

Proposition pour la création d'une Action de Recherche Coopérative : Assimilation de données en qualité de l'air (ADOQA)

M. Bocquet, I. Herlin, F. Legland, F.X. Le Dimet, L. Rouil, B. Sportisse

2 décembre 2004

1 Introduction

Les modèles de qualité de l'air (de transport réactif de polluants dans l'atmosphère de manière générale) ont à présent atteint un niveau de qualité suffisant pour qu'ils puissent être utilisés de manière prédictive dans un contexte opérationnel. Un bon exemple est fourni par la plate-forme de prévision Prév'Air mise en oeuvre par l'INERIS (Institut National de l'Environnement Industriel et des Risques, www.prevoir.org).

De nombreux problèmes à caractère méthodologique restent néanmoins posés : un point clé est le couplage entre modèles numériques (ce que l'on désigne communément dans ce domaine comme les modèles de Chimie-Transport) et données mesurées. Ceci nécessite, de manière relativement classique à présent, le recours à des méthodes d'assimilation de données (qu'elles soient variationnelles, relevant du contrôle optimal, ou séquentielles, relevant de la théorie de l'estimation).

L'objectif de cette proposition d'ARC est d'explorer deux aspects différents liés à cette problématique, que nous pensons être stratégiques pour les applications et scientifiquement pertinents car encore largement non résolus.

Le premier point concerne la méthodologie en général et la recherche d'une densité de probabilité (plutôt qu'une valeur moyenne ou vraisemblable) pour décrire l'état du système : c'est ce que l'on désigne parfois par l'assimilation de données non linéaire. Ce point est évidemment crucial dans une optique de "risque environnemental" (au delà de la qualité de l'air d'ailleurs) pour laquelle la prévision d' "événements extrêmes" peut être la plus pertinente.

Le second point est spécifique à une source de données en plein essor, les données satellitaires. Actuellement, les données de capteurs terrestres sont largement utilisées mais restent lacunaires (en couverture spatiale et temporelle). L'enjeu est donc de définir de bonnes méthodologies d'assimilation de données satellitaires dans un tel contexte.

Ces deux points, largement indépendants, font intervenir plusieurs projets de compétence complémentaire à l'INRIA et il nous a semblé important de les fédérer au sein de cette ARC : CLIME pour la donnée satellitaire et la thématique visée, IDOPT pour l'assimilation de données et ASPI pour les méthodes probabilistes. La cohérence applicative est par ailleurs donnée d'une part par CLIME, d'autre part par l'implication d'un

acteur opérationnel l'INERIS, pouvant être intéressé par les sorties de ce programme de recherche.

2 Contexte Scientifique

La dynamique de l'atmosphère est gouvernée par les équations primitives. Celles-ci sont non-linéaires (second terme de la dérivée convective de la vitesse principalement). Elles sont connues pour, de ce fait, être de nature chaotique. Pour cette raison, l'utilisation d'un système d'assimilation est essentielle dans l'amélioration des performances des centres de prévisions météorologiques.

Lorsqu'on utilise l'assimilation 4D-Var, la prise en compte des non-linéarités de la dynamique est naturelle et sans approximation. On peut aussi avoir recours à une assimilation séquentielle du type filtre de Kalman, qui présente l'avantage d'être naturelle dans son interprétation et simple dans son implémentation numérique. Cependant, elle présente deux inconvénients majeurs. Le premier est que le filtre de Kalman et ses variantes utilisent de façon explicite les matrices de covariances d'ébauche, de taille considérable. Ce écueil peut être évité si l'on a recours des méthodes de réduction. Le second écueil est que le filtre requiert la linéarisation des équations de la dynamique du système. C'est donc une approximation pour un système non-linéaire. Le système d'assimilation peut en conséquence ne pas parvenir à prévoir des transitions dynamiques (par exemple le passage d'un attracteur à l'autre dans le cas du modèle de Lorenz). Cette réflexion s'applique également à la dynamique de l'océan, et son assimilation, mais sur une échelle de temps plus grande.

Une question importante est donc de savoir comment, dans le cadre d'une assimilation séquentielle, prendre en compte pleinement la non-linéarité d'un système. Nous souhaitons plus spécifiquement étudier ce point dans le cas du transport réactif atmosphérique. C'est le premier point de notre proposition.

Le deuxième point de notre proposition concerne l'enjeu opérationnel très important de l'assimilation de mesures de concentrations de constituants de l'atmosphère mesurées par satellite. Les performances des capteurs spatiaux de la composition de l'atmosphère augmentent continuellement, et permettent à présent des mesures de la composition de la troposphère (voir par exemple [5, 4] pour l'ozone et [3] pour les aérosols). Cependant, l'obtention des concentrations dans la troposphère est le résultat d'un processus complexe de modélisation inverse de la mesure radiative brute des sondeurs spatiaux. Il s'agit d'un problème notoirement mal posé, rendu plus difficile encore par le fait que la contribution de la troposphère au signal mesurée est petite comparée à celle de la stratosphère. La résolution de ce problème inverse fait intervenir des hypothèses sur la répartition verticale des constituants, sur la composition de la stratosphère, et fait appel à des méthodes de type *look-up table*, où l'on compare la mesure du satellite à un ensemble de sorties d'un modèle radiatif pour des configurations typiques de l'atmosphère. Nous maîtrisons donc mal l'ensemble des hypothèses sous-tendant le processus d'inversion, ni a fortiori l'incertitude attachée à son résultat. Un enjeu majeur, qui constitue le deuxième point de notre proposition, concerne l'étude de l'assimilation des mesures radiatives acquises par satellite dans les modèles de chimie transport, au lieu des mesures de concentrations chimiques obtenues par inversion d'un modèle radiatif, dans la lignée des récents développements en météorologie [1]. Cela peut s'aborder en considérant le

modèle de transfert radiatif comme un opérateur d'observation fortement non-linéaire. Se pose alors la question de la discrétisation optimale du modèle radiatif, étant données d'une part la nature des informations chimiques issues du modèle de chimie, et d'autre part la nature des mesures satellitaires.

3 Proposition de recherche

3.1 méthodes avancées en assimilation de données

La question est, nous l'avons dit de savoir comment, dans le cadre de l'assimilation séquentielle qui aille au delà du filtre de Kalman, prendre en compte la non-linéarité du système.

Le système d'assimilation parfait devrait propager la densité de probabilité [pdf] des variables d'état. Mais actuellement, le filtre de Kalman et ses variantes les plus simples, ne propagent que les espérances et les seconds moments. En conséquence :

- Ce type de méthode rend assez mal compte de l'évolution des pdfs lorsque le système est chaotique. Partant d'une distribution gaussienne, rien ne garantit que la distribution reste unimodale. Enfin, si le système est non-linéaire, mais pas nécessairement chaotique, le filtre de Kalman ne peut pas rendre compte fidèlement de l'évolution d'une distribution multimodale.
- Même si la distribution rester unimodale, le filtre de Kalman ne décrit pas correctement l'apparition de queues de distributions (événements rares mais importants dans le contexte du risque environnemental où l'on peut être intéressé par la prévision d'un état extrême).

La transposition algorithmique des équations exactes de l'assimilation séquentielle est ce qu'on appelle *filtre particulière*. Une implémentation simple est connue sous le nom de filtre SIS (Sequential Importance Sampling) et une version plus évoluée sous le terme SIR (Sequential Importance Resampling). Lorsque l'on veut faire porter l'assimilation séquentielle sur une variable particulière, on échantillonne sa pdf, puis on propage les membres de l'échantillon par la dynamique temporelle. Pour l'étape d'analyse, on applique la formule de Bayes, ce qui a pour effet d'affecter un poids statistique à chacun des membres de l'échantillon. Ces deux étapes sont rigoureuses. La difficulté réside alors dans le choix d'une méthode permettant de re-tirer l'échantillon affecté par les poids, avant de renouveler le cycle. Cela est nécessaire car l'absence de ré-échantillonnage entraîne une dégénérescence du processus (les poids se concentrent sur quelques membres) [8]. Pour un système de grande taille (état physique de l'atmosphère, de l'océan, composition chimique de l'atmosphère), l'application directe de cette stratégie semble impossible car la représentativité d'un échantillon statistique (et sa survie à travers le filtre) exigerait un nombre trop important d'états. La définition du ré-échantillonnage serait un problème épineux qui exige que la dynamique du système puisse être fortement réduite.

Il y a peu de tentatives en géosciences d'adaptation de cette approche. On peut toutefois citer

- Le filtre de Kalman d'ensemble de Geir Evensen (Norsk Hydro, Bergen, Norvège), mais qui reste un filtre gaussien fondé sur une approche Monte Carlo empirique (et insuffisamment fondée mathématiquement).
- Le filtre SEEK (Singular Evolutive Extended Kalman) et surtout le filtre SEIK

(Singular Evolutive Interpolated Kalman) [7, 6] que l'on doit à Dinh Tuan Pham (Projet IDOPT de l'INRIA), qui demeurent essentiellement des filtres de type Kalman, mais dont la construction est fondée sur le constat fait plus haut sur le filtre particulière.

- La tentative de P. J. Leeuwen (Université d'Utrecht) sur un modèle simplifié l'océan [9]. Parce que le système est de taille très importante, il insiste sur la nécessité de l'appliquer localement en décomposant l'état de l'océan sur une base d'EOFs (Empirical Orthogonal Functions).
- On peut également citer les travaux académiques non publiés de Hans Wackernagel (École des Mines de Paris) et d'Olivier Talagrand (Laboratoire de Météorologie Dynamique).

La dynamique du système de prévision de la qualité de l'air est celle de la composition chimique de la basse troposphère :

$$\frac{\partial c}{\partial t} + \operatorname{div}(\mathbf{u}c) - \operatorname{div}(\mathbf{K}\nabla c) + \Lambda c = \sigma + \chi(c), \quad (1)$$

où c est le vecteur des champs des espèces chimiques. Cette dynamique est foncièrement non-linéaire : elle provient du terme de chimie $\chi(c)$ dès lors qu'une réaction est au moins d'ordre deux. En revanche, dans la limite d'un transport négligeable par rapport à la chimie, elle n'est pas chaotique. Son spectre de Lyapunov est donc surtout négatif. (On ne peut toutefois pas exclure à certaines échelles, un régime chaotique dû à l'advection, lorsque le transport concurrence la chimie). Contrairement à la dynamique météorologique dont l'incertitude est intrinsèque, l'incertitude provient ici des inventaires d'émissions (autrement dit des forçages), et des champs météorologiques (qui sont considérés comme donnés).

Au travers de cette ARC, notre ambition est d'étudier l'intérêt de l'application de méthodes d'assimilation avancées aux équations de la chimie-transport et de proposer des solutions sur des systèmes prototypes.

Parce que la dynamique est de nature foncièrement différente de celle de la dynamique atmosphérique, il semble important de débiter par une étude théorique qui devra évaluer l'intérêt d'une bonne représentation de la non-linéarité du système, respectivement pour :

1. le suivi de la dynamique du modèle direct (contexte de la prévision d'ensemble, propagations d'incertitudes),
2. l'assimilation de données à proprement parler.

L'application directe sans réflexion amont des méthodes avancées de prise en compte de la non-linéarité sur un modèle de chimie-transport serait hasardeuse. Toutefois, des études numériques exploratoires concernant le deuxième point mentionné (prévision d'ensemble) sont possibles et d'un grand secours dans la réflexion.

Cette étude théorique sera suivie et complétée par l'étude et la conception d'algorithmes pertinents appliqués à des systèmes réels : en fonction des réponses apportées lors de la première phase, on pourra développer de nouvelles méthodes et les appliquer à des systèmes de grande taille. En tout état de cause, leur application sera concomitante à des méthodes de réduction de la dynamique (filtres locaux, décomposition en EOFs, ou autre).

3.2 Problématique de l’assimilation de données provenant de l’acquisition satellitaire

Comme indiqué précédemment, nous nous intéressons ici au problème de l’assimilation des mesures radiatives issues des capteurs spatiaux de sondage de l’atmosphère.

Du point de vue de la modélisation directe, la radiance mesurée en sortie de l’atmosphère dépend de la répartition verticale de ses constituants et de la température, par le biais de modèles de transfert radiatif. Ces modèles expriment la radiance monochromatique $I_\lambda(\Omega)$ pour une direction de visée Ω par une équation différentielle intégrale de type :

$$\frac{dI_\lambda}{d\tau}(\tau, \Omega) = I_\lambda(\tau) - \omega_{a,\lambda}B(\lambda, T) + \frac{\omega_{d,\lambda}}{4\pi} \int_{\Omega'} P(\Omega, \Omega') I_\lambda(\Omega') d\Omega' \quad (2)$$

dont les termes expliquent respectivement l’absorption du rayonnement, l’émission thermique à la température T , et enfin (terme intégral), la diffusion par les constituants de l’atmosphère (gaz, aérosols) des rayonnements venant des directions Ω' . Cette équation appelle plusieurs commentaires :

- La dépendance envers la composition de l’atmosphère intervient principalement par le biais de la variable d’intégration, la profondeur optique τ , qui dépend de la concentration du milieu en absorbants et diffusants et de leurs propriétés spectrales. Par ailleurs, les coefficients spectraux, tels ω_a , ω_b et $P(\Omega, \Omega')$ (respectivement albédo d’absorption, albédo de diffusion et fonction de phase) dépendent des propriétés spectrales des constituants de l’atmosphère et de son indice de réfraction. C’est principalement ces coefficients qui permettent une distinction entre la contribution de la troposphère et celle de la stratosphère, car les conditions différentes de température et pression dans ces deux milieux induisent une différence de comportement spectral des constituants. On constate donc une dépendance fortement non linéaire envers la composition chimique de l’atmosphère.
- Le terme intégral de cette équation est le plus problématique, car il implique une résolution en fonction de la direction de propagation du rayonnement, donc implicitement suppose une connaissance 3D de la composition chimique. Si on peut négliger ce terme pour les grandes longueurs d’onde, où l’absorption est prédominante, il n’en est pas de même dans les domaines du visible et de l’ultra violet, qui sont utilisés pour la mesure chimique de l’atmosphère. On utilise classiquement des méthodes dites à n flux créant une discrétisation des directions de propagation, ce qui abouti à un système couplé de n équations différentielles.

On constate ainsi que l’information apportée par un modèle de dispersion atmosphérique incluant les aérosols permet de modéliser la radiance exo-atmosphérique pour peu que les propriétés spectrales des composants soient connues, et que le modèle de dispersion fournisse la connaissance de la composition de l’atmosphère pour toutes les altitudes.

Notre proposition vise à étudier les problèmes théoriques de l’assimilation de radiance. Comme le modèle POLAIR3D utilisé pour nos études est limité à la basse troposphère, il ne sera pas possible dans cette proposition d’assimiler pratiquement les radiances mesurées par satellite, car il nous faudrait connaître la composition de la stratosphère. Néanmoins, cela n’interdit pas d’utiliser des mesures radiatives obtenues par capteurs aéroportés, ou de mener des expériences numériques jumelles.

Nous nous proposons d’étudier les points suivants :

- **Assimilation de radiance dans un modèle de dispersion** : dans ce cas,

l'équation du transfert radiatif (2) est considérée comme un opérateur d'observation, qui fournit, en fonction de la connaissance de l'atmosphère apportée par le modèle, l'équivalent modèle de la radiance. La difficulté est ici liée au caractère non linéaire de l'opérateur d'observation, ce qui a un impact direct sur les méthodes d'assimilation. Nous proposons d'étudier ce point dans le cadre de l'assimilation variationnelle, ce qui nécessitera le développement de l'adjoint de l'opérateur d'observation.

- **Choix d'un modèle radiatif optimal** : l'assimilation de données est un problème d'optimalité portant sur l'ensemble constitué du modèle, de ses données d'entrées, des observations. A observation fixée, le problème du choix optimal d'un modèle est encore un problème largement ouvert, des expériences préliminaires prouvant qu'un compromis est à trouver entre un modèle trop simple, représentant mal la réalité physique, et un modèle trop complexe, pour lequel nous ne disposons pas d'assez de données. Ce problème s'avère crucial dans le cas du choix du modèle radiatif, car dans ce cas, ses entrées sont fixées (par le modèle de dispersion), et les observations de radiances sont celles fournies par les satellites. Il est donc important de se poser la question de la discrétisation optimale (en terme de nombre de flux notamment) du modèle radiatif.
- **Expériences numériques** : la faisabilité de l'assimilation de radiance sera testée sur des expériences numériques jumelles, où l'on utilisera des radiances sorties de modèles puis perturbées comme simulation d'observations réelles. Nous nous efforcerons enfin de prendre contact avec des équipes d'optique atmosphérique (comme le LOA) pour disposer de procédure d'inversion de modèles radiatifs, afin de comparer les résultats de l'assimilation de variance et de d'assimilation de mesures chimiques obtenues par inversion de mesures radiatives. Notons que nous avons d'ores et déjà engagé des travaux dans cette dernière direction, notamment par le biais d'une soumission à l'appel d'offre de l'ESA et d'EUMETSAT sur l'exploitation des futures données du satellite MetOp. Dans ce cadre, nous projetons d'assimiler les profils d'ozone troposphérique fournis par ce capteur.

4 Le consortium

La présente proposition pour une action de recherche coopérative regroupe les équipes suivantes :

1. Equipe CLIME, Inria-Rocquencourt/ENPC CLIME effectue des travaux dans trois directions principales, les deux premières à vocation recherche, la troisième visant à faciliter le transfert vers les applications :
 - Traitement de la donnée et de l'image environnementale,
 - Assimilation de données et modélisation inverse,
 - Chaînes logicielles pour les applications environnementales.CLIME, a une expérience forte sur la simulation numérique de la qualité de l'air et la problématique d'assimilation de données associée. Les participants à cette proposition sont :
 - M. Bocquet : Chercheur de l'ENPC.
 - J.-P. Berroir : Chargé de Recherche, INRIA.
 - I. Herlin : Directeur de Recherche, INRIA.

– B. Sportisse : Chercheur de l'ENPC.

2. Equipe ASPI

L'équipe Aspi, Inria-Rennes/Irisa, a pour principal domaine d'investigation les méthodes probabilistes utilisant des particules aléatoires en interaction. L'activité de l'équipe se fait autant dans le domaine théorique (résultats de convergence, développement de nouvelles méthodes statistiques), que dans le domaine des applications (positionnement de terminaux mobiles, événements rares pour le trafic aérien). Les membres d'Aspi ont une grande expérience des méthodes probabilistes pour les modèles non-linéaires (notamment les méthodes SIR et apparentés). Les participants à cette proposition sont :

- F. Le Gland, Directeur de Recherche, INRIA.
- F. Campillo, Chargé de Recherche, INRIA.
- F. Céro, Chargé de Recherche, INRIA.

3. Projet IDOPT

Ce projet a été créé à Grenoble en 1993, c'est un projet commun à l'INRIA, au CNRS, à l'Université Joseph Fourier et à l'Institut National Polytechnique de Grenoble.

Le plus souvent on a d'un phénomène physique une information hétérogène tant en nature, qu'en qualité.

Information mathématique : fournie par les équations régissant les phénomènes. Ces équations contiennent, le plus souvent des quantités non directement accessibles à la mesure, : paramètres conditions initiales ;

Information physique : mesures, observations.

Une partie des travaux d'IDOPT est consacrée à la reconstruction de l'état d'un système à partir de sources d'information hétérogènes : problèmes d'assimilation de données en environnement, problèmes inverses en physique.

Ces problèmes sont étudiés à la fois dans un cadre déterministe et dans un cadre stochastique.

La collaboration très fructueuse avec les océanographes a aussi permis de développer une composante modélisation notamment consacrée au raffinement de maillage.

4. INERIS

L'INERIS (Institut National de l'Environnement Industriel et des Risques) assure une mission d'appui scientifique et technique au Ministère de l'Ecologie et du Développement Durable (MEDD), relative à l'évaluation de l'impact sanitaire et environnemental des substances chimiques rejetées dans l'environnement par les activités anthropiques. A ce titre, dans le domaine de la surveillance et de la gestion de la qualité de l'air, l'Institut participe au développement et met en oeuvre de manière opérationnelle le système national de prévision et de cartographie de la qualité de l'air, PREV'AIR (www.prevair.org). L'assimilation de données d'observation issues de mesures in-situ ou satellitaires, constitue désormais une voie d'amélioration des résultats des modèles de simulation de la chimie et du transport des polluants atmosphériques. L'INERIS souhaite faire évoluer les modèles de qualité de l'air utilisés en interne, et dans le système PREV'AIR en ce sens.

Sa contribution au projet proposé serait donc essentiellement applicative, et déboucherait logiquement sur l'intégration des méthodes d'assimilation de données développées, dans les modèles utilisés par l'INERIS de manière opérationnelle pour des études dédiées à la gestion de la qualité de l'air, et dans le cadre de PREV'AIR.

Les participants à cette proposition sont :

- L. Rouil : ingénieur Etudes et Recherches, INERIS.
- Recrutement à venir.

5 Budget indicatif

La responsabilité administrative de l'ARC sera assurée par I. Herlin.

Nous demandons sur cette proposition le financement de 3 post-docs et de 2 stages qui se répartiront entre les équipes selon les besoins scientifiques. Des moyens complémentaires nécessaires pour mener à bien l'ensemble des travaux seront recherchés par les partenaires dans le cadre de collaborations extérieures.

Les équipes se proposent également d'organiser un workshop en 2006 (10Keuros).

Par ailleurs, un budget missions est prévu, pour l'organisation des réunions et le déplacement entre les trois sites : Grenoble, Paris et Rennes, ainsi que pour tous les déplacements professionnels (conférences, séminaires...) des personnels recrutés.

Le budget par année civile est donc le suivant :

- **Année civile 2005**

Missions : 7keuros

Personnel post-doc + stage : 63Keuros

Total 2005 : 70keuros

- **Année civile 2006**

Missions : 7keuros

Personnel post-doc + stage : 63Keuros

Workshop : 10keuros

Total 2006 : 80keuros

Budget Global demandé : 150keuros

Annexe : les budgets 2004 des équipes INRIA impliquées :

- CLIME : 100 keuros
- ASPI : 25 keuros
- IDOPT : 50 keuros

6 Collaborations

Des collaborations académiques sont envisageables, par exemple avec Hans Wackernagel (ENSMP), Olivier Boucher (LOA) et Cathy Clerbaux (IPSL). L'articulation avec le Programme National de Chimie Atmosphérique du CNRS (CNRS, projet ADO-MOCA : Assimilation de Données pour la Modélisation de la Chimie Atmosphérique) sera recherchée.

Références

- [1] E. Andersson, J. Pailleux, J.N. Thépaut, J.R. Eyre, A.P. McNally, and P. Courtier. Use of cloud-cleared radiances in three/four variational data assimilation. *Q.J.R. Meteor. Soc.*, 120 :627–653, 1994.
- [2] Bertino, L., Evensen, G. and Wackernagel, H. (2003) Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71, p. 223-241.
- [3] O. Boucher and D. Tanré. Estimation of the aerosol perturbation to the earth's radiative budget over oceans using polder satellite aerosols retrieval. *Geophys. Res. Lett.*, 27 :1103–1106, 2000.
- [4] O.P. Hasekamp and J. Landgraf. Ozone profile retrieval from backscattered ultra-violet radiances : the inverse problem solved by regularisation. *J. Geophys. Res.*, 106 :8077–8088, 2001.
- [5] R. Munro, R. Siddans, W.J. Reburn, and B. Kerridge. Direct measurements of tropospheric ozone from space. *Nature*, 392 :168–171, 1998.
- [6] Pham, D. T. (2000) Stochastics methods for sequential data assimilation in strongly nonlinear systems *Mon. Weather Rev.* 129(5), p. 1194-1207.
- [7] Pham, D. T., Verron, J., Roubaud, M. (1998) A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Marine Syst.* 16(3-4), p. 323-340.
- [8] Doucet, A., de Freitas, N. Gordon, N. (2001) *Sequential Monte Carlo methods in practice*. New York, Springer.
- [9] Van Leeuwen, P. J. (2002) A variance-minimizing filter for large-scale applications. Submitted to *Mon. Weather Rev.*