

Principes de la modélisation inverse pour les émissions de polluants

Marc Bocquet

CEREA, joint lab École des Ponts ParisTech and EdF R&D,
Université Paris-Est, Champs-sur-Marne, France

(marc.bocquet@enpc.fr)

Collaborateurs

M. R. Koohkan, J.-M. Haussaire, Y. Liu, Y. Roustan, C. Seigneur, A. Mathieu, O. Saunier.



Outline

1 Principe de la modélisation inverse

2 Méthodes

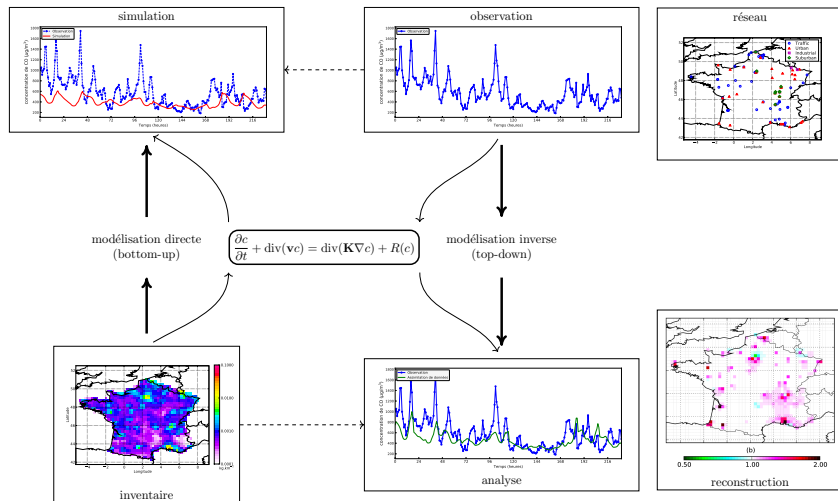
- Méthodes variationnelles
- Méthodes de filtrage d'ensemble
- Estimation des statistiques d'erreur

3 Exemples

- Gaz à effet de serre (global/régional)
- Monoxyde de carbone (erreur de représentativité)
- Composés organiques volatils (estimation d'erreur a priori)
- Radionucléides (quantification d'incertitude)

4 Conclusions

Approches *bottom-up* et approche *top-down*



Modélisation directe et inverse

► Modélisation directe :

$$\mu = \mathbf{H}\sigma + \varepsilon$$

- \mathbf{H} est le modèle (supposé ici linéaire : relation source-récepteur)
- μ est le vecteur des mesures
- σ est le vecteur des émissions
- ε est le vecteur des erreurs

► Modélisation inverse :

μ connu, mais ni σ , ni ε . Il faut définir une procédure objective permettant de passer de μ à σ , voire ε . Mais c'est un problème non-trivial (et même *mal-posé*) car

- \mathbf{H} n'est pas carrée en général,
- même si \mathbf{H} est carrée, \mathbf{H}^{-1} n'existe pas en général,
- même si \mathbf{H}^{-1} existe en théorie, \mathbf{H} est mal conditionnée et \mathbf{H}^{-1} n'a pas d'existence numérique,
- et le problème est bruité par ε .

Modélisation statistique du problème inverse

Ce que l'on suppose savoir :

- Modélisation directe :

$$\mu = \mathbf{H}\sigma_t + \varepsilon.$$

- Hypothèse statistique sur les erreurs d'observation :

$$\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R}).$$

- Ébauche (utilisation d'un inventaire a priori) :

$$\sigma \sim \mathcal{N}(\sigma_b, \mathbf{B}).$$

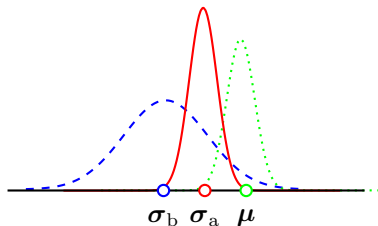
Estimation statistique optimale (BLUE)

- ▶ On veut trouver un estimateur σ_a qui
 - ▶ soit sans biais : $\mathbb{E}[\sigma_a - \sigma_t] = \mathbf{0}$,
 - ▶ minimise l'erreur a posteriori $\mathbb{E}[(\sigma_a - \sigma_t)(\sigma_a - \sigma_t)^T]$.

- ▶ On le choisit linéaire, de la forme $\sigma_a = \mathbf{L}\sigma_b + \mathbf{K}\mu$.

- ▶ Alors on montre que la solution (BLUE) est :

$$\begin{aligned}\sigma_a &= \sigma_b + \mathbf{K}(\mu - \mathbf{H}\sigma_b), \\ \mathbf{K} &= \mathbf{B}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}, \\ \mathbf{P}^a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}.\end{aligned}$$



Généralisation: loi de Bayes; le cadre mathématique idéal

- Inférence par la loi de Bayes/Laplace :

$$p(\sigma|\mu) = \frac{p(\mu|\sigma)p(\sigma)}{p(\mu)},$$

avec $p(\mu|\sigma)$ la **vraisemblance** des observations, $p(\sigma)$ l'**ébauche/prior** sur les émissions, et $p(\mu)$ est l'**évidence**. L'évidence correspond à une normalisation:

$$p(\mu) = \int d\sigma p(\mu|\sigma)p(\sigma).$$

- C'est une approche probabiliste. Elle quantifie l'**incertitude/l'information**. Elle ne fournit pas d'**estimateur** déterministe. Cela requiert un choix supplémentaire.
- L'approche bayésienne est très satisfaisante [Jaynes et al., 2003]. La plupart des méthodes d'assimilation de données et de modélisation inverse sont construites sur la loi de Bayes. Généralisation naturelle aux statistiques non-gaussiennes et modèles non-linéaires.
- Mais rarement facile à mettre en œuvre.

Outline

1 Principe de la modélisation inverse

2 Méthodes

- Méthodes variationnelles
- Méthodes de filtrage d'ensemble
- Estimation des statistiques d'erreur

3 Exemples

- Gaz à effet de serre (global/régional)
- Monoxyde de carbone (erreur de représentativité)
- Composés organiques volatils (estimation d'erreur a priori)
- Radionucléides (quantification d'incertitude)

4 Conclusions

3D-Var et interpolation optimale

- Formulation variationnelle du problème :

$$J(\sigma) = \frac{1}{2} \|\sigma - \sigma_b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \|\mu - \mathbf{H}\sigma\|_{\mathbf{R}^{-1}}^2,$$

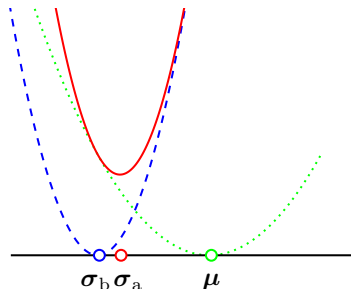
où $\|\sigma\|_{\mathbf{A}}^2 = \sigma^T \mathbf{A} \sigma$, qui est équivalent à BLUE.

- Interprétation probabiliste/bayésienne :

$$p(\sigma|\mu) \propto e^{-J(\sigma)},$$

- Capable de prendre en compte les opérateurs d'observation non-linéaires en utilisant des techniques d'optimisation non-linéaires standards :

$$J(\sigma) = \frac{1}{2} \|\sigma - \sigma_b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \|\mu - H(\sigma)\|_{\mathbf{R}^{-1}}^2.$$

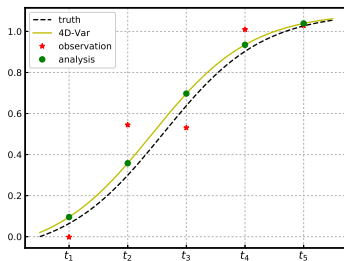
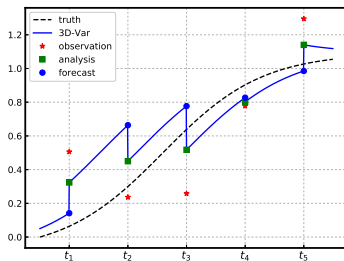


4D-Var

- 4D-Var : généralisation du 3D-Var où l'analyse est effectuée sur une fenêtre temporelle (typiquement 24h en météorologie). Fonction de coût :

$$J(\mathbf{x}_0, \sigma_{1:K}) = \frac{1}{2} \|\mathbf{x}_0 - \mathbf{x}_0^b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \sum_{k=1}^K \|\mathbf{y}_k - H_k(\mathbf{x}_k)\|_{\mathbf{R}_k^{-1}}^2 \quad \text{tq} \quad \mathbf{x}_k \stackrel{=}{=} M_k(\mathbf{x}_{k-1}, \sigma_k).$$

- du 3D-Var au 4D-Var :



- 4D-Var bien adapté à l'estimation de paramètres.

Le filtre de Kalman (d'ensemble)

► Similaire à l'interpolation optimale. Mais, \mathbf{B} est remplacé par un \mathbf{P}^f dynamique qui doit être mis à jour et propagé.

► Étape d'analyse :

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f \right),$$

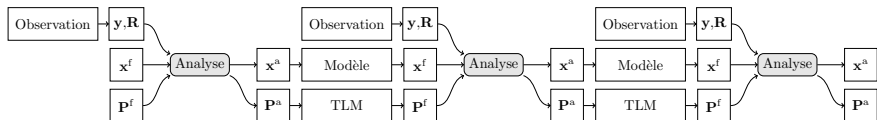
$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T \left(\mathbf{R}_k + \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T \right)^{-1},$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f.$$

► Étape de prévision :

$$\mathbf{x}_{k+1}^f = \mathbf{M}_{k+1:k} \mathbf{x}_k^a,$$

$$\mathbf{P}_{k+1}^f = \mathbf{M}_{k+1:k} \mathbf{P}_k^a \mathbf{M}_{k+1:k}^T + \mathbf{Q}_{k+1}.$$



Le filtre de Kalman (d'ensemble)

► Efficacité en grand dimension :

► Filtre de Kalman d'ensemble : remplacer approximativement $\mathbf{x}^f, \mathbf{P}^f$ par un ensemble dont les statistiques empiriques coïncident avec $\mathbf{x}^f, \mathbf{P}^f$.

► Estimation de paramètres avec un filtre :

► Le filtre de Kalman (d'ensemble) est adapté à l'estimation de l'état, pas nécessairement les paramètres comme les émissions. Le problème peut être contourné grâce à l'astuce de l'**état augmenté** :

$$\mathbf{x} \leftarrow \begin{pmatrix} \mathbf{x} \\ \sigma \end{pmatrix}$$

Estimer les statistiques d'erreur

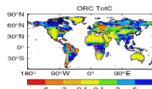
- Toutes ces méthodes reposent sur les statistiques d'erreur, **R**, **B**, dont la connaissance est **imparfaite**. Ces paramètres statistiques sont appelés **hyperparamètres**.
- Seul le filtre de Kalman (d'ensemble) s'appuie sur une estimation dynamique de **P^f**, mais également partiellement erronée.
- Les erreurs étant très significatives en chimie atmosphérique, aucune des méthodes présentées ne peut fonctionner efficacement avec des observations réelles sans une estimation plus fine des erreurs.
- Solution large gamme de méthodes empiriques et mathématiques pour les estimer : **χ^2** , **validation croisée**, **maximum de vraisemblance**, **estimateur non-biaisé du risque**, **courbe en L**, etc.

Outline

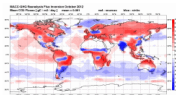
- 1 Principe de la modélisation inverse
- 2 Méthodes
 - Méthodes variationnelles
 - Méthodes de filtrage d'ensemble
 - Estimation des statistiques d'erreur
- 3 Exemples
 - Gaz à effet de serre (global/régional)
 - Monoxyde de carbone (erreur de représentativité)
 - Composés organiques volatils (estimation d'erreur a priori)
 - Radionucléides (quantification d'incertitude)
- 4 Conclusions

Gaz à effet de serre (échelle globale)

**Flux GES a priori
(inventaires "bottom-up",
process de modèles) avec
incertitude**

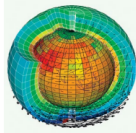


**Flux GES inversés
avec incertitude**

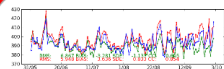


**Simulation des GES par le
transport atmosphérique**

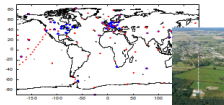
**Modèles/proxy de transport
avec des "erreurs modèle"**



**Inversion statistique :
corrections optimales qui
minimisent une somme
d'écarts aux mesures et à
l'ébauche**



**Comparaison
aux mesures**


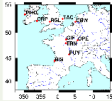
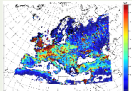

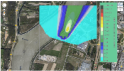
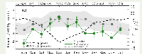
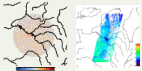
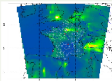
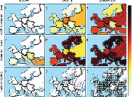
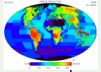


**Mesures de GES dans
l'atmosphère avec des
erreurs de mesure**

- Utilisé pendant 15 ans pour estimer les flux naturels de CO_2 à l'échelle globale.
- Émergence de systèmes d'inversion régionaux : estimation locale plus robuste, et meilleure analyse du transport.

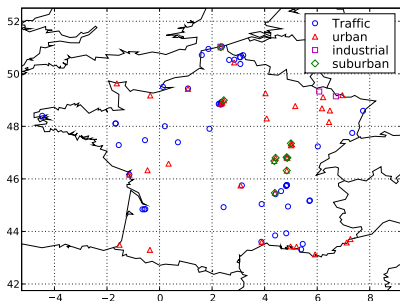
[Crédit : G. Broquet et F. Chevallier, LSCE]

Gaz à effet de serre (jusqu'à l'échelle urbaine)

Echelle	Sites industriels	Villes	National	Continental	Global
CO ₂		<p>Inversions (IO) sur des agglomération avec mesures continues et OSSES d'observation satellitaires (Paris et Recife)</p> 	<p>Inversions (IO) des flux naturels et anthropogéniques avec des mesures continues en France.</p> 	<p>Inversions (varionnelle et IO) dans flux naturels et anthropogéniques avec des mesures continues, avion et satellite.</p> 	<p>Inversions (varionnelle et IO) dans flux naturels et anthropogéniques avec des mesures continues et satellite.</p> 
CH ₄	<p>Inversion (IO) des émissions de décharges, eaux usées, raffinerie, fermes (instruments mobiles).</p> 	 			 <p>GCP-CH4 budget Global C-Atlas</p>

► Inversion des flux de gaz à effet de serre de la petite à l'échelle globale, utilisant des observations in situ, mobiles, et satellite. Panorama des activités au LSCE (G. Broquet).

Modélisation inverse des flux de CO à l'échelle régionale



► Stations du réseau BDQA : concentrations horaires de CO; mesures sur 80 stations.

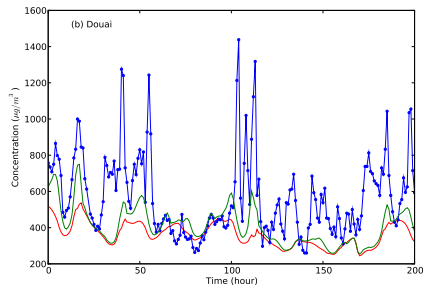
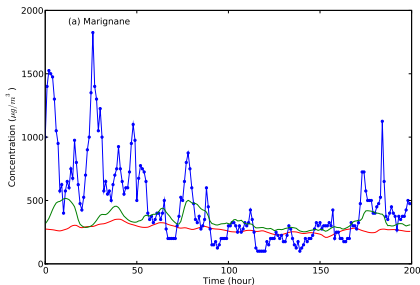
► Observations très fortement impactées par l'**erreur de représentativité** (trafic, stations urbaines).

► Grand nombre d'observations (environ 10^5 assimilées ici, 5×10^5 utilisées pour la validation).

► Espace de contrôle : flux et sources volumiques paramétrés avec environ 70×10^3 variables à une résolution de $0.25^\circ \times 0.25^\circ$.

→ Même avec une physique/chimie linéaire, le 4D-Var est une méthode de choix.

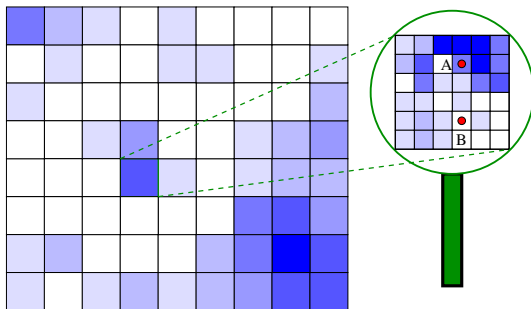
Résultats du 4D-Var traditionnel



► Courbes : observation, simulation, analyse.

► Impact colossal de l'erreur de représentativité !

Couplage du 4D-Var avec un modèle statistique sous-maille



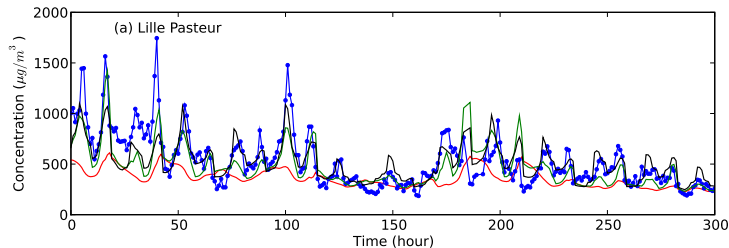
- On veut prendre en compte l'impact des sources proches qui engendrent les pics de concentration de CO visibles sur les chronologies :

$$\varepsilon_{\text{rep}} \simeq \xi \cdot \mathbf{\Pi} \longrightarrow \mu = \mathbf{H}\sigma + \xi \cdot \mathbf{\Pi} + \widehat{\varepsilon}.$$

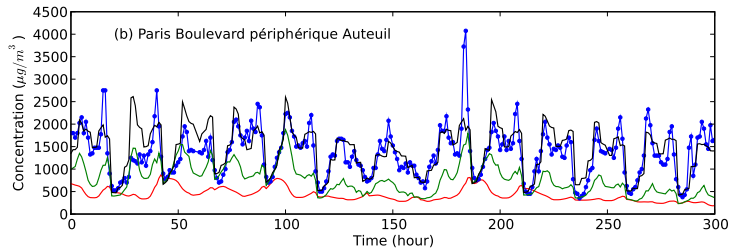
ξ : ensemble de coefficient statistiques (facteurs d'influence).

[Koohkan et Bocquet, 2012]

Résultats du 4D-Var- ξ : profils (1/2)



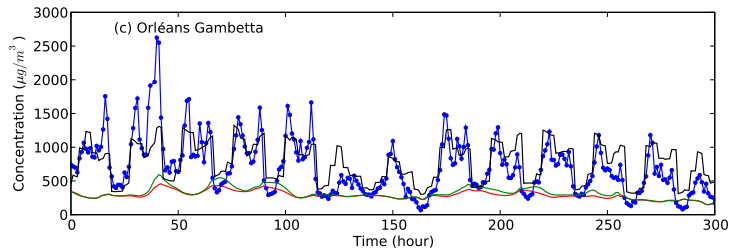
$$\xi_i = 0.6 \text{ h.}$$



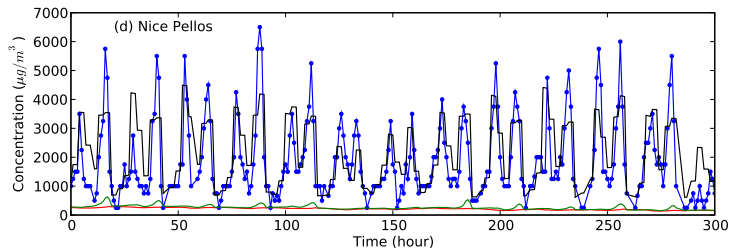
$$\xi_i = 2.7 \text{ h.}$$

► Courbes : observation, simulation, analyse, analyse multi-échelle

Résultats du 4D-Var- ξ : profils (2/2)



$$\xi_i = 11.9 \text{ h.}$$

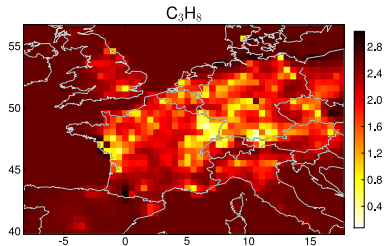
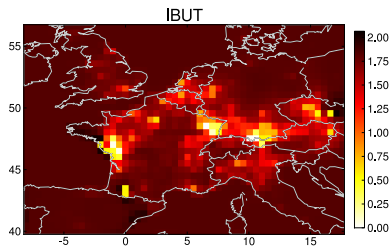


$$\xi_i = 45.8 \text{ h.}$$

► Courbes : observation, simulation, analyse, analyse multi-échelle

Modélisation inverse des émissions de COV non-méthaniques

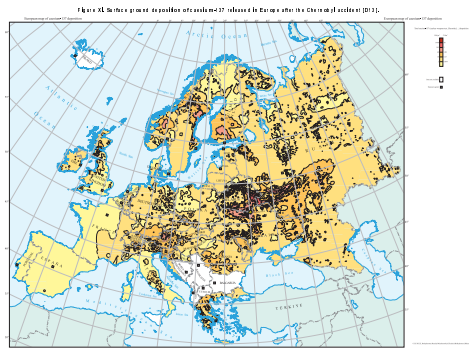
- Estimation des hyperparamètres (les statistiques d'erreur a priori).
- Approche dite **bayésienne empirique**.



- Premières inversions des flux COV à l'échelle européenne, et estimation de l'erreur a posteriori.

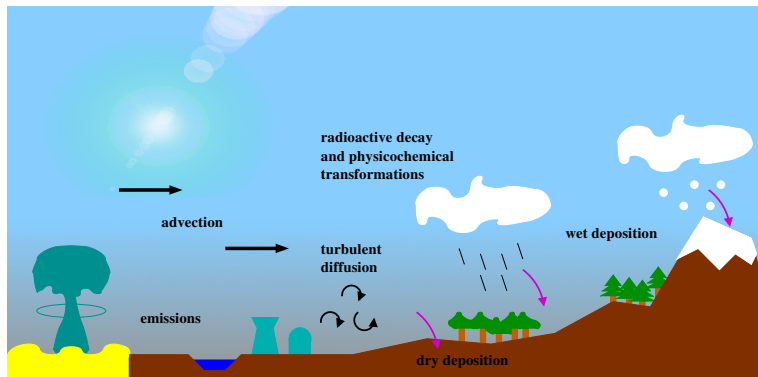
[Koohkan et al., 2013]

Accident de Tchernobyl



- ▶ 30 morts dans les premiers jours de l'accident
- ▶ Évacuation de 200 000 civils
- ▶ Zone d'exclusion de 30 km
- ▶ Impact sanitaire à moyen/long terme : cancer de la thyroïde (milliers de cas).

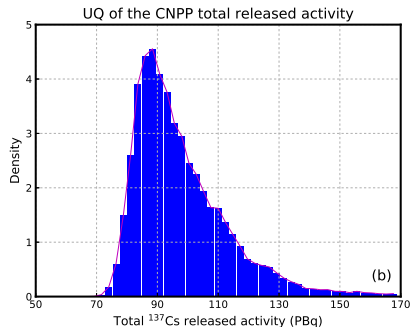
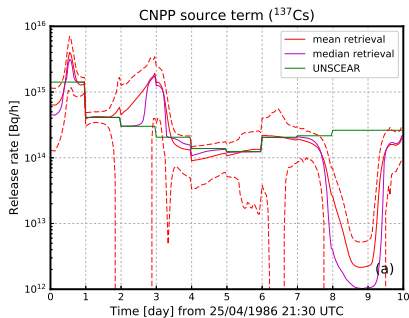
Dispersion des radionucléides : processus



- Modélisation par des EDPs des processus de transport, et paramétrisations micro-physiques et chimiques.
- Terme source (émission de radionucléides) inconnu.
- Paramètres des paramétrisations physiques souvent mal connus (diffusion turbulente effective, paramètres du lessivage et du dépôt sec).

Application de techniques avancées de quantification d'incertitude

- Modélisation inverse sous hypothèses log-normales sur les erreurs + analyse bayésienne hiérarchique qui inclue les hyperparamètres + solution numérique par MCMC + analyse transdimensionnelle pour réduire le coût calcul.



[Liu et al., 2017]

Outline

- 1 Principe de la modélisation inverse
- 2 Méthodes
 - Méthodes variationnelles
 - Méthodes de filtrage d'ensemble
 - Estimation des statistiques d'erreur
- 3 Exemples
 - Gaz à effet de serre (global/régional)
 - Monoxyde de carbone (erreur de représentativité)
 - Composés organiques volatils (estimation d'erreur a priori)
 - Radionucléides (quantification d'incertitude)
- 4 Conclusions

Conclusions

- ▶ Les techniques de **modélisation inverse** sont désormais appliquées à l'estimation des émissions de polluants, avec plus de **rigueur**.
- ▶ Les statistiques erreurs sont maintenant également estimées, ce qui rend beaucoup plus **fiables** les inversions.
- ▶ Quantification d'incertitude pour la **modélisation directe**, mais aussi pour la **modélisation inverse**, i.e. quelle confiance attribuer à des émissions inversées ?
- ▶ Piste explorée actuellement : **réduction dimensionnelle** pour les plus grands systèmes (grand nombre de flux à inverser).
- ▶ Modélisation inverse à toutes les échelles : **du global à l'urbain** !

References

- [1] M. Bocquet.
Parameter field estimation for atmospheric dispersion: Application to the Chernobyl accident using 4D-Var.
Q. J. R. Meteorol. Soc., 138:664–681, 2012.
- [2] A. Farchi, M. Bocquet, Y. Roustan, A. Mathieu, and A. Quérel.
Using the Wasserstein distance to compare fields of pollutants: application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident.
Tellus B, 68:31682, 2016.
- [3] M. R. Koohkan and M. Bocquet.
Accounting for representativeness errors in the inversion of atmospheric constituent emissions: Application to the retrieval of regional carbon monoxide fluxes.
Tellus B, 64:19047, 2012.
- [4] M. R. Koohkan, M. Bocquet, Y. Roustan, Y. Kim, and C. Seigneur.
Estimation of volatile organic compound emissions for Europe using data assimilation.
Atmos. Chem. Phys., 13:5887–5905, 2013.
- [5] Y. Liu, J.-M. Haussaire, M. Bocquet, Y. Roustan, O. Saunier, and A. Mathieu.
Uncertainty quantification of pollutant source retrieval: comparison of bayesian methods with application to the Chernobyl and Fukushima-Daiichi accidental releases of radionuclides.
Q. J. R. Meteorol. Soc., 0:0–0, 2017.
Submitted.
- [6] O. Saunier, A. Mathieu, D. Didier, M. Tombette, D. Quélo, V. Winiarek, and M. Bocquet.
An inverse modeling method to assess the source term of the Fukushima Nuclear Power Plant accident using gamma dose rate observations.
Atmos. Chem. Phys., 13:11403–11421, 2013.
- [7] V. Winiarek, M. Bocquet, N. Duhanyan, Y. Roustan, O. Saunier, and A. Mathieu.
Estimation of the caesium-137 source term from the Fukushima Daiichi nuclear power plant using a consistent joint assimilation of air concentration and deposition observations.
Atmos. Env., 82:268–279, 2014.
- [8] V. Winiarek, M. Bocquet, O. Saunier, and A. Mathieu.
Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: Application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant.
J. Geophys. Res., 117:D05122, 2012.