


Variabilité spatiale des concentrations de NO<sub>2</sub> due à l'hétérogénéité spatiale des émissions routières (Airparif)

Sources ponctuelles et hétérogénéités du vent

# Différentes échelles météorologiques

Échelle	Taille caractéristique	Phénomène atmosphérique
Micro-échelle	< 1 km	Météorologie locale, turbulence
Méso-échelle	1 km–1000 km	Mouvements convectifs
Échelle synoptique	$10^4$ km	Mouvements synoptiques



Représentation de la couche limite atmosphérique (CLA), de hauteur  $H \sim 100 \text{ m} - 2 \text{ km}$ , et du cisaillement de vent.

La CLA est la part de l'atmosphère directement influencée par le sol, dans un temps caractéristique d'une heure.

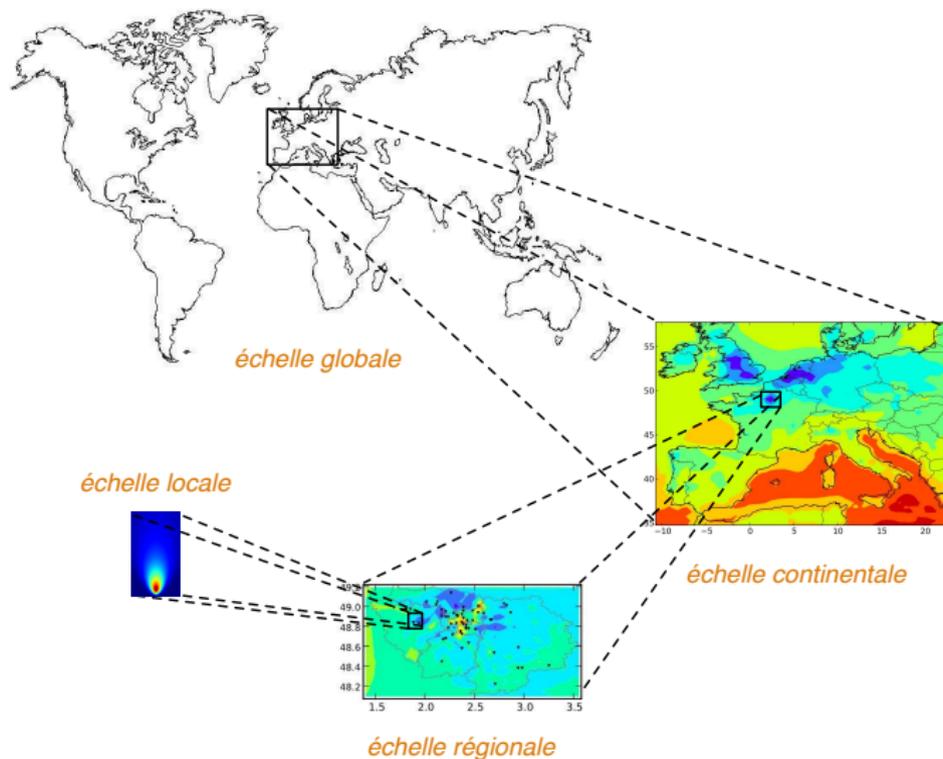
# Temps de résidence des espèces dans l'atmosphère

Déterminé par les processus de perte :

- Dépôt sec, lessivage,
- Décroissance radioactive,
- Réactions chimiques,
- Condensation/évaporation, coagulation...

Phénomène atmosphérique	Temps carac. (s)	Espèce chimique
Turbulence	$10^{-1}$ –1	OH, HO <sub>2</sub>
Mouvements convectifs	$10^2$ – $10^3$	NO, NO <sub>2</sub>
Convection, orage	$10^3$ – $10^4$	CH <sub>2</sub> O, Isoprène
Mouvements synoptiques	$10^5$ – $10^6$	H <sub>2</sub> O, SO <sub>2</sub>
Circulation générale	$10^7$	O <sub>3</sub> troposphérique, CO
Changement climatique	$10^8$ – $10^9$	CH <sub>4</sub>

# Hiérarchie d'applications, hiérarchie de modèles



# Modèles eulériens : phénomènes sous-maille

## Équation d'advection-diffusion-réaction

$$\frac{\partial c_i}{\partial t} + \underbrace{\operatorname{div}(\mathbf{u}c_i)}_{\text{Advection}} = \underbrace{\operatorname{div}\left(\rho \mathbf{K} \nabla \frac{c_i}{\rho}\right)}_{\text{Diffusion}} + \underbrace{\chi_i(\mathbf{c})}_{\text{Chimie}} + \underbrace{S_i}_{\text{Sources}} - \underbrace{\Lambda_i c_i}_{\text{Pertes}}$$

## Méthodes de changements d'échelles

- 1 Méthodes numériques
- 2 Couplage de modèles
- 3 Paramétrisations sous-maille
- 4 Méthodes stochastiques
- 5 Réduction d'échelle statistique

# Modèles eulériens : phénomènes sous-maille

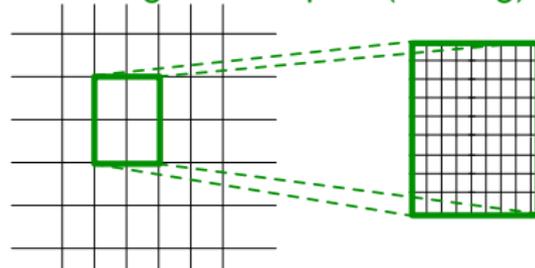
## Équation d'advection-diffusion-réaction

$$\frac{\partial c_i}{\partial t} + \underbrace{\operatorname{div}(\mathbf{u}c_i)}_{\text{Advection}} = \underbrace{\operatorname{div}\left(\rho \mathbf{K} \nabla \frac{c_i}{\rho}\right)}_{\text{Diffusion}} + \underbrace{\chi_i(\mathbf{c})}_{\text{Chimie}} + \underbrace{S_i}_{\text{Sources}} - \underbrace{\Lambda_i c_i}_{\text{Pertes}}$$

## Méthodes de changements d'échelles

- 1 Méthodes numériques
- 2 Couplage de modèles
- 3 Paramétrisations sous-maille
- 4 Méthodes stochastiques
- 5 Réduction d'échelle statistique

### Maillages imbriqués (nesting)



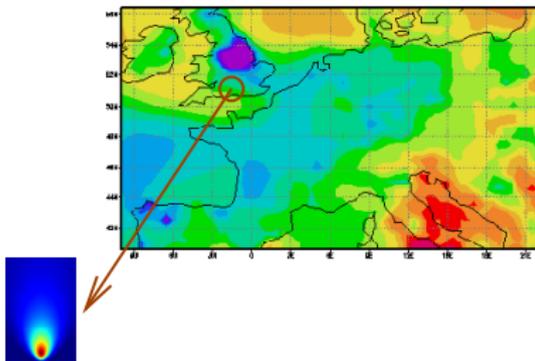
# Modèles eulériens : phénomènes sous-maille

## Équation d'advection-diffusion-réaction

$$\frac{\partial c_i}{\partial t} + \underbrace{\operatorname{div}(\mathbf{u}c_i)}_{\text{Advection}} = \underbrace{\operatorname{div}\left(\rho \mathbf{K} \nabla \frac{c_i}{\rho}\right)}_{\text{Diffusion}} + \underbrace{\chi_i(\mathbf{c})}_{\text{Chimie}} + \underbrace{S_i}_{\text{Sources}} - \underbrace{\Lambda_i c_i}_{\text{Pertes}}$$

## Méthodes de changements d'échelles

- 1 Méthodes numériques
- 2 Couplage de modèles
- 3 Paramétrisations sous-maille
- 4 Méthodes stochastiques
- 5 Réduction d'échelle statistique



# Modèles eulériens : phénomènes sous-maille

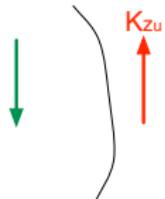
## Équation d'advection-diffusion-réaction

$$\frac{\partial c_i}{\partial t} + \underbrace{\operatorname{div}(\mathbf{u}c_i)}_{\text{Advection}} = \underbrace{\operatorname{div}\left(\rho \mathbf{K} \nabla \frac{c_i}{\rho}\right)}_{\text{Diffusion}} + \underbrace{\chi_i(\mathbf{c})}_{\text{Chimie}} + \underbrace{S_i}_{\text{Sources}} - \underbrace{\Lambda_i c_i}_{\text{Pertes}}$$

## Méthodes de changements d'échelles

- 1 Méthodes numériques
- 2 Couplage de modèles
- 3 Paramétrisations sous-maille
- 4 Méthodes stochastiques
- 5 Réduction d'échelle statistique

Mouvement ascendant (updraft)



Mouvements  
descendants  
(downdraft)  $K_{zd}$

Maille eulérienne

# Modèles eulériens : phénomènes sous-maille

## Équation d'advection-diffusion-réaction

$$\frac{\partial c_i}{\partial t} + \underbrace{\operatorname{div}(\mathbf{u}c_i)}_{\text{Advection}} = \underbrace{\operatorname{div}\left(\rho \mathbf{K} \nabla \frac{c_i}{\rho}\right)}_{\text{Diffusion}} + \underbrace{\chi_i(\mathbf{c})}_{\text{Chimie}} + \underbrace{S_i}_{\text{Sources}} - \underbrace{\Lambda_i c_i}_{\text{Pertes}}$$

## Méthodes de changements d'échelles

- 1 Méthodes numériques
- 2 Couplage de modèles
- 3 Paramétrisations sous-maille
- 4 **Méthodes stochastiques**
- 5 Réduction d'échelle statistique

$$\underbrace{\langle \Lambda(p) \rangle}_{\text{Moyenne}} = \int \underbrace{P(p)}_{\text{PDF}} \Lambda(p) dp,$$

$p$  paramètre physique (gouttes de pluie...)

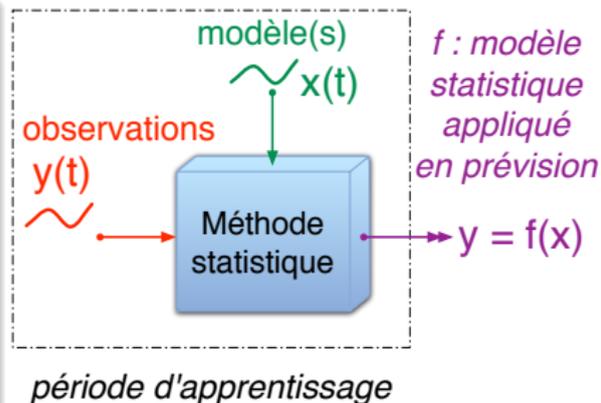
# Modèles eulériens : phénomènes sous-maille

## Équation d'advection-diffusion-réaction

$$\frac{\partial c_i}{\partial t} + \underbrace{\operatorname{div}(\mathbf{u}c_i)}_{\text{Advection}} = \underbrace{\operatorname{div}\left(\rho \mathbf{K} \nabla \frac{c_i}{\rho}\right)}_{\text{Diffusion}} + \underbrace{\chi_i(\mathbf{c})}_{\text{Chimie}} + \underbrace{S_i}_{\text{Sources}} - \underbrace{\Lambda_i c_i}_{\text{Pertes}}$$

## Méthodes de changements d'échelles

- 1 Méthodes numériques
- 2 Couplage de modèles
- 3 Paramétrisations sous-maille
- 4 Méthodes stochastiques
- 5 Réduction d'échelle statistique



# Plan

## 1 Couplage : modèle de panache sous-maille

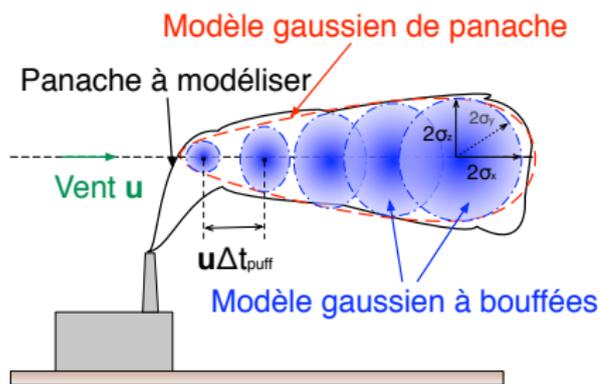
- Description : modèle à bouffées et couplage
- Panache sous-maille passif à l'échelle continentale
- Panache sous-maille réactif à l'échelle régionale

## 2 Réduction d'échelle statistique

- Principes de la réduction d'échelle pour la qualité de l'air
- Application à l'échelle européenne
- Réduction de dimension

# Description : modèle à bouffées gaussiennes

- 1 Émission ponctuelle continue, de débit  $Q_s$  : discrétisation en une série de bouffées, séparées par un pas de temps  $\Delta t_{\text{puff}}$
- 2 Distribution gaussienne dans les trois directions pour chaque bouffée,
- 3 Conditions météorologiques uniformes au sein d'une bouffée.



- Bouffée  $i$  émise au temps  $t_i = i\Delta t_{\text{puff}}$ ,
- Coordonnées  $x_c^i$ ,  $y_c^i$  et  $z_c^i$ ,
- Si vent constant  $\bar{u}$  selon  $x$  :  $x_c^i(t) = x_s + \bar{u}(t - t_i)$ ,
- Écart types  $\sigma_x^i(t - t_i)$ ,  $\sigma_y^i(t - t_i)$ ,  $\sigma_z^i(t - t_i)$ .

# Description : modèle à bouffées gaussiennes

## Écarts types gaussiens

- Paramétrisations empiriques (expériences de dispersion),
- Dépendent de la distance  $x$  (ou du temps  $t$ ), et de la météorologie,
- Trois paramétrisations : **Briggs, Doury, théorie de la similitude**

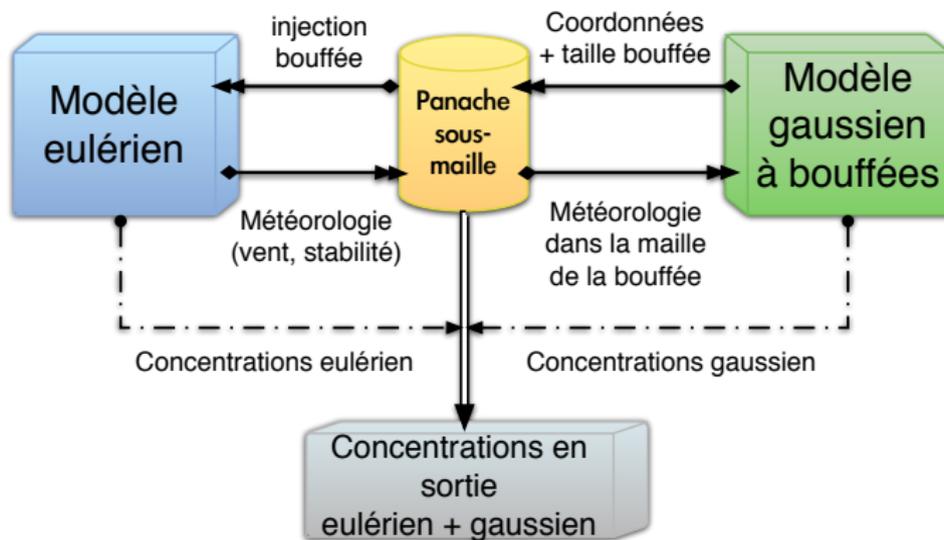
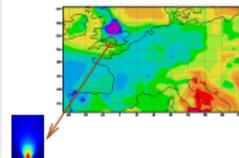
## Modèles gaussiens de Polyphemus – *première partie de la thèse*

- Modèle gaussien de panache (« plume ») et à bouffées (« puff »),
- Évaluation sur des expériences de dispersion (Prairie Grass, Kincaid), bons résultats
- **Sensibilité aux paramétrisations, incertitudes**

Référence : Korsakissok, I. et Mallet, V. (2009a). Comparative study of Gaussian dispersion formulae within the Polyphemus platform : evaluation with Prairie Grass and Kincaid experiments. J. Applied Meteor. DOI : 10.1175/2009JAMC2160.1. In Press.

## Description : couplage – principe

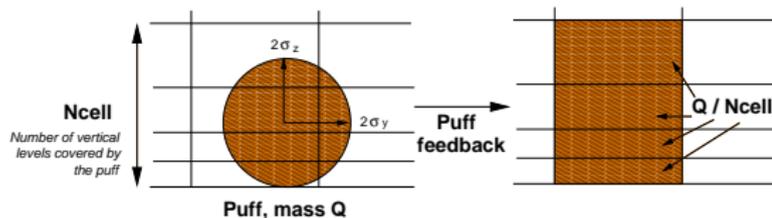
- Couplage modèle eulérien – modèle à bouffées pour les sources ponctuelles
- Bouffées « injectées » dans le modèle eulérien au bout d'un certain temps (« temps d'injection »)



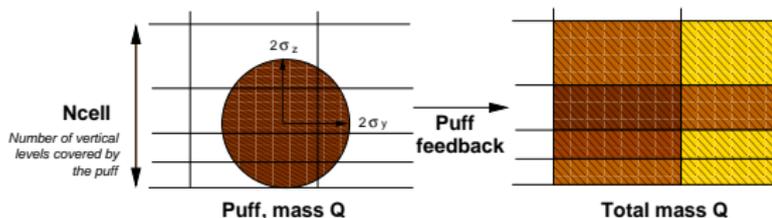
## Description : couplage – méthodes d'injection

Critères d'injection : (1) temps fixé ou (2) taille de la bouffée

- 1 Injection sur une colonne** : la masse de la bouffée est répartie dans les mailles sur la verticale, dans la colonne où se situe son centre.



- 2 Injection répartie** : La répartition de la masse est faite sur la verticale et l'horizontale, sur les mailles couvertes par la bouffée.



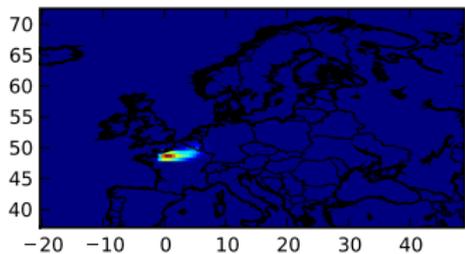
# Application à l'échelle continentale : ETEX

## Questions posées

- 1 Quel est l'impact d'une meilleure représentation des émissions sur les concentrations ? Sur la diffusion ?
  - 2 Quelle est la sensibilité aux paramétrisations locales ?
  - 3 Quelle est l'échelle (spatiale et temporelle) appropriée pour utiliser le modèle local ? Quelle est l'influence de la résolution du maillage ?
- ETEX : traceur passif, échelle européenne, 7 jours de mesure
  - Comparaison Polair3D (eulérien) avec/sans traitement sous-maille
  - Temps d'injection : une heure

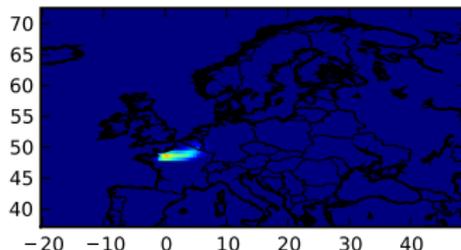
Référence : Korsakissok, I. et Mallet, V. (2009b). Subgrid-scale treatment for major point sources in an Eulerian model : a sensitivity study on the ETEX and Chernobyl cases. Journal of Geophysical Research – Atmospheres. DOI : 10.1029/2009JD012774. In Press.

# Impact spatial du traitement sous-maille (horizontal)



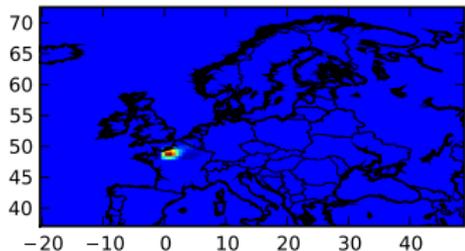
0.0 1.5 3.0 4.5 6.0 7.5 9.0 10.5 12.0 13.5

Polair3D



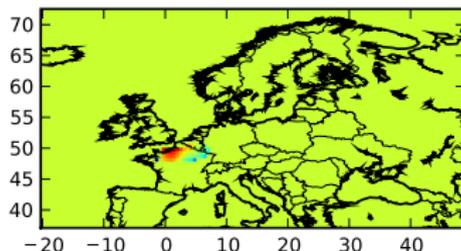
0.0 1.5 3.0 4.5 6.0 7.5 9.0 10.5 12.0 13.5

Panache sous-maille (théorie de la similitude)



0.0 0.6 1.2 1.8 2.4 3.0 3.6 4.2

Différence (Polair3D - panache sous-maille)

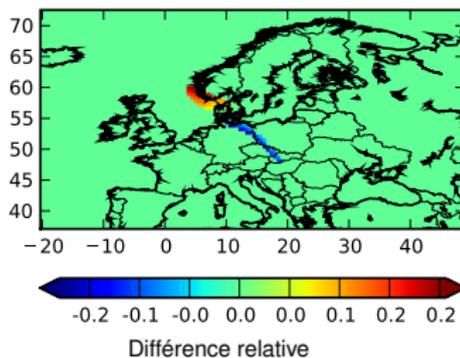
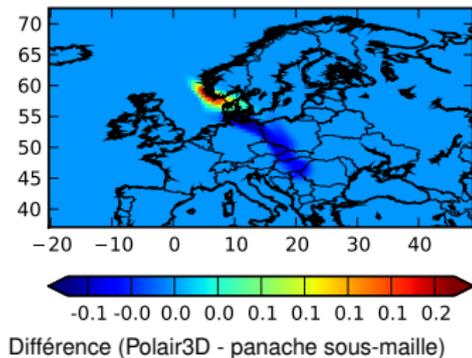
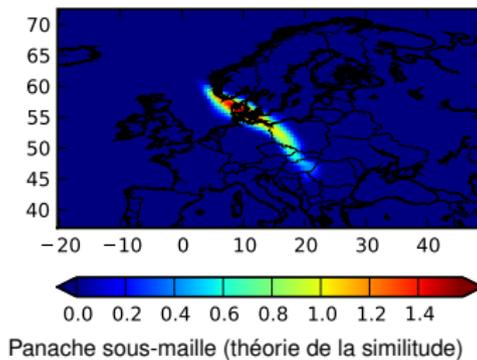
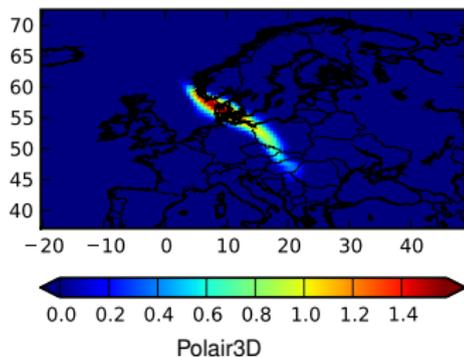


-0.8 -0.6 -0.5 -0.3 -0.2 0.0 0.2 0.3 0.4

Différence relative

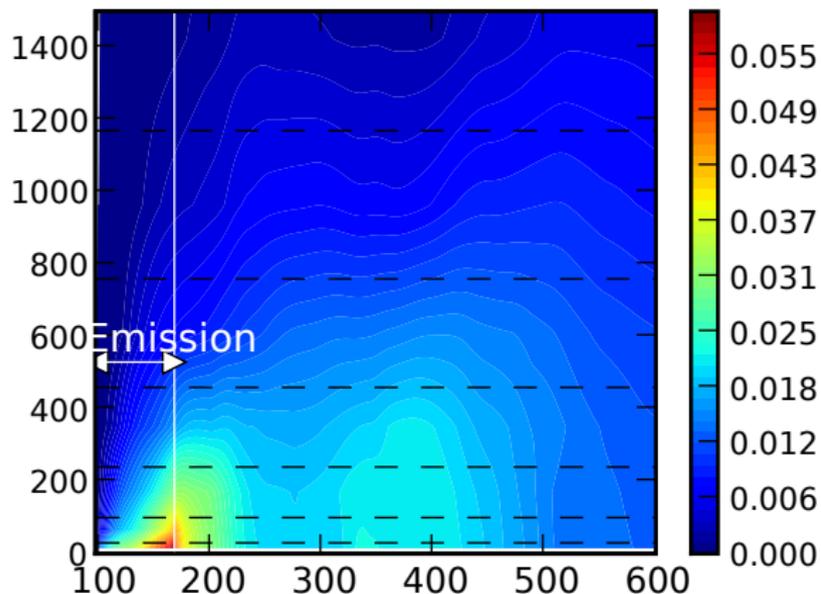
Concentrations au sol en  $\mu\text{g m}^{-3}$ , 5 h après la fin de l'émission (pas de temps 200).

# Impact spatial du traitement sous-maille (horizontal)



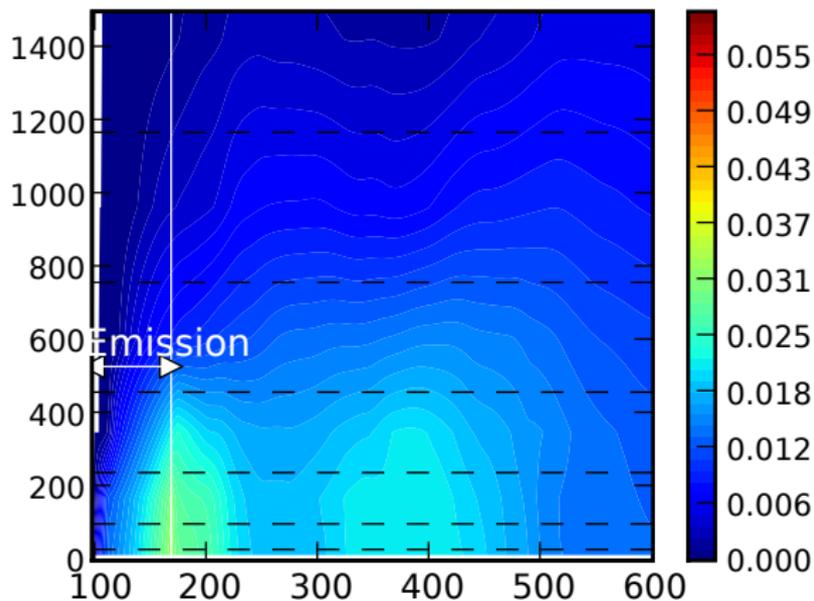
Concentrations au sol en  $\mu\text{g m}^{-3}$ , 38 h après la fin de l'émission (pas de temps 400).

# Impact spatial du traitement sous-maille (vertical)



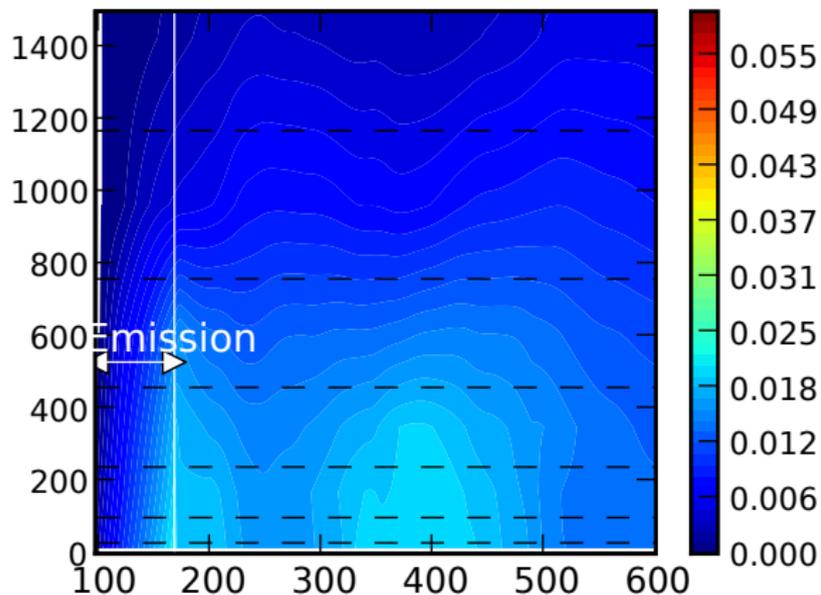
Profils verticaux des concentrations (moyenne sur tout l'espace) en fonction du temps, en  $\mu\text{g m}^{-3}$ . Référence (Polair3D).

# Impact spatial du traitement sous-maille (vertical)



Profils verticaux des concentrations (moyenne sur tout l'espace) en fonction du temps, en  $\mu\text{g m}^{-3}$ . Panache sous-maille, similitude.

# Impact spatial du traitement sous-maille (vertical)



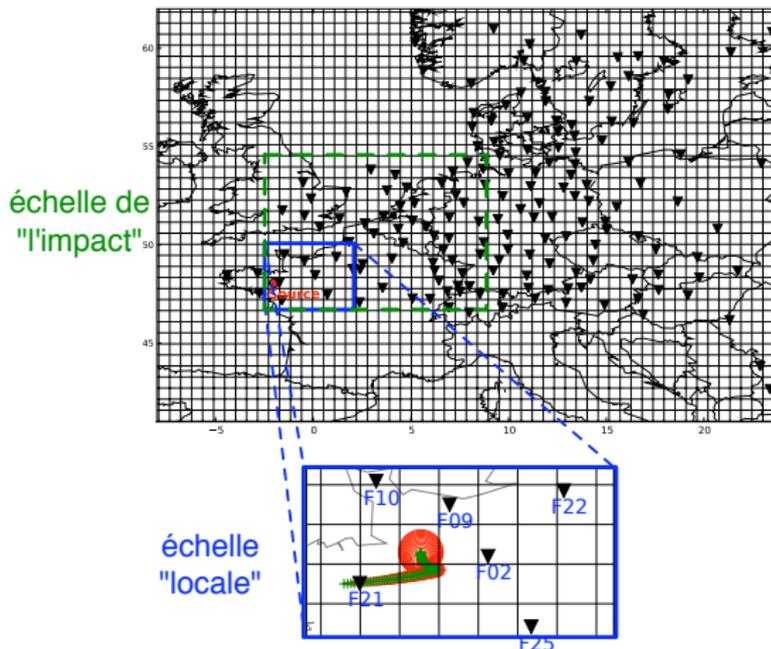
Profils verticaux des concentrations (moyenne sur tout l'espace) en fonction du temps, en  $\mu\text{g m}^{-3}$ . Panache sous-maille, Doury.

## Comparaison aux observations

- MFBE =  $\frac{2}{n} \sum_{i=1}^n \frac{x_i - y_i}{x_i + y_i}$  (biais fractionnel), avec  $x_i$  valeurs simulées, et  $y_i$  valeurs observées ( $n$  observations),
- Corrélation =  $\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \times \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$ , avec  $\bar{\alpha}$  la moyenne des  $\alpha_i$  sur les  $n$  observations ( $\alpha$  étant  $x$  ou  $y$ ),
- FAC2(FAC5) = proportion de valeurs simulées à moins d'un facteur 2(5) des observations.

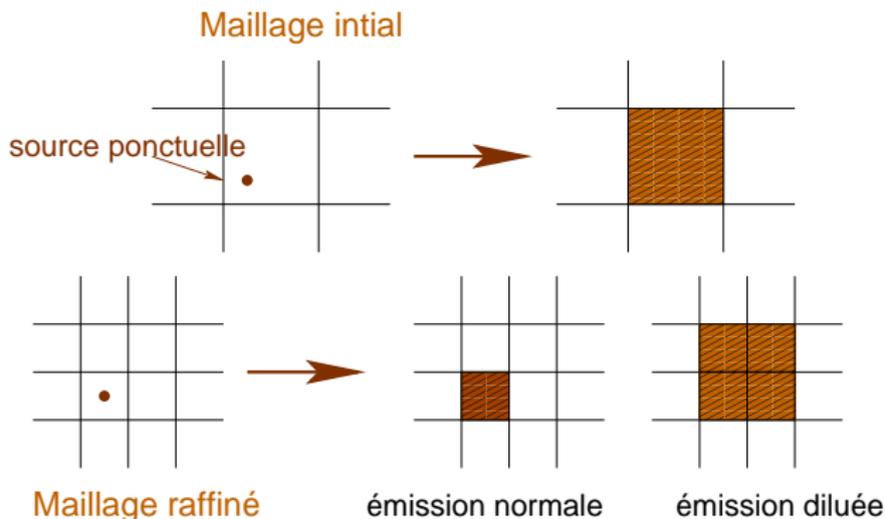
Modèle	Moyenne	MFBE	Corrélation	FAC2	FAC5
Obs	0.21	0.00	1.00	1.00	1.00
Polair3D	0.68	1.07	0.60	0.19	0.36
Similitude	0.48	0.79	0.65	0.19	0.37
Doury	<b>0.39</b>	<b>0.61</b>	<b>0.68</b>	0.20	<b>0.39</b>
Briggs	0.43	0.70	0.51	<b>0.22</b>	<b>0.39</b>

Comparaison aux observations sur 168 stations. Moyenne en  $\mu\text{g m}^{-3}$ . Panache sous-maille avec l'injection sur une colonne et un temps d'injection d'une heure.



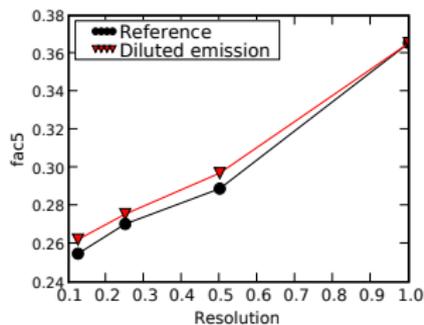
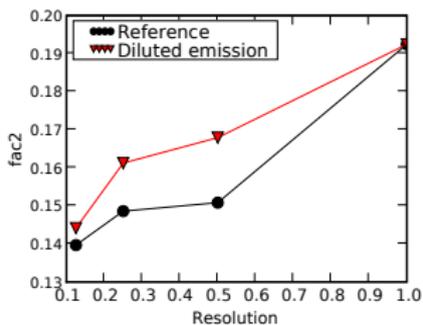
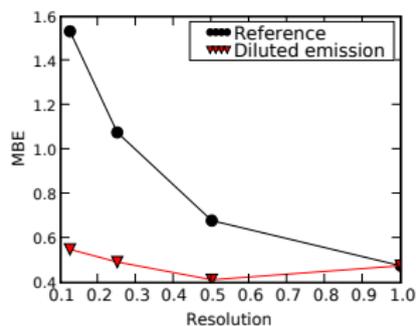
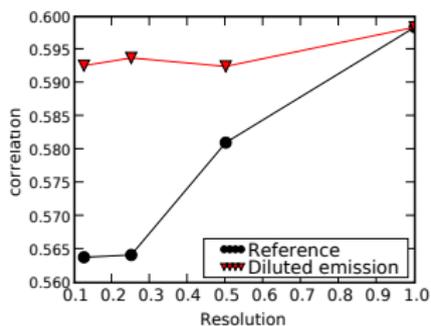
- Modèle « local », impact « global »
- Effet du panache sous-maille : diffusion verticale en champ proche
- Meilleurs résultats si plus de diffusion (Doury, colonne)

# Influence de la résolution du maillage sur l'émission



- Résolutions plus fines :  $\div 8$ ,  $\div 4$ ,  $\div 2$
- Emission dans une maille, ou répartie sur plusieurs mailles

# Influence de la résolution du maillage sur l'émission

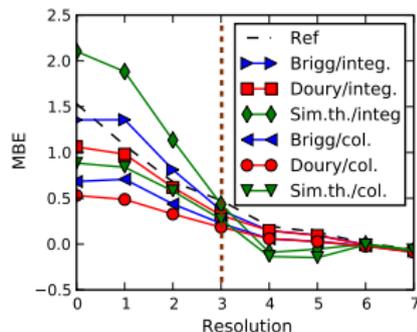
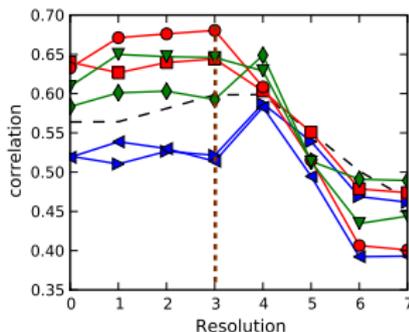


- Résolutions plus fines :  $\div 8$ ,  $\div 4$ ,  $\div 2$
- Performances moins bonnes à fine résolution

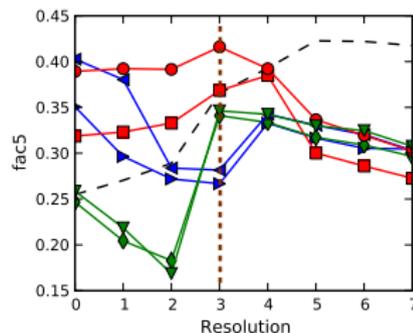
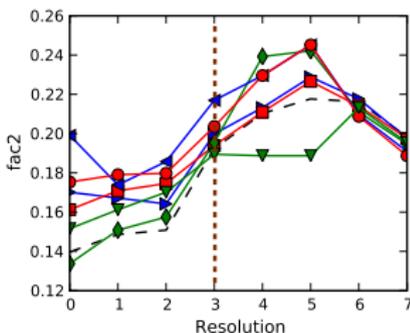
# Influence de la résolution sur le panache sous-maille

## Temps d'injection fixé (une heure)

Corrélation, biais (MBE), FAC2 et FAC5 pour plusieurs résolutions de maillage, pour Polair3D (---) et le panache sous-maille (différente configurations).  
Résolution initiale  $0.5^\circ$ , abscisse 3 (.....).



Impact du panache sous-maille plus grand pour un maillage fin

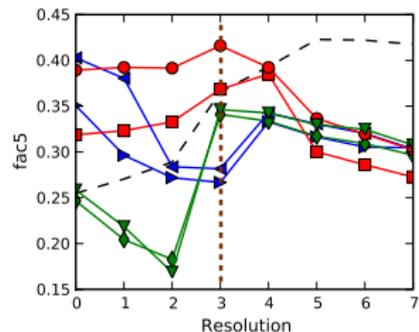
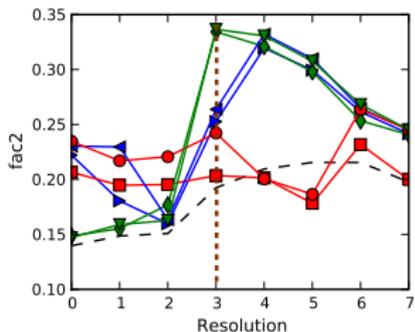
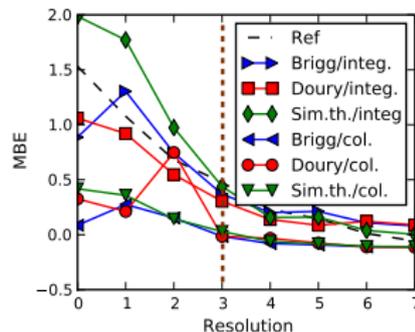
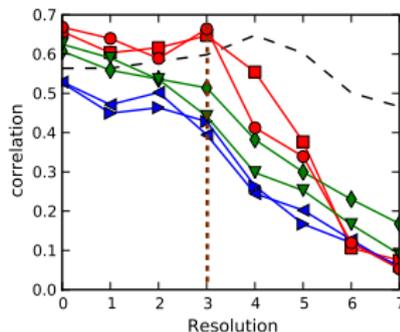


# Influence de la résolution sur le panache sous-maille

Corrélation, biais (MBE), FAC2 et FAC5 pour plusieurs résolutions de maillage, pour Polair3D (---) et le panache sous-maille (différente configurations).  
Résolution initiale  $0.5^\circ$ ,  
abscisse 3 (-----).

Impact du panache sous-maille plus grand pour un maillage fin

## Critère sur la taille des bouffées



# Échelle régionale : panache sous-maille réactif

## Questions posées

- 1 Quel est l'impact d'une meilleure représentation des émissions sur un panache réactif ?
- 2 Quel est l'impact sur les espèces primaires, secondaires ?
- 3 Impact sur les résultats moyens pour six mois ?
- 4 Variabilité spatiale et/ou impact aux stations ?

Référence : Korsakissok, I. et Mallet, V. (2009c). Development and application of a reactive plume-in-grid model : Evaluation over Greater Paris. Atmospheric Chemistry and Physics. Submitted.

# Chimie dans les bouffées : bouffées superposées

- Les espèces d'une bouffée  $\alpha$  réagissent entre elles
- Les espèces de bouffées  $\alpha$  et  $\beta$  superposées réagissent entre elles
- Les espèces d'une bouffée réagissent avec les espèces de fond

- Volume de la bouffée  $\alpha$

$$V_{\alpha} = \frac{\langle c_A^{\alpha} \rangle^2}{\langle c_A^{\alpha 2} \rangle}$$

$c_A^{\alpha}$  concentration moyenne de l'espèce  $A$  dans la bouffée  $\alpha$

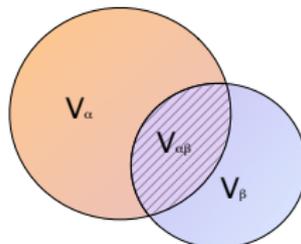
$Q_A^{\alpha} = \langle c_A^{\alpha} \rangle$  quantité totale de  $A$  dans  $\alpha$  avec  $\langle \cdot \rangle$  intégrale sur tout l'espace

- Volume de superposition entre deux bouffées  $\alpha$  et  $\beta$

$$\frac{V_{\alpha\beta}}{V_{\alpha} V_{\beta}} = \frac{\langle c_A^{\alpha} c_A^{\beta} \rangle}{\langle c_A^{\alpha} \rangle \langle c_A^{\beta} \rangle}$$

- Concentration de superposition de la bouffée  $\alpha$ , sur toutes les bouffées  $\beta$  avec lesquelles  $V_{\alpha\beta} \neq 0$

$$\widehat{c}_A^{\alpha} = \sum_{\beta} Q_A^{\beta} \frac{V_{\alpha\beta}}{V_{\alpha} V_{\beta}} = \sum_{\beta} c_A^{\beta} \frac{V_{\alpha\beta}}{V_{\alpha}}$$



# Chimie dans les bouffées : interaction avec le fond



$c_A^\alpha, c_B^\alpha$  bouffée

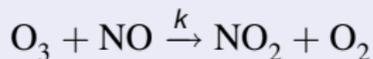
$c_A^b, c_B^b$  fond

$$\frac{d(c_A^\alpha + c_A^b)}{dt} = -k \underbrace{(c_A^\alpha c_B^\alpha)}_{\text{bouffée}} + \underbrace{c_A^b c_B^b}_{\text{fond}} + \underbrace{c_A^\alpha c_B^b + c_B^\alpha c_A^b}_{\text{interactions}}$$

$$\frac{dc_A^b}{dt} = -kc_A^b c_B^b \quad \text{chimie dans le fond (eulérien)}$$

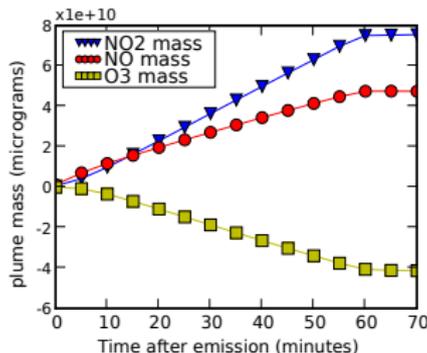
$$\frac{dc_A^\alpha}{dt} = \frac{d(c_A^\alpha + c_A^b)}{dt} - \frac{dc_A^b}{dt} \quad \text{bouffée = perturbation du fond}$$

## Titration de l'ozone



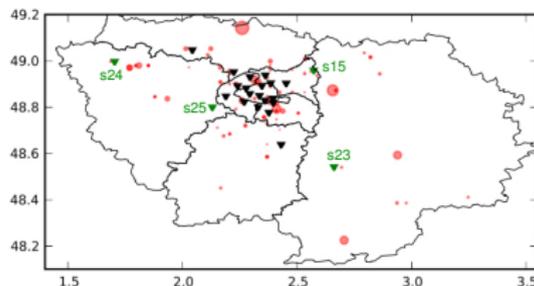
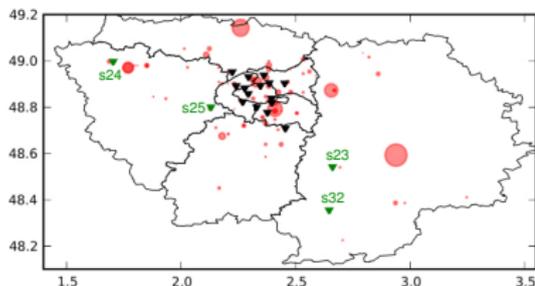
- panache de  $NO_x$  ( $NO+NO_2$ )
- Fond (homogène) d'ozone  $O_3$

Dans le panache,  $O_3$  diminue



Évolution de la masse totale de différentes espèces ( $\mu\text{g}$ ) dans un panache continu de  $NO_x$  émis dans un fond d' $O_3$ .

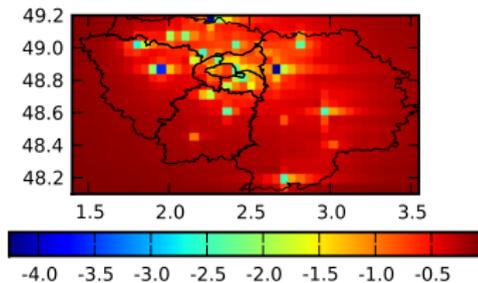
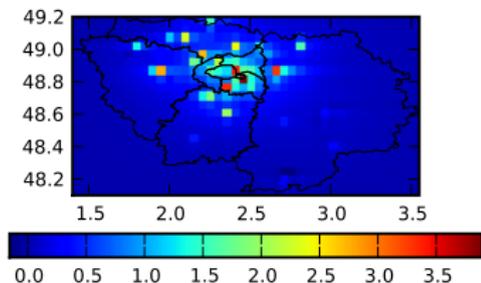
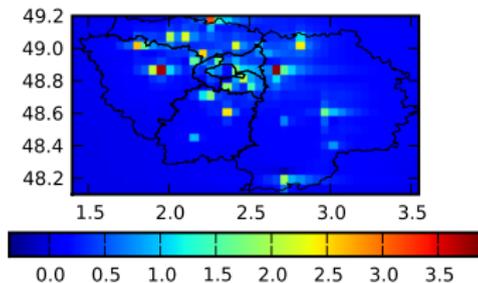
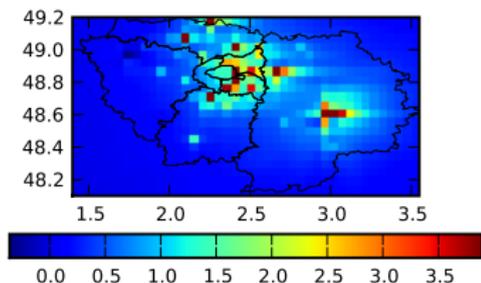
# Échelle régionale : photochimie en Île-de-France



Sources sélectionnées (●), et stations de mesure rurales (▼) et urbaines (▼). Gauche : SO<sub>2</sub>, droite : NO. Le diamètre des cercles est proportionnel au taux d'émission des sources.

- Île-de-France, été 2001 (1<sup>er</sup> avril – 27 septembre)
- Modèle de panache sous-maille avec la chimie des bouffées (RACM, chimie gazeuse complexe), comparaison avec Polair3D
- Injection des bouffées à 20 minutes, similitude, injection colonne
- 89 sources ponctuelles :  $Q_s > 10^6 \mu\text{g s}^{-1}$  pour NO<sub>x</sub> ou SO<sub>2</sub>

# Impact spatial du traitement sous-maille



Différences de concentrations au sol : Polair3D - panache sous-maille.  
Moyenne sur six mois ( $\mu\text{g m}^{-3}$ ).

## Comparaison aux observations

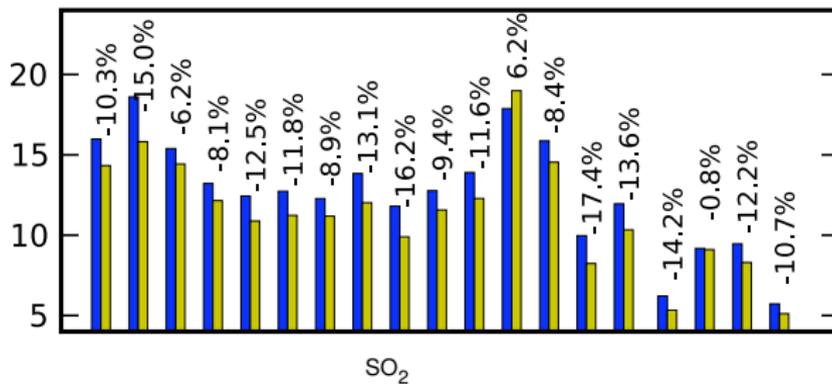
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}, \quad \text{avec } x_i \text{ valeurs simulées, et } y_i \text{ valeurs observées.}$$

Espèce	Moyenne	MFBE	Corrélation	RMSE
SO <sub>2</sub>	13.76/11.98	<b>0.66/0.55</b>	0.35/0.31	<b>13.08/11.88</b>
NO	20.93/19.64	<b>0.41/0.36</b>	0.47/0.46	<b>33.08/31.62</b>
NO <sub>2</sub>	35.84/35.23	<b>0.06/0.04</b>	0.58/0.57	<b>35.23/35.84</b>
O <sub>3</sub>	40.24/41.05	<b>-0.46/-0.44</b>	0.68/0.68	<b>30.58/30.18</b>

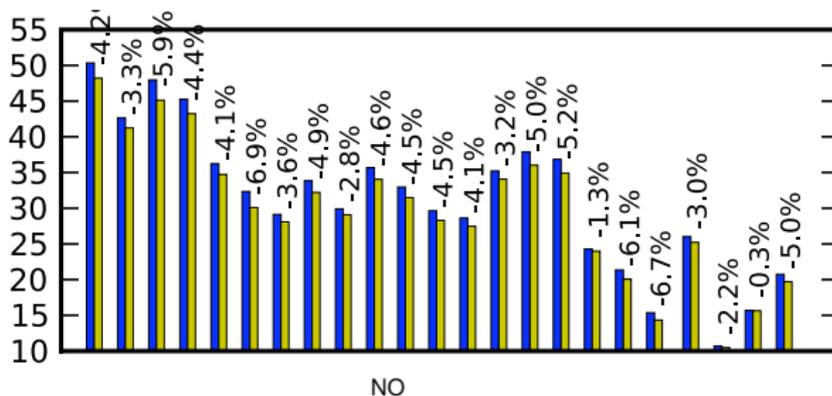
Comparaison aux observations sur six mois. Moyenne et RMSE en ( $\mu\text{g m}^{-3}$ ). Polair3D/panache sous-maille.

- Impact relatif : part des sources ponctuelles dans les émissions
- Espèces primaires vs espèces secondaires

# RMSE aux stations



■ Polair3D,  
■ panache  
sous-maille



# Conclusion : modèle de panache sous-maille

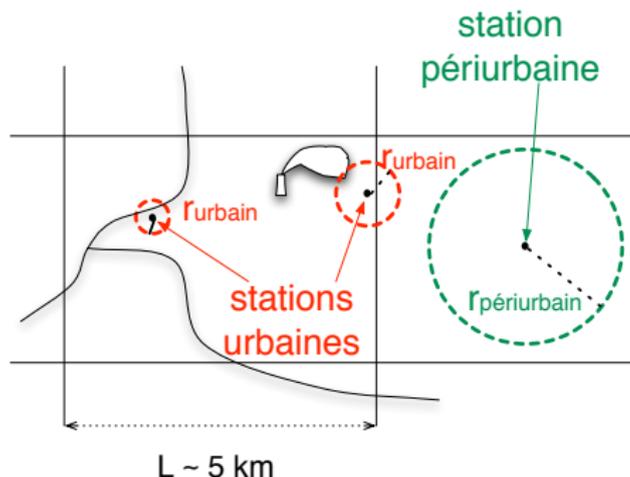
## Conclusions

- Mise en évidence de l'importance la diffusion verticale
- Variabilité due aux paramétrisations locales
- **Importance de la météorologie (trajectoire et cisaillement)**
- Échelles spatiales et temporelles de l'impact pour les espèces réactives/passives

## Perspectives

- Scission des bouffées (météorologie)
- Résolution des champs météorologiques : couplage avec un modèle météorologique local
- **Extension du modèle de panache sous-maille aux sources linéiques : émissions routières**

# Échelle de représentativité des stations



Rayon de représentativité  $r$  des stations : émissions, topographie...

$$r_{\text{urbain}} < r_{\text{périurbain}} < r_{\text{rural}} < r_{\text{fond}}$$

# Erreur de représentativité

## Erreur de représentativité du modèle

- Échelle de représentativité  $r$  d'une station
- Échelle « bien représentée » par le modèle eulérien :  $L \sim$  quelques mailles.

$r \ll L \rightarrow$  erreur de représentativité

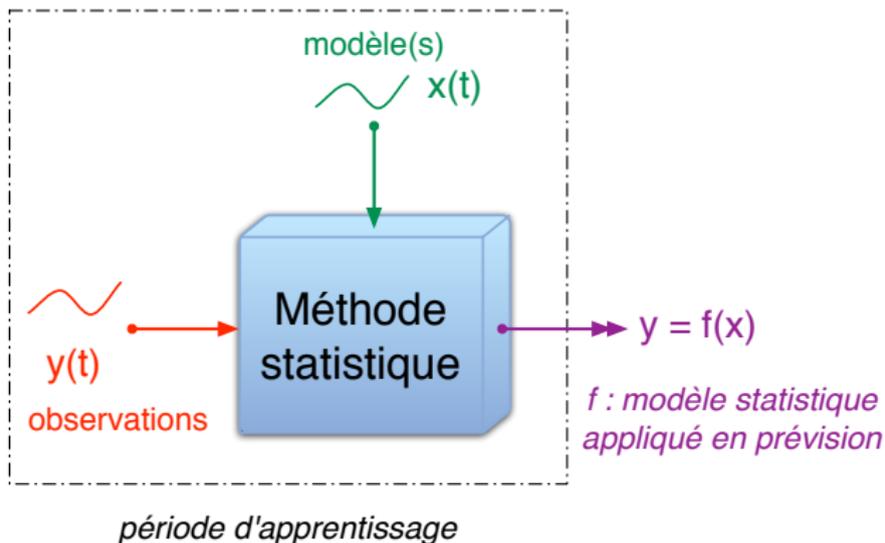
- Erreur totale à la station  $k$  : écart entre le modèle  $x_k$  et les mesures  $y_k$  (au temps  $t$ )

$$\epsilon_k(t) = x_k(t) - y_k(t) = \epsilon_{k_{\text{représentativité}}}(t) + \epsilon_{k_{\text{modèle}}}(t) + \epsilon_{k_{\text{mesure}}}(t)$$

- Même un modèle « parfait » ( $\epsilon_{k_{\text{modèle}}} = 0$ ) ferait une erreur aux stations...
- Réduction d'échelle : réduire l'erreur de représentativité

# Réduction d'échelle statistique

Principe : déterminer une relation statistique entre la variable à **grande échelle** (sortie du modèle) et la variable à **petite échelle** (observations)



Période d'apprentissage suffisamment longue (variations météorologiques) : relation « systématique » entre  $x$  et  $y$

# Réduction d'échelle statistique

Principe : déterminer une relation statistique entre la variable à **grande échelle** (sortie du modèle) et la variable à **petite échelle** (observations)

## Régression linéaire à une station

$$\mathbf{y} = \mathbf{X}^T \mathbf{b} + \epsilon,$$

- $\mathbf{y} = (y_1, \dots, y_n)^T$  observations à la station aux temps  $(t_1, \dots, t_n)$
- $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m)^T$  matrice de taille  $m \times n$ , sorties de  $m$  modèles, avec  $\mathbf{x}_s$  sorties du modèle  $s$  (concentrations interpolées)
- $\mathbf{y}$ ,  $\mathbf{x}_s$  centrés :  $\bar{\mathbf{y}} = 0$ ,  $\bar{\mathbf{x}}_s = 0$
- $\mathbf{b}$  coefficients de régression pour les  $m$  modèles (taille  $m$ )
- $\epsilon$  résidu de la régression (taille  $n$ ) à **minimiser**.

Estimation des moindres carrés :  $\hat{\mathbf{b}} = (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X} \mathbf{y}$  et  $\hat{\mathbf{y}} = \mathbf{X}^T \hat{\mathbf{b}}$

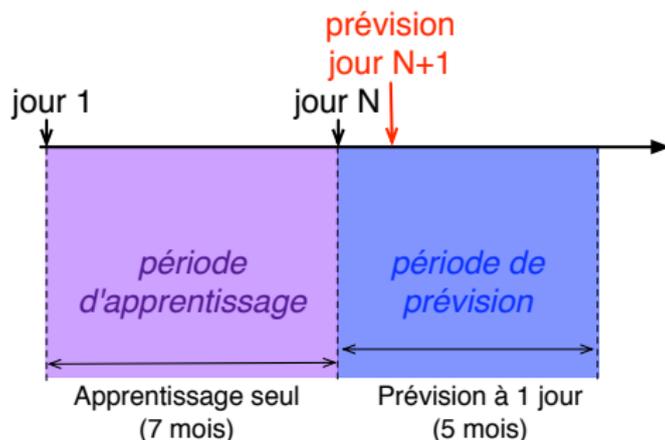
# Application à l'échelle continentale

## Questions posées

- 1 Quel est l'impact de la réduction d'échelle sur les performances ?
- 2 Quelle est l'influence de la réduction d'échelle en fonction du type de station (rural/urbain), donc de son échelle de représentativité ?
- 3 Quel est l'apport d'un ensemble de modèles ?

# Application à l'échelle continentale

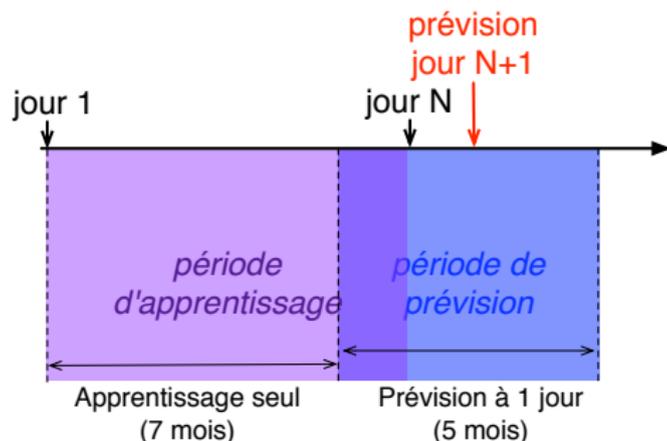
- Prédiction des pics d'ozone à l'échelle européenne, année 2001
- Réseaux de mesure : Pioneer (154 stations) et Airbase (1247 stations)
- Réduction d'échelle
  - 1 pour un seul modèle (« modèle de référence »),
  - 2 pour un ensemble de modèles



Prédiction à un jour : apprentissage du début de l'année au jour  $N$ , prédiction pour le jour  $N + 1$  (cinq mois)

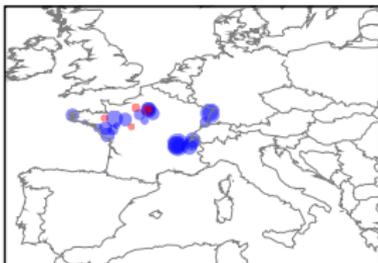
# Application à l'échelle continentale

- Prédiction des pics d'ozone à l'échelle européenne, année 2001
- Réseaux de mesure : Pioneer (154 stations) et Airbase (1247 stations)
- Réduction d'échelle
  - 1 pour un seul modèle (« modèle de référence »),
  - 2 pour un ensemble de modèles

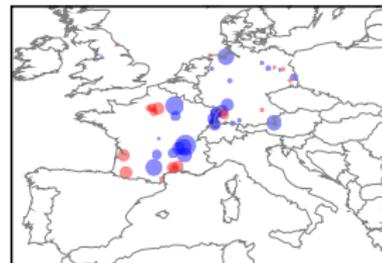


Prévision à un jour :  
 apprentissage du  
 début de l'année au  
 jour  $N$ , prévision pour  
 le jour  $N + 1$  (cinq  
 mois)

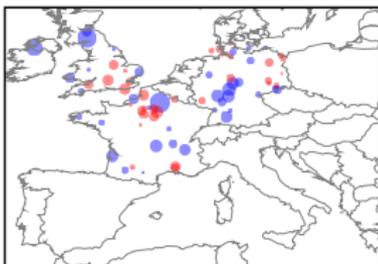
# Réduction d'échelle sur un seul modèle



Stations urbaines. (min : -9.6 ; max : 1.8)



Stations périurbaines. (min : -8.4 ; max : 2.8)



Stations rurales. (min : -7.8 ; max : 2.4)

Stations	RMSE modèle	RMSE prévision
Toutes (154)	22.4	21.1 (-6%)
Urbaines (35)	25.4	22.6 (-11%)
Péri-urbaines (47)	24.0	22.9 (-4%)
Rurales (72)	19.5	19.1 (-2%)

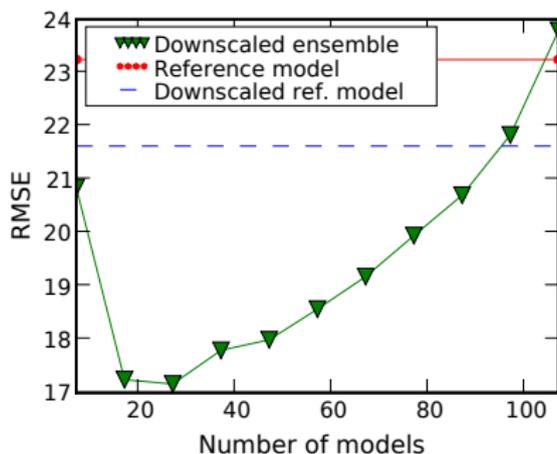
Différences de RMSE aux stations, avec le modèle de référence, et après réduction d'échelle (prévision à 1 jour). RMSE sur cinq mois ( $\mu\text{g m}^{-3}$ ).

Différence négative (●) : la réduction d'échelle diminue la RMSE.

Différence positive (●) : la réduction d'échelle augmente la RMSE.

## Prise en compte de l'ensemble

- Ensemble de 107 modèles : paramétrisations, données d'entrée
- Modèle de référence : « bonne » configuration (Polair3D)
- Choix aléatoire de  $N_s$  modèles parmi les 107



RMSE (cinq mois, 93 stations) sur le réseau Pioneer : modèle de référence avec et sans réduction d'échelle, et RMSE pour l'ensemble en fonction du nombre de modèles ( $\mu\text{g m}^{-3}$ ). Prévission à un jour.

## Prise en compte de l'ensemble : colinéarité

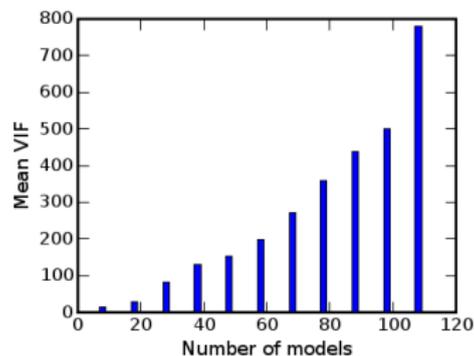
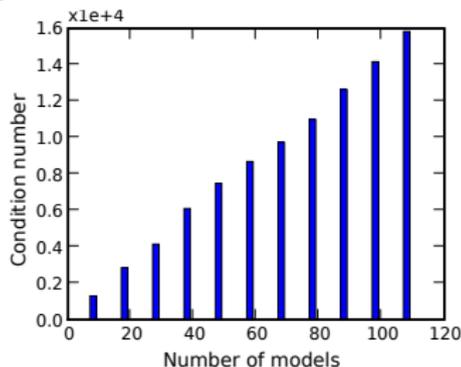
Estimation des coefficients de régression :  $\hat{\mathbf{b}} = (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X} \mathbf{y}$

Variance des coefficients de régression :  $\text{var}(\hat{\mathbf{b}}) = \sigma_\epsilon^2 (\mathbf{X}\mathbf{X}^T)^{-1}$

Si  $\mathbf{X}\mathbf{X}^T$  mal conditionnée,  $\text{var}(\hat{\mathbf{b}})$  très grand : estimation non fiable...

**Indicateurs de colinéarité** : nombre de conditionnement  $\kappa$  et facteur

d'inflation de la variance  $\text{vif}(\mathbf{s}) = \frac{1}{1-R_s^2}$  avec  $R_s$  corrélation multiple entre le modèle  $s$  et les autres modèles.



Indicateurs de colinéarité en fonction du nombre de modèles de l'ensemble :  
moyenne sur les stations, et moyenne sur les modèles pour le vif

# Réduction de dimension

- 1 Choix de  $N_s$  modèles au hasard, ou calibration d'un ensemble
- 2 Décomposition en composantes principales « PCA »
- 3 Décomposition en composantes principales ajustées « PFC »

## Décomposition en composantes principales

- Principe : éliminer les redondances, garder les « modes » dominants
- Vecteurs propres  $\mathbf{p}_k$  de la matrice  $C_{XX} = XX^T$  : **directions principales**
- Projection de  $\mathbf{x}(t)$  sur la base des  $\mathbf{p}_k$

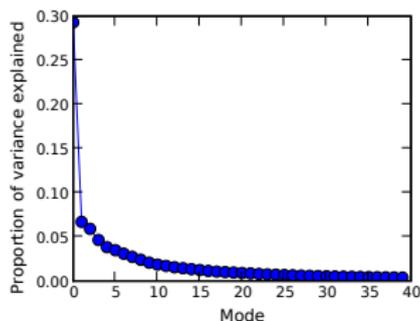
$$\mathbf{x}(t) = \sum_{k=1}^m \alpha_k(t) \mathbf{p}_k.$$

- Les coefficients  $\alpha_k(t)$  sont les composantes principales (orthogonales)
- Les valeurs propres  $\lambda_k$  donnent l'importance des modes (part expliquée de variance de  $X$ ) → **tronquer aux  $N_m$  premiers modes**

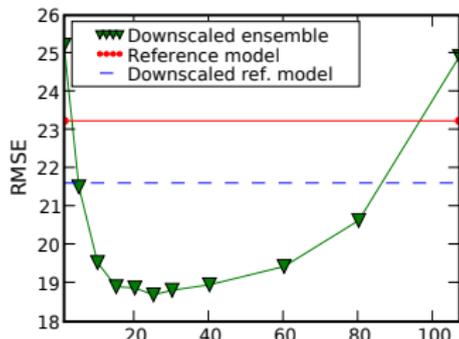
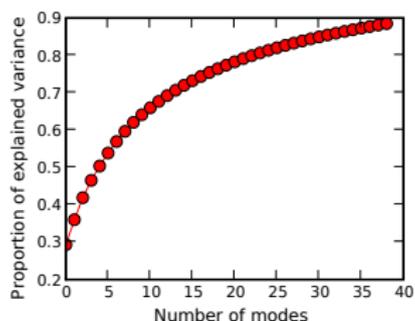
# Décomposition en composantes principales

Variance expliquée par un mode  $k$  :  $\lambda_k / \sum_j \lambda_j$  (puis moyenne sur les stations)

Variance expliquée par chacun des modes



Variance expliquée cumulée



Réduction d'échelle sur les  $N_m$  premières composantes principales : RMSE (cinq mois, Pioneer, 93 stations) en fonction du nombre de modes inclus ( $\mu\text{g m}^{-3}$ ). Prédiction à un jour.

# Décomposition en composantes principales ajustées

Référence : Cook, R. D. (2007). Fisher Lecture : Dimension Reduction in Regression. *Stat. Sci.*, 22(1) :1-26.

- Principe : composantes principales dépendant de  $X$  et de  $\mathbf{y}$
- PCA sur une matrice  $\hat{X}$  « ajustée » : projection de  $X$  sur un espace dépendant de  $\mathbf{y}$  :  $\hat{X}^T = P_F X^T$  avec  $P_F = F_y^T (F_y F_y^T)^{-1} F_y$ ,
- $F_y$  matrice contenant  $N_r$  fonctions de  $\mathbf{y}$  (centré), de taille  $(N_r \times N_t)$

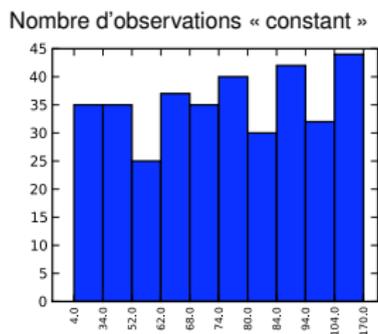
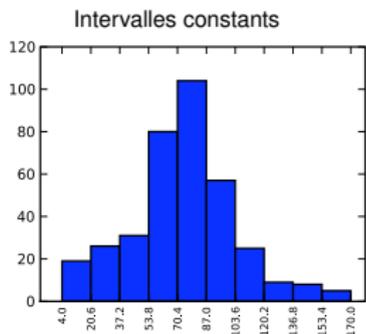
- 1  $F_y = \mathbf{y}^T$  : on a alors  $N_r = 1$  (une seule composante)
- 2 Fonctions polynomiales :  $F_y = (\mathbf{y}^T, \mathbf{y}^{2T}, \dots)^T$  relation entre  $X$  et  $\mathbf{y}$
- 3  $F_y$  contenant  $N_r$  fonctions continues par morceaux, définies par

$$f_r(y_k) = \begin{cases} 1 - n_r/N_t & \text{si } y_k \in I_r, \\ -n_r/N_t & \text{sinon,} \end{cases}$$

avec  $n_r$  le nombre d'observations dans l'intervalle  $I_r$   
 $f_r(y_k) > 0$  si l'observation au temps  $t_k$  est dans  $I_r$ ,  $< 0$  sinon.

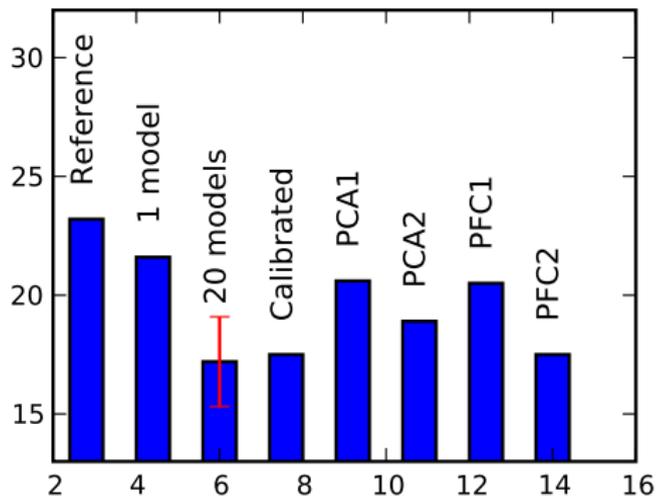
# PFC avec les fonctions continues par morceaux

- 1 « PFC1 » : intervalles constants choisis sur l'ensemble des valeurs de  $\mathbf{y}$ .
- 2 « PFC2 » : nombre d'observations « constant »  $n_r = \frac{N_t}{N_r}$  ( $r \in \llbracket 0, N_r \rrbracket$ )
- 3  $N_r = 10$  : si  $N_r$  grand, peu d'observations dans les intervalles ( $n_r \sim 0$ ) :  $f_r$  prend des valeurs  $\sim 0$  et  $F_y$  est mal conditionnée.



Nombre d'observations inclus dans chacun des intervalles. Abscisse : bornes de chaque intervalle (valeur de l'observation en  $\mu\text{g m}^{-3}$ ).

# Bilan : évaluation des différentes méthodes



Comparaison des performances en prévision (à un jour) aux stations du réseau Pioneer (93 stations).

- « **Reference** » : modèle de référence (sans réduction d'échelle),
- « **1 model** » : réduction d'échelle sur le modèle de référence
- « **20 models** » : réduction d'échelle sur un ensemble (20 modèles)
- « **Calibrated** » : réduction d'échelle sur un ensemble calibré sur les stations rurales (23 modèles)
- « **PCA1** » : décomposition en composantes principales (40 modes expliquant 90% de la variance)
- « **PCA2** » : décomposition en composantes principales (20 modes)
- « **PFC1** » : décomposition en composantes principales ajustées à intervalles constants ( $N_r = 10$ )
- « **PFC2** » : décomposition en composantes principales ajustées à nombre d'observations constant ( $N_r = 10$ )

# Conclusion : réduction d'échelle

## Conclusions

- Amélioration en fonction de la représentativité des stations
- **Apport de l'ensemble** très notable
- Réduction de dimension : intérêt des **composantes principales ajustées** (meilleurs résultats, plus faible nombre de régresseurs)
- Problème : déterminer les « bonnes » fonctions d'ajustement

## Perspectives

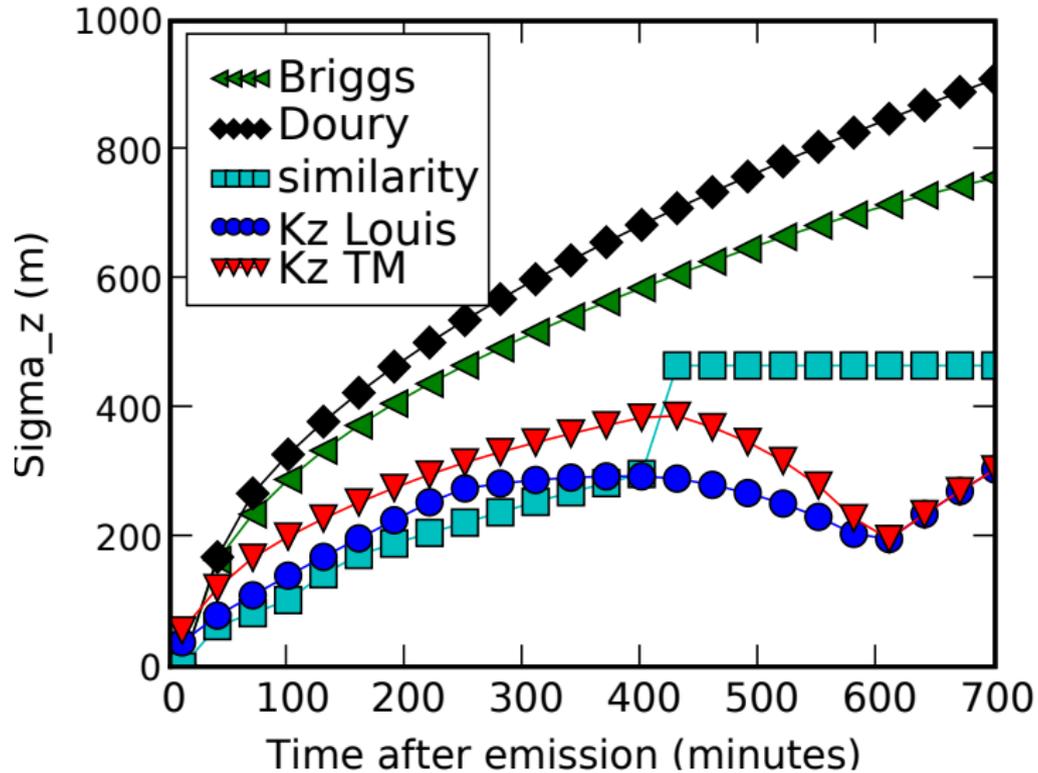
- Différentes échelles, différents polluants (plus « locaux », e.g.  $\text{NO}_x$ )
- **Prise en compte de l'erreur modèle via l'ensemble calibré**
- Calcul plus fin du rayon de représentativité (covariogrammes)
- **Estimation spatiales des concentrations (krigeage)**

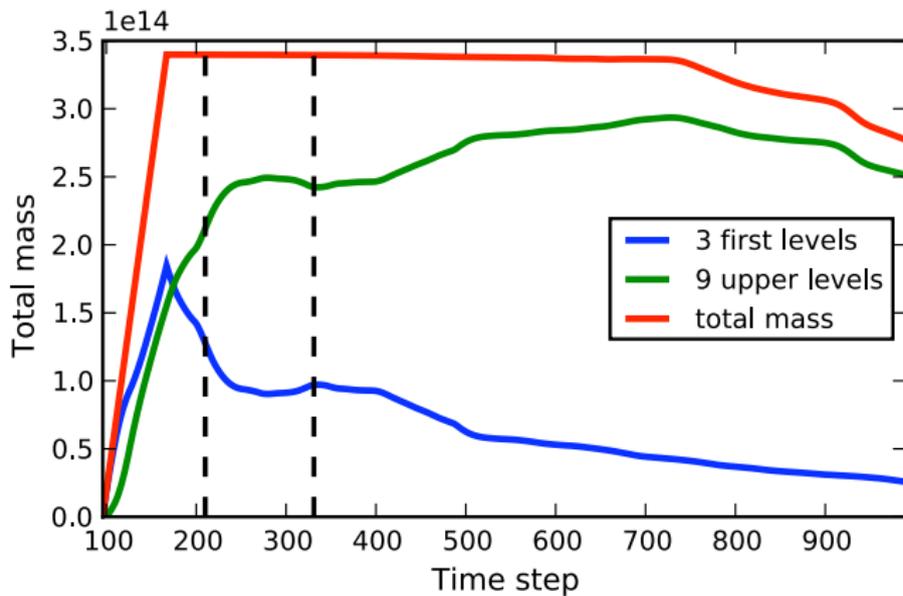
# Conclusions

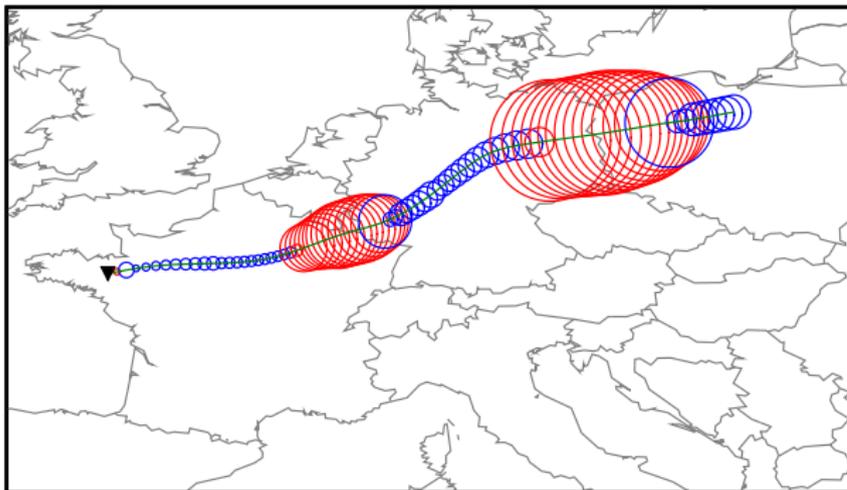
## Bilan et perspectives

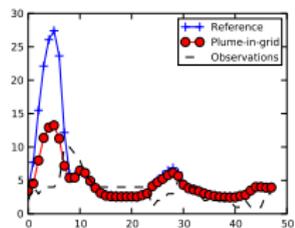
- 1 Deux méthodes développées et évaluées, opérationnelles
- 2 Meilleure prise en compte de la variabilité locale : impact aux stations, impact spatial
- 3 Complémentarité : modélisation et traitement statistique
- 4 Importance de modéliser une échelle plus fine pour *plusieurs* processus (émissions et météorologie...)
- 5 Autres perspectives : maillage adaptatif, méthodes stochastiques

Merci de votre attention

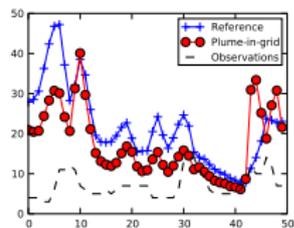




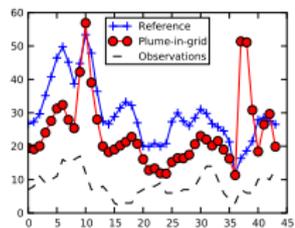




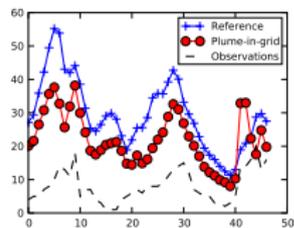
(a) MELUN



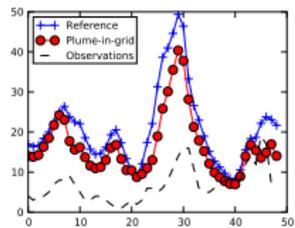
(b) VITRY-SUR-SEINE



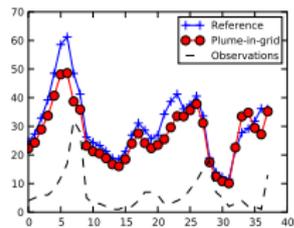
(c) IVRY-SUR-SEINE



(d) PARIS12eme



(e) AUBERVILLIERS



(f) LADEFENSE