Localisation de sources polluantes depuis un réseau de capteurs

A. Ickowicz, F. Septier, Y. Delignon, P. Armand, H. Snoussi, P. Honeine
{ickowicz,septier,delignon}@telecom-lille1.eu, armand@cea.fr, {snoussi,honeine}@utt.fr

Le Problème

L'estimation inverse des caractéristiques de la source polluante depuis un nombre fini de mesures bruitées de concentration de polluants provenant d'un réseau de capteurs a reçu beaucoup d'attention ces dernières années. Néanmoins, de nombreux problèmes persistent lorsque l'objectif est d'avoir une identification complète et précise lors de la présence de multiples sources et ce en un minimum de temps de traitement. Nous présentons dans ce poster un algorithme permettant de réduire ce temps de traitement tout en conservant une grande précision.

L'approche stochastique

A cause de la complexité du modèle atmosphérique, l'approche bayésienne est la plus utilisée pour l'estimation des termes de source. Si l'on note ζ les erreurs de mesure de fonction de probabilité $\mathcal{L}_\zeta$, θ les termes de source à estimer, $R^{obs(i)}$ les observations du capteur i, et $D^{true(i)}$ le modèle 1 appliqué en i avec θ, alors la loi a posteriori de θ s'écrit:

$$p(\theta|R^{obs(i)}, I) \quad \propto \quad \mathcal{L}_\zeta(R^{obs(i)}, D^{true(i)}(\theta))p_0(\theta)$$

L'utilisation d'un algorithme d'échantillonnage fournira la distribution a posteriori de θ

L'approche déterministe

Basée sur la théorie de l'optimisation, la détermination des termes de source est traitée comme un problème d'optimisation. On construit un modèle de dispersion pour générer les observations. Les techniques d'optimisation sont ensuite employées pour ameliorer itérativement l'adéquation des observations et des données générées jusqu'à satisfaction d'un critère de convergence appliqué à une fonction du type:

$$f(\theta) \quad = \quad \min_\theta \|R^{obs(i)} - D^{true(i)}(\theta)\|$$

La conclusion

Les deux catégories de méthodes existantes pour l'estimation des termes de source ont leurs défauts. Les méthodes statistiques sont très lourdes à mettre en oeuvre, ce qui est rédibitoire dans le cas d'une réponse urgente. Les méthodes d'optimisation, plus rapides, souffrent de par leur considération déterministe, leurs utilisateurs niant l'existence d'un aléatoire suffisamment puissant pour influencer les observations. Toutefois, les capteurs ne sont pas si fiables. Le modèle de dispersion utilisé n'est pas si près de la réalité, en particulier dans des configurations géographiques complexes. Les solutions envisagées semblent répondre positivement aux griefs émis ci-dessus.

[1] Wilson, J.D., Sawford, B.L., *Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere*. Boundary-Layer Meteorology, 87,191–210,1996

[2] Zheng, X., Chen, Z., *Inverse calculation approaches for source determination in hazardous chemical releases*. Journal of Loss Prevention in the Process Industries, 2011

[3] Cornuet, J.M., Marin, J.M., Mira, A., Robert, C.P., *Adaptive multiple importance sampling*. Scandinavian Journal of Statistics, to appear

[4] Yee, E., *Theory for reconstruction of an unknown number of contaminant sources using probabilistic inference*. Boundary-layer meteorology, 127,359–394,2008

Le modèle de dispersion

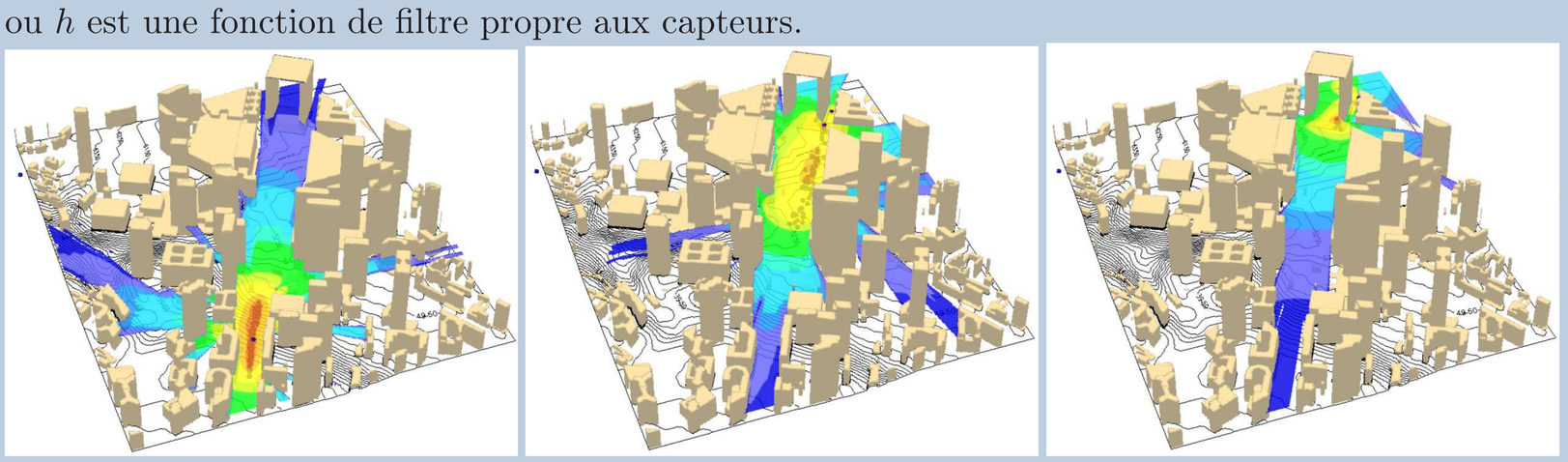
Les observations des capteurs se présentant sous la forme d'une concentration du produit CBR relaché, nous devons nous appuyer sur les modèles de dispersion atmosphérique de type Lagrangien [1]. Ceux-ci, largement étudiés dans la litterature, se présentent sous la forme:

$$\begin{aligned} \frac{\partial C}{\partial t} + u\nabla C - \nabla(K\nabla C) \quad &= \quad Q \\ \text{s.c. } \nabla_n C = 0 \text{ sur } \partial\Omega & \end{aligned} \tag{1}$$

ou C représente la concentration de produit, u le champ de vent, K la constante de Kolmogorov, Q la fonction d'émission et Ω le domaine d'intérêt. Les observations sont de la forme:

$$D_{t_j}^{true(i)}(\theta) = \int_0^T \int_\Omega C(x,t)h(x,t|x_i,t_j)dxdt \tag{2}$$

ou h est une fonction de filtre propre aux capteurs.



Les méthodes existantes

Pour estimer les termes de sources, deux tendances majeures peuvent se dégager [2], dites *optimisation*, et *stochastique*. Parmi les méthodes stochastiques, on notera:

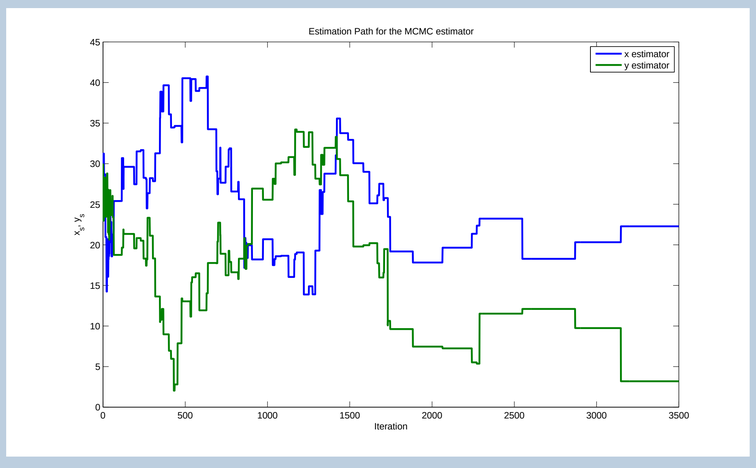
- **Les méthodes MCMC.** Elles sont basées sur le parcours d'une chaine de Markov ayant pour loi stationnaire la distribution cible.
- **Les méthodes MCMC associées aux équations backwards**. L'expression D^{true} se retrouve simplifiée en une intégrale simple du temps, puis les MCMC sont utilisées.

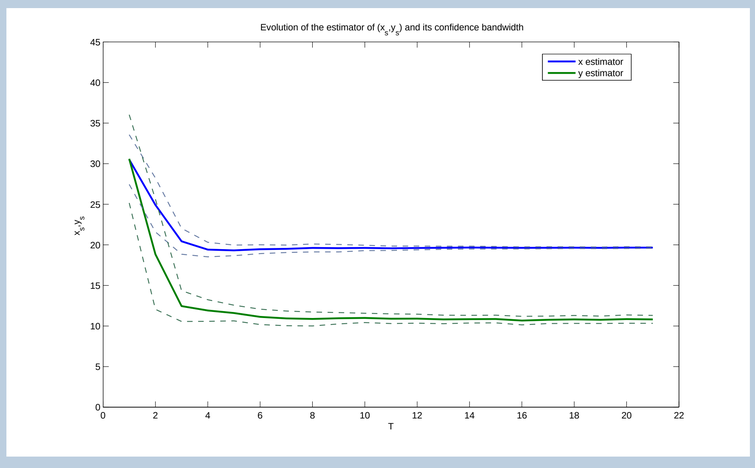
Pour les méthodes déterministes:

- **La méthode du gradient.** L'optimisation se fait à l'aide des dérivées de la fonction f.
- **Les méthodes d'apprentissage.** On citera notamment les réseaux de neurones. Un modèle inverse est entrainé sur un grand nombre d'observations (générées par le modèle de dispersion), puis utilisé pour minimiser f.
- **Les méthodes basées sur le principe de l'algorithme génétique.** Des échantillons sont générés, les plus adaptés aux observations sont conservés, et servent de base pour la génération d'autres échantillons. L'adequation étant jugée à l'aide de f.

Les contributions et perspectives

- **Les méthodes d'échantillonnage adaptatives.** On citera notemment l'algorithme AMIS [3], basé sur les techniques dites d'*importance sampling*,qui adapte la loi de proposition à chaque itération. De cette façon, la convergence vers les paramètres d'intérêt se fait plus rapidement que par un algorithme MCMC tel que proposé par Yee [4]:





- **L'apprentissage statistique.** L'utilisation des processus Gaussiens, notamment, permettrait l'apprentissage d'un modèle de dispersion inverse moins *gourmand*, tout en conservant un caractère alátoire.