

# Stage de fin d'études : *Prévision d'ensemble en qualité de l'air avec agrégation séquentielle de modèles*

Vivien Mallet (ENPC) et Gilles Stoltz (ENS Paris)

Décembre 2006

## Sujet

La prévision de la qualité de l'air repose traditionnellement sur les simulations numériques de modèles de chimie-transport. En s'appuyant sur des prévisions météorologiques, ces modèles simulent les concentrations de polluants (par exemple, l'ozone sur l'Europe) pour le lendemain et le surlendemain. Une limitation importante réside dans les incertitudes élevées à la fois dans les données d'entrée (champs météorologiques, émissions, ...) et dans les paramétrisations physiques (turbulence, ...). Dans ce contexte, la prévision d'un seul modèle est insuffisante. La prévision doit être fondée sur un ensemble de modèles représentatif de l'incertitude : on parle de prévision d'ensemble. L'objectif du stage est d'étudier des méthodes d'agrégation (combinaisons linéaires) de multiples prévisions afin de constituer une meilleure prévision.

Le travail sera effectué autour du système de simulation Polyphemus (<http://www.enpc.fr/cerea/polyphemus/>) qui effectue des prévisions d'ensemble quotidiennement. Il s'agira d'implémenter, de développer et de proposer des méthodes de combinaison séquentielle de modèles, qui soient performantes en prévision. Une première série de méthodes sera constituée par des algorithmes récents de prédiction issus de la communauté des suites individuelles ; ces méthodes combinent toutes linéairement les modèles avec des poids appris sur la base de leurs performances passées (écart aux observations). Elles garantissent que quelle que soit la suite des observations, la performance de l'algorithme d'apprentissage est presque aussi bonne que la meilleure combinaison linéaire constante des modèles (et donc, en particulier, que le meilleur modèle). Dans un deuxième temps, il s'agira de développer et de proposer de nouvelles méthodes, qui puissent corriger les éventuels défauts des méthodes de la première série, ou de s'intéresser à la sélection séquentielle d'un sous-ensemble de modèles performants.

Les méthodes implémentées seront appliquées à la prévision des concentrations d'ozone, et peut-être d'aérosols (selon la disponibilité de prévisions d'ensemble). Elles pourront vraisemblablement être mises en œuvre sur la plate-forme de prévision opérationnelle Prév'air (<http://www.prevair.org/>), ce qui constituerait une valorisation supplémentaire du travail.

## Conditions et contacts

**Profil :** dernière année d'école d'ingénieur ou équivalent, avec un goût pour les mathématiques appliquées à la modélisation numérique et/ou statistique

**Poursuite en thèse :** envisageable au CEREAs (voir ci-dessous) sur l'assimilation de données

**Début du stage :** avant juin 2007

**Durée :** 5 à 6 mois

**Rémunération et statut :** 800 à 1000 euros nets par mois, financement sur le projet ANR (Agence nationale de la recherche) ATLAS (« from Applications to Theory in Learning and Adaptive Statistics »)

**Localisation (double) :**

- École des Ponts (Champs sur Marne, près de la ligne de RER A), laboratoire CEREAL (<http://www.enpc.fr/cerea/>, laboratoire commun ENPC / EDF R&D, projet CLIME commun avec l'INRIA), et,
- École normale supérieure (Paris, un à deux jours par semaine), laboratoire DMA (<http://www.dma.ens.fr/>)

**Responsables du stage :** Vivien Mallet (ENPC) et Gilles Stoltz (ENS Paris)

**Contacts :** [Vivien.Mallet@cerea.enpc.fr](mailto:Vivien.Mallet@cerea.enpc.fr) (01 64 15 21 47)  
[Gilles.Stoltz@ens.fr](mailto:Gilles.Stoltz@ens.fr) (01 44 32 32 77)