



Proposition de stage de M2  
Mathématiques appliquées

## Prévision de qualité de l'air par agrégation automatique de modèles numériques



Vivien Mallet, INRIA & Gilles Stoltz, CNRS

---

### Contexte

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 comme à la figure 1) 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, etc.) et dans les paramétrisations physiques (turbulence, etc.). 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 (figure 1). Une approche développée par l'équipe d'accueil et aujourd'hui bien maîtrisée consiste à agréger (par combinaison linéaire) les prévisions d'un ensemble de modèles. Les méthodes employées, dites d'apprentissage automatique, garantissent que la prévision agrégée est en moyenne meilleure que toute prévision de l'ensemble. Leurs propriétés mathématiques et leurs performances pratiques en font les méthodes opérationnelles de référence.

Les travaux antérieurs, notamment conduits lors de précédents stages, sont venus enrichir le système de simulation **Polyphemus**, qui est à la fois un outil de recherche et un outil opérationnel (il fournit quotidiennement des prévisions de la qualité de l'air sur l'Europe). À ce titre, les travaux intéressent l'INERIS et sa **plate-forme Prév'air** qui utilise déjà, de manière opérationnelle, les méthodes d'apprentissage automatique implémentées dans Polyphemus.

### Objectifs possibles de ce stage

Le stage commencera bien sûr par une mise à niveau bibliographique, tant sur les méthodes d'apprentissage séquentiel étudiées que sur les résultats précédemment obtenus. Plusieurs problèmes de recherche se posent aujourd'hui. **Un ou plusieurs d'entre eux pourront être abordés** par le stagiaire, selon ses compétences et ses intérêts.

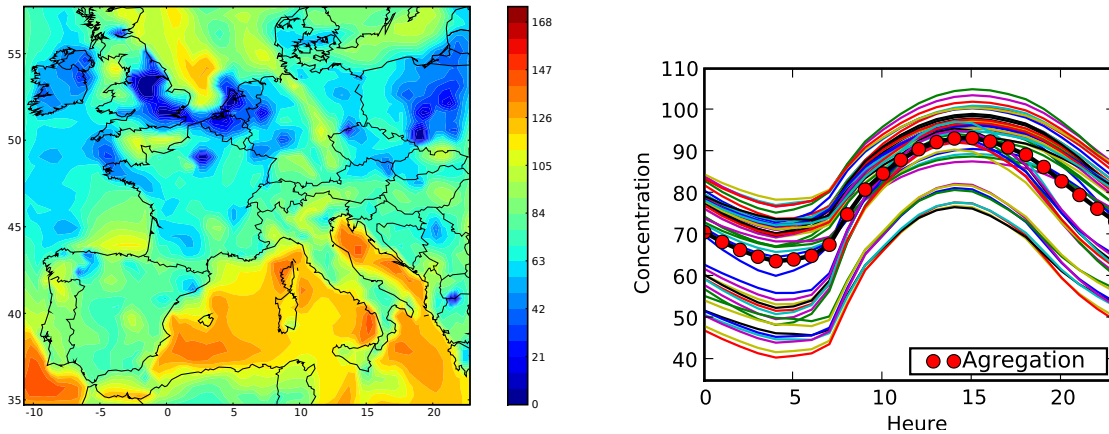


FIGURE 1 – (Gauche) Exemple de prévision de la pollution à l’ozone (concentration au sol, en  $\mu\text{g m}^{-3}$ ). (Droite) Profils journaliers d’ozone pour 48 modèles d’un ensemble (concentrations en  $\mu\text{g m}^{-3}$ ). À chaque échéance de prévision, les méthodes d’apprentissage automatique associent un poids à chaque modèle et agrègent leurs prévisions par combinaison linéaire.

### Problème 1 : Gestion des incertitudes

Les prévisions des modèles sont accompagnées d’une mesure de leur incertitude. Par exemple, une prévision est fournie avec une mesure de déviation possible donnée par un écart type. Les observations attendues se situent alors dans une certaine région de l’espace (un ellipsoïde de confiance). On peut interpréter chaque élément de cet ellipsoïde comme correspondant à un certain pseudo-expert et agréger les prévisions de ces experts selon leurs performances passées. Il s’agira de formaliser mathématiquement ce cadre, de l’implémenter en pratique et de comparer les résultats obtenus avec les méthodes plus classiques.

### Problème 2 : Agrégation de champs structurés

Il s’agit de prédire des champs de concentration d’ozone sur l’Europe. Des travaux spécifiques ont été précédemment menés afin d’agréger les champs de concentration prévus par les experts. Les champs agrégés sont capables de reconstituer en moyenne les structures spatiales attendues, mais les prévisions agrégés pour une seule échéance ne présentent pas toujours les bonnes structures. On souhaite donc développer une approche pour agréger des champs en prédisant mieux les structures. Il se pose notamment la question de la bonne métrique d’erreur entre deux champs structurés (la « perte », dans le vocabulaire de l’apprentissage) et des performances théoriques associées.

### Problème 3 : Comparaison avec l’agrégation par filtrage

Une autre forme d’agrégation par filtrage, dans un cadre stochastique ou purement déterministe, est à l’étude. Cette approche permet de donner une dynamique aux poids et d’appliquer des techniques de filtrage (Kalman ou minimax) pour les estimer. De plus, elle permet potentiellement d’apporter une information sur les incertitudes. Le stage pourra être l’occasion de développer ces méthodes et de les comparer aux méthodes d’apprentissage automatique classiques.

## Poursuite possible en thèse

Une continuation en thèse, sur les mêmes thèmes et dans la même lignée de recherche, est envisageable (sous réserve de financement).

## Conditions et contacts

**Profil** : M2 recherche ou dernière année d'école d'ingénieur, avec goût pour la modélisation (numérique ou statistique)

**Début du stage** : dès que possible en 2012

**Durée** : 5 à 6 mois

**Rémunération** : 417 euros nets par mois

**Statut** : convention de stage entre l'établissement de rattachement du M2 et l'INRIA

**Bi-localisation** : INRIA, [site de Rocquencourt](#) (près de Versailles, navettes depuis Paris), dans l'équipe-projet [CLIME](#), et [ENS de Paris](#)

**Responsables du stage** : Vivien Mallet (INRIA – CEREAs) et Gilles Stoltz (CNRS – ENS Paris – INRIA – HEC Paris)

**Contacts** :

– [vivien.mallet@inria.fr](mailto:vivien.mallet@inria.fr) (01 39 63 55 76)

– [gilles.stoltz@ens.fr](mailto:gilles.stoltz@ens.fr) (01 44 32 32 77)

N'hésitez pas à demander des précisions supplémentaires ! Les conditions ci-dessus peuvent bien sûr être aménagées.