

## ACHILLE Newsletter

Lettre trimestrielle à l'attention  
de la communauté ADOMOCA

Numéro 8, Février 2012

### Edito

La deuxième réunion annuelle de la communauté ADOMOCA-2 s'est déroulée les 22 et 23 novembre 2011 aux Lecques, organisée par S. Bekki du LATMOS. Elle a réuni une quinzaine de participants provenant de 6 laboratoires impliqués dans l'assimilation de constituants chimiques atmosphériques. De nouveaux développements d'outils (VALENTINA V6, inversion de sources, kriegeage, haute résolution horizontale) ont été présentés. De nouvelles thématiques ont montré des résultats prometteurs : pollution/qualité de l'air, aérosols, bilan  $H_2O$ , CO et  $O_3$  dans l'UTLS. La troisième réunion du groupe est prévue à la fin de l'été 2012 pour finaliser ADOMOCA-2 et se projeter vers un éventuel nouveau programme ADOMOCA-3.

Philippe Ricaud

## Modélisation inverse

**Estimation de champs de paramètres en dispersion atmosphérique : application à l'accident de Tchernobyl à l'aide d'un 4D-Var**

M. Bocquet<sup>1,2</sup>

La modélisation inverse est de plus en plus utilisée en chimie et dispersion atmosphérique pour identifier les sources de polluants ou les champs d'émissions. D'autres paramètres qui caractérisent les processus du modèle peuvent également être estimés par modélisation inverse même si ce thème est très peu exploré. Une des grandes questions des problèmes inverses en général est de savoir si l'estimation des paramètres inconnus produit des valeurs physiques, ou bien corrige également d'autres biais liés à d'autres processus.

<sup>1</sup> Université Paris-Est, CEREAs, laboratoire commun École des Ponts ParisTech et EDF R&D, France.

<sup>2</sup> INRIA, centre de recherche de Paris Rocquencourt, France.

Nous avons exploré cette question sur le cas de l'accident de Tchernobyl, au cours duquel des concentrations d'activité radiologique de radionucléides (césium 137, césium 134 et iode 131 en particulier) ont été mesurées. Nous avons utilisé un 4D-Var pour estimer non seulement le terme source de ces trois espèces, mais également les paramètres liés à la paramétrisation du dépôt sec et humide, de la diffusion horizontale et verticale (Kh et Kz). L'utilisation du 4D-Var est absolument nécessaire dans ce contexte compte tenu du nombre d'optimisations testées, et de la dépendance non-linéaire des observations en les paramètres. Les paramètres estimés ne sont pas seulement des valeurs uniformes, mais également des champs complets de paramètres.

Un des résultats de cette étude est la modélisation inverse simultanée du terme source, des paramètres des processus de dépôts, et du Kz, de sorte que tous les processus sont débiaisés et n'impactent plus l'estimation du terme source. Par ailleurs, une régularisation du terme source doit être simultanément utilisée afin d'éviter la semi-convergence de l'estimation de la source, c'est-à-dire afin d'éviter que les variables du terme source ne sur-explicitent les observations (overtuning). Sur la Figure 1 sont représentés 14 facteurs d'échelle multipliant les Kz (un par niveau vertical) et ce pour le jour et la nuit (soit 28 facteurs au total).

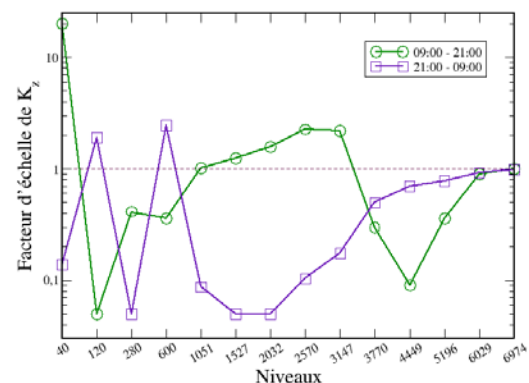
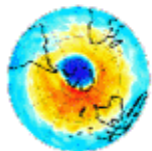


Figure 1: Facteurs d'échelles par niveau qui multiplient le champ  $K_z$  de référence (paramétrisation de Louis). 14 facteurs sont calculés pour le jour, et 14 pour la nuit. Comme on pouvait s'y attendre compte tenu des déficiences actuelles des modèles de dispersion, les valeurs optimales des  $K_z$  sont plus faibles que les valeurs de référence la nuit et légèrement plus fortes le jour.

Un second résultat est la reconstruction d'un champ 2D de vitesse de dépôt sec. La reconstruction n'est pas régularisée. L'inversion n'est donc a priori pas physique, c'est-à-dire que les paramètres retrouvés ne correspondent pas nécessairement à des vitesses de dépôt réalistes. Toutefois, ces



paramètres reconstruits permettent de bien expliquer les observations. Si cela ne présente donc pas d'intérêt pour déterminer des valeurs physiques des valeurs de dépôt, cela peut permettre de paramétrer d'autres erreurs modèle et biais présents dans la simulation. Nous avons calculé ce champ 2D optimal pour la dispersion du césium 137. Les indicateurs statistiques caractérisant la qualité de la simulation qui en résulte sont très au-dessus des scores moyens en dispersion accidentelle : par exemple 80% pour la corrélation de Pearson à comparer à 60% pour l'état de l'art d'une simulation Tchernobyl. Ceci était espéré. Toutefois l'utilisation de ce champ 2D dans la simulation du césium 134 (validation croisée) mène étonnement à des scores tout aussi satisfaisants. Il est donc fort probable que cette optimisation contribue à corriger significativement l'erreur modèle dans le système, en particulier l'erreur de représentativité. Le champ de vitesse optimal est représenté sur la Figure 2.

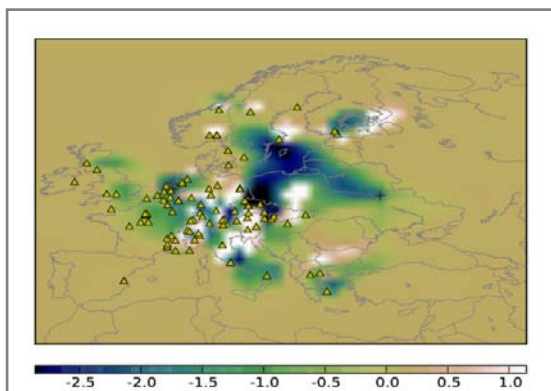


Figure 2: Facteur d'échelle du champ de vitesse de dépôt 2D optimal, en échelle logarithmique. La valeur 0 correspond à une vitesse de référence pour le césium 137, soit  $v_d=0.2 \text{ cm.s}^{-1}$ .

Compte tenu des scores obtenus, nous émettons l'hypothèse que ce type de modélisation inverse peut permettre dans un contexte dispersion ou chimie atmosphérique d'obtenir des scores de prévision très élevés grâce à une correction significative de l'erreur modèle. L'investissement technique (adjoint par rapport aux paramètres) serait cependant important.

Référence :

Bocquet, M. Parameter field estimation for atmospheric dispersion: Application to the Chernobyl accident using 4D-Var, M. Bocquet., *Q. J. R. Meteorol. Soc.*, sous presse, 2012.

## Estimation des erreurs et modélisation inverse de rejets accidentels de polluants atmosphériques : application au cas de l'accident de Fukushima Daiichi

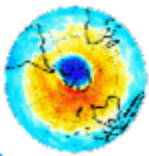
V. Winiarek<sup>1,2</sup>, M. Bocquet<sup>1,2</sup>, O. Saunier<sup>3</sup>, A. Mathieu<sup>3</sup>

Lorsqu'un accident industriel rejetant dans l'atmosphère des polluants (gaz toxiques, matières radioactives, ...) survient, il est nécessaire, pour la protection des populations, d'anticiper les zones qui vont être impactées par le rejet. Pour ce faire, les modèles de transport atmosphérique sont largement utilisés. Mais ces modèles sont naturellement très sensibles au terme source utilisé. Si le lieu précis de l'accident est connu, il s'agit de préciser un profil temporel et éventuellement une distribution verticale du rejet (cas de Tchernobyl et de Fukushima); mais le lieu de l'accident peut aussi être inconnu (cas de l'incident d'Algésiras). Les mesures sur site ou la modélisation directe sont rarement possibles pour déterminer le terme source. Les techniques de modélisation inverse ont montré leur efficacité pour résoudre ce type de problèmes.

Nous avons utilisé ce type de techniques pour retrouver le profil temporel du terme source de césium 137 et d'iode 131 consécutif à l'accident sur la centrale nucléaire de Fukushima Daiichi, lui-même conséquence du tremblement de terre et du tsunami du 11 mars 2011. Nous avons pour cela utilisé des mesures de concentration d'activité réalisées au Japon, mais aussi aux Etats-Unis et au Canada ainsi que sur certaines îles du Pacifique. Le profil temporel des termes sources recherchés est discrétisé en 384 pas de temps d'une heure; le nombre de mesures disponibles est de 267 pour le césium 137 et 428 pour l'iode 131.

Dans le cadre d'une modélisation Gaussienne des erreurs, la méthode repose essentiellement sur la minimisation d'une fonction de coût de type moindre carré. Néanmoins, le faible nombre de mesures disponibles rend nécessaire l'utilisation d'un terme de régularisation. Le terme d'ébauche étant choisi nul, deux paramètres *a priori* sont alors nécessaires : la variance *a priori* des erreurs d'observation et la variance *a priori* des erreurs d'ébauche. Le résultat de la modélisation inverse est très sensible à ces deux paramètres et il convient donc d'estimer leurs valeurs optimales. Pour

<sup>3</sup> Institut de Radioprotection et de Sécurité Nucléaire, Fontenay-aux-Roses, France.



cela, nous avons comparé la méthode de vraisemblance maximale, capable de prendre en compte sans approximation la positivité de la source, et une méthode basée sur la L-curve, couplée à un diagnostic du  $\chi^2$ .

Les résultats varient d'une méthode à l'autre, aussi bien sur les paramètres que sur la masse totale retrouvée. Néanmoins, pour cette dernière, les ordres de grandeur sont cohérents. Pour le césium 137, une limite inférieure de la masse totale rejetée est estimée à  $1,2 \cdot 10^{16}$  Bq; pour l'iode 131 une limite inférieure de la masse rejetée est estimée entre  $1,9 \cdot 10^{17}$  Bq et  $3,8 \cdot 10^{17}$  Bq. Ces résultats montrent des quantités émises entre 5 fois et 10 fois plus faibles que lors de l'accident de Tchernobyl.

Un deuxième résultat de l'étude est l'estimation formelle de l'incertitude *a posteriori* des termes sources reconstruits. Les résultats montrent que les conditions météorologiques particulières, qui ont poussé le panache radioactif le plus souvent directement vers l'océan Pacifique, ont pour conséquence des fenêtres temporelles peu observées, pendant lesquelles l'incertitude sur l'estimation du terme source reste élevée (à hauteur de l'incertitude *a priori*). Les quantités de matières radioactives effectivement rejetées sont donc certainement légèrement plus élevées.

La méthode peut encore être améliorée par exemple par l'optimisation des paramétrisations physiques du modèle de transport atmosphérique utilisé (vitesses de dépôt sec par exemple). Il serait également certainement intéressant d'implémenter des méthodes reposant sur des statistiques non-Gaussiennes pour les erreurs d'observations.

Enfin il serait particulièrement bénéfique d'implémenter des méthodes capables de prendre en compte différents types d'observation (mesures de dépôts, de doses, ...) afin d'augmenter considérablement le nombre de données disponibles, la contrepartie serait

certainement l'augmentation des incertitudes *a priori* introduites dans le système.

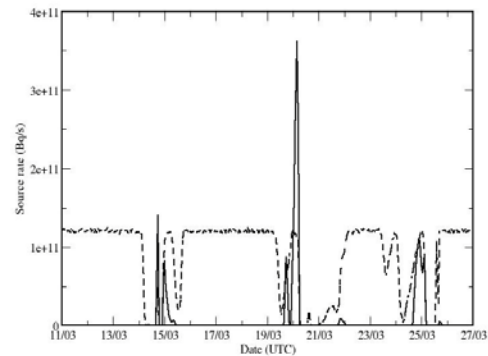


Figure 3: Profil temporel du terme source de césium 137 reconstruit avec la méthode vraisemblance maximale en contrainte de positivité de la source et en utilisant les observations japonaises uniquement. En pointillés : l'incertitude *a posteriori* du terme source.

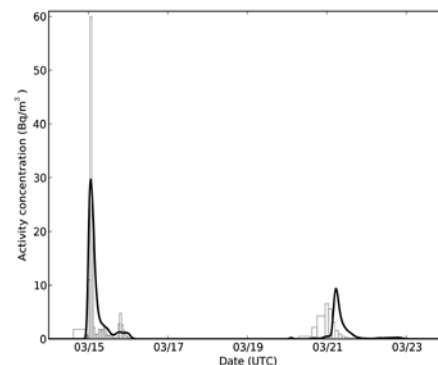


Figure 4: Cartes de concentrations d'activité de césium 137 simulées (en Bq/m<sup>3</sup>) en utilisant le terme source reconstruit avec la méthode de vraisemblance maximale et toutes les observations disponibles. Modèle de transport atmosphérique : POLAIR3D, modèle eulérien de la plateforme POLYPHEMUS (CEREA, École des Ponts ParisTech et EdF R&D).

Référence:

Winiarek, V., Bocquet, M., Saunier, O., Mathieu, A. 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.*, sous presse. 2012.

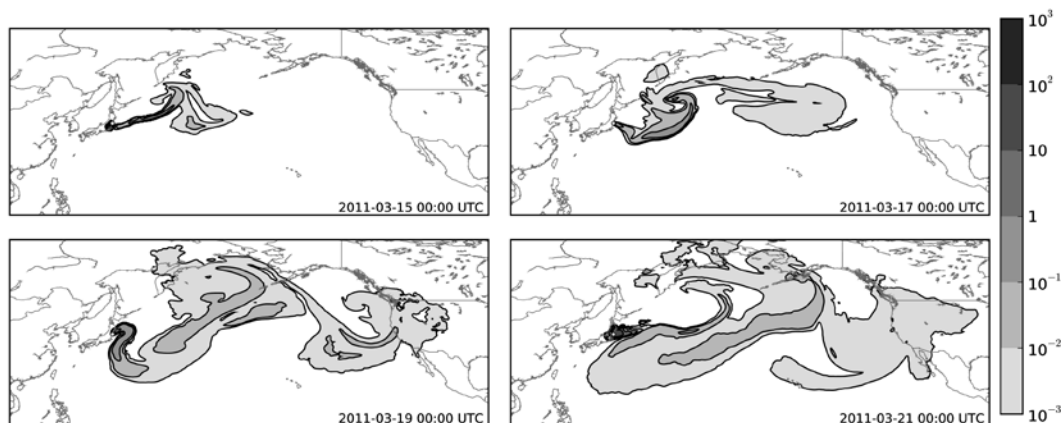
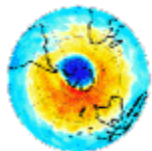


Figure 5: Comparaison entre les mesures réalisées à Tokyo (histogrammes dont la largeur représente la durée de la mesure) et les concentrations simulées de césium 137 en utilisant le terme source reconstruit avec la méthode de vraisemblance maximale et les observations japonaises.





## Assimilation continentale

### Assimilation de colonnes partielles d'ozone (0-6 km) IASI dans un modèle de chimie-transport à l'échelle continentale

A. Coman<sup>4</sup>, G. Forêt<sup>4</sup>, M. Beekmann<sup>4</sup>, M. Eremenko<sup>4</sup>, G. Dufour<sup>4</sup>, B. Gaubert<sup>4</sup>

Dans le domaine de la pollution, la synergie entre les modèles régionaux de chimie-transport et observations conduit à la mise en place de systèmes d'assimilation de plus en plus performants. Ces systèmes s'appuient sur des méthodes d'assimilation variationnelles ou d'ensemble avec pour but d'améliorer les performances des modèles autant à la surface que dans la troposphère libre. Au LISA, nous développons depuis 2006 un système d'assimilation des observations qui couple un Filtre de Kalman d'Ensemble (Evensen, 1994) au modèle CHIMERE. Le système a été développé initialement pour assimiler les observations d'ozone de surface.

L'apparition des sondeurs infrarouge de dernière génération comme IASI permet maintenant de mesurer précisément les concentrations d'ozone dans la troposphère libre avec une fréquence bi-journalière (Eremenko et al., 2008). Nous avons pu montrer qu'une partie de l'information apportée par IASI pouvait être transférée de la troposphère libre vers la surface (Foret et al., 2009).

Un exercice d'assimilation portant sur le mois de Juillet 2007 a été mis en place visant à assimiler les colonnes partielles (0-6 km) d'ozone IASI, produites par inversion, en utilisant la méthode de Tikhonov (Eremenko et al., 2008). Les résultats montrent une correction du biais positif du modèle CHIMERE (Figure 7, extraite de Coman et al., 2011). Cela correspond aussi à une réduction d'environ 25% sur l'erreur quadratique moyenne. Globalement, on observe que les corrections sont plus importantes dans la moitié nord du domaine où les conditions aux limites surestiment les concentrations d'ozone advectées depuis l'Atlantique Nord. Ces résultats sont cohérents avec les résultats obtenus à partir d'un ensemble de modèles

dans le projet FP7/GEMS (Zyryanov et al., 2011). Par ailleurs, la validation effectuée en utilisant les données MOZAIC disponibles (83 profils verticaux) montre que la correction a été plus efficace là où la sensibilité de l'instrument IASI est maximale, c'est-à-dire dans la partie comprise entre 2 et 5 km d'altitude (Figure 6, extraite de Coman et al., 2011). En revanche, pour les 5 sites de sondages utilisés dans l'exercice de validation, les résultats sont plus mitigés. Une moyenne effectuée sur les 46 profils disponibles (sur l'ensemble de 5 sites) ne montre pas d'amélioration dans la troposphère libre. Cependant, à Payerne (en Suisse), le site où on dispose de plus de mesures (14 profils), on observe une faible amélioration. Des investigations plus approfondies seront nécessaires pour arriver à expliquer ces résultats.

Quant aux données de surface, utilisées uniquement pour la validation, elles montrent quand même une diminution de l'erreur quadratique moyenne à la surface (pour plus de 90% de stations AIRBASE) allant jusqu'à 7ppb par comparaison avec une simulation de référence. Ceci est un signal fortement positif, car l'instrument IASI ne montre a priori pas une grande sensibilité à la surface.

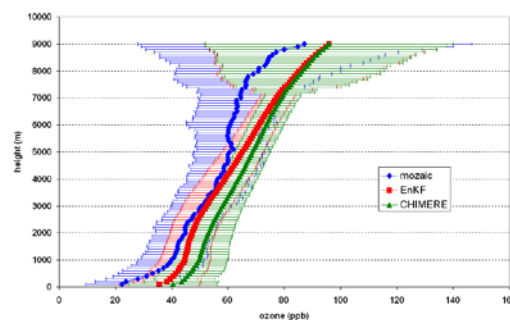


Figure 6: Comparaisons des profils verticaux moyens provenant de vols MOZAIC (83 profils au départ de Francfort) en bleu, le profil moyen calculé à partir d'une simulation de référence du modèle CHIMERE (en vert) et celui calculé sur l'ensemble analysé (en rouge). Les barres d'erreur correspondent à l'écart-type à la moyenne pour chaque type de profil.

Cet exercice initial sera complété en travaillant sur l'été 2009 et notamment en s'appuyant sur les cas de pollution intense détectés en Europe durant cette période. Cette étude sera menée dans le cadre d'ADOMOCA et conjointement avec le Laboratoire d'Aérodologie et le CNRM qui mettront en œuvre la chaîne d'assimilation MOCAGE-Valentina. Par la suite, une approche d'assimilation conjointe pour tenir compte de la complémentarité des observations sol/satellite sera envisageable, ce qui permettra au final d'avoir une vision plus complète du

<sup>4</sup> Laboratoire Inter-universitaire des Systèmes Atmosphériques, LISA/IPSL, Université Paris Est et 7, CNRS/INSU UMR 7583, Créteil, France.

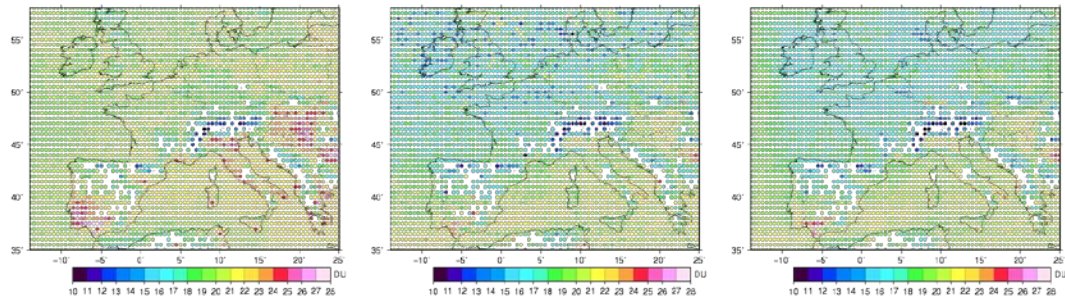
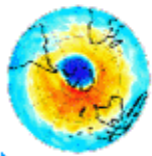


Figure 7: Comparaison des moyennes de colonnes partielles (0-6 km) d'ozone (DU) pour le mois de Juillet 2007 à 9 heures (passage matinal du satellite) pour a) a priori CHIMERE (moyenne calculée sur 40 membres d'ensemble) b) les observations IASI disponibles, c) analyse obtenue en utilisant un filtre de Kalman d'ensemble (utilisant une analyse locale).

potentiel de ces observations pour contraindre le modèle et notamment pour améliorer la prévision et la surveillance des pics d'ozone.

Références:

Coman, A., Foret, G., Beekmann, M., Eremenko, M., Dufour, G., Gaubert, B., Ung, A., Flaud, J.-M., Bergametti, G., Schmechtig, C.: Assimilation of IASI partial tropospheric columns with an Ensemble Kalman Filter over Europe, *Atmos. Chem. Phys. Discuss.*, 11, 26943-26997, 2011, doi:10.5194/acpd-11-26943-2011.

Eremenko, M., Dufour, G., Foret, G., Keim, C., Orphal, J., Beekmann, M., Bergametti, G., and Flaud, J.-M.: Tropospheric ozone distributions over Europe during the heat wave in July 2007 observed from infrared nadir spectra recorded by IASI, *Geophys. Res. Lett.*, 35, L18805, doi:10.1029/2008GL034803, 2008.

Evensen G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 143-162, 1994.

Foret, G., Hamaoui, L., Schmechtig, C., Eremenko, M., Keim, C., Dufour, G., Boynard, A., Coman, A., Ung, A. and Beekmann, M.: Evaluating the potential of IASI ozone observations to constrain simulated surface ozone concentrations, *Atmos. Chem. Phys.*, 9, 8479-8491, 2009.

Zyryanov, D., G. Foret, M. Eremenko, M. Beekmann, J.-P. Cammas, M. D'Isidoro, H. Elbern, J. Flemming, E. Friese, I. Kioutsioutkis, A. Maurizi, D. Melas, F. Meleux, L. Menut, P. Moinat, V.-H. Peuch, A. Poupkou, M. Razinger, M. Schultz, O. Stein, A. M. Suttie, A. Valdebenito, C. Zerefos, G. Dufour, G. Bergametti, and J.-M. Flaud, 3-D evaluation of tropospheric ozone simulations by an ensemble of regional, Chemistry Transport Model, *Atmos. Chem. Phys. Discuss.*, 11, 28797-28849, 2011.

## Choix d'une configuration optimale pour la matrice de covariances d'erreur d'ébauche dans le cadre de réanalyses annuelles de l'ozone de surface

E. Jaumouillé<sup>5</sup>, S. Massart<sup>5</sup>, A. Piacentini<sup>5</sup>

Dans le cadre du projet Européen MACC (Monitoring Atmospheric Composition and Climate), des réanalyses du champ d'ozone à

la surface ont été faites au Cerfacs et au CNRM pour les années 2008 et 2009. Elles sont basées sur la configuration régionale de Morage-Valentina qui a été utilisée pour assimiler des observations d'ozone à la surface sur l'Europe. Les observations proviennent de trois types de stations (urbaines, suburbaines et rurales) et des stations non assimilées ont été choisies pour être des stations de validation.

Nous avons montré lors d'une étude des champs d'ozone à la surface sur l'Europe (cf lettre Achille N°6, novembre 2010) que les paramètres de la matrice des covariances d'erreur d'ébauche  $\mathbf{B}$  (i.e. les écarts-types et les longueurs de portées) varient suivant les saisons et au cours d'une journée pour une configuration de Morage-Valentina similaire à celle utilisée pour les réanalyses. Pour les réanalyses 2009, nous avons fait une étude de sensibilité des analyses obtenues avec des configurations différentes de la matrice  $\mathbf{B}$ . Voici les trois configurations choisies :

- version 1 : la matrice  $\mathbf{B}$  est mensuelle, elle est issue des diagnostics *a posteriori* faits à partir des réanalyses 2008;
- version 2 : la matrice  $\mathbf{B}$  est journalière, elle est issue des diagnostics *a posteriori* faits à partir de la version 1;
- version 3 : la matrice  $\mathbf{B}$  est journalière et varie au cours de la journée, elle est issue des diagnostics *a posteriori* faits à partir de la version 1.

Pour chacune de ces trois configurations, seule la variation spatiale des écarts-types est prise en compte. Par exemple, on voit sur les cartes suivantes (Figure 8) que la variation spatiale des écarts-types et leur intensité sont différentes dans les trois versions de la matrice  $\mathbf{B}$ . Les longueurs de portée sont quant à elles constantes sur toute la grille quelque soit la version.

<sup>5</sup> CERFACS/SUC CNRS-URA 1875, Toulouse, France

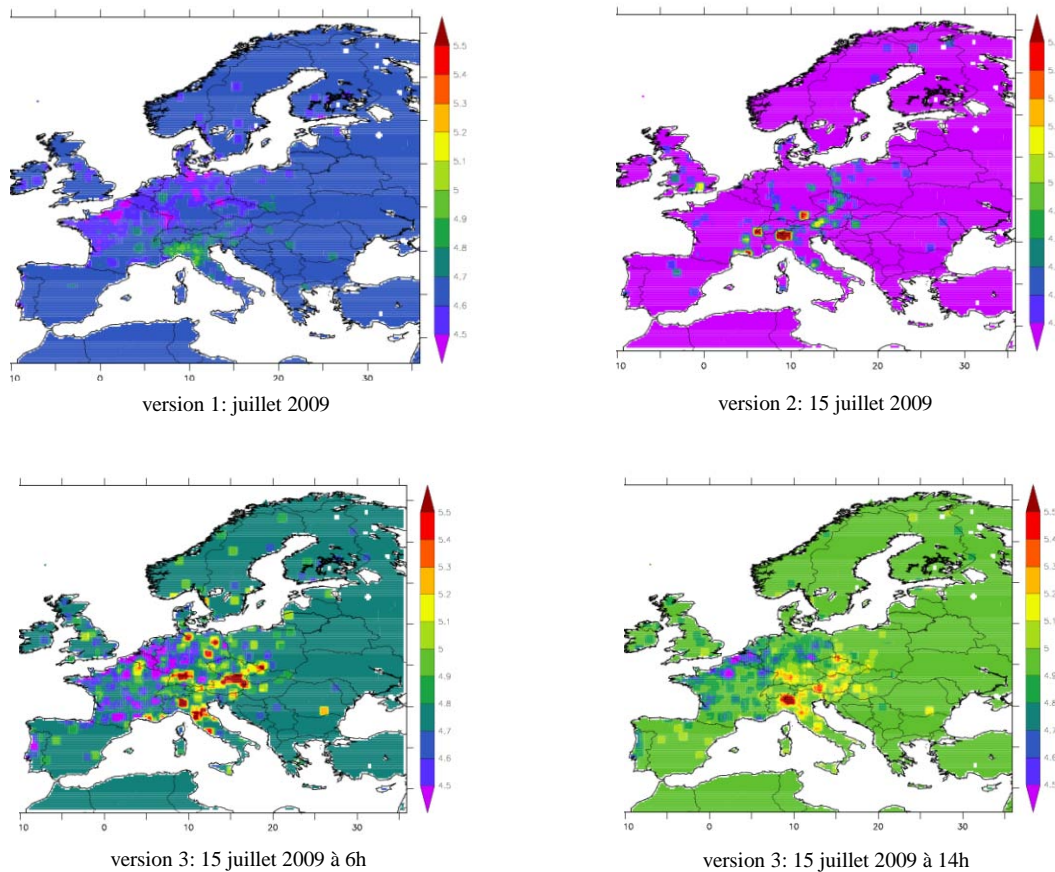
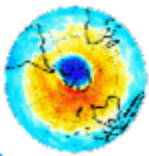


Figure 8: Cartes des écarts-types utilisés dans la matrice **B** pour les différentes versions des réanalyses (en ppbv).

Comme première validation, on regarde la valeur du  $\chi^2$  qui représente le rapport entre la valeur de la fonction coût et le nombre d'observations divisé par deux. Lorsque la valeur du  $\chi^2$  est 1 cela signifie que la matrice **B** prise pour l'assimilation est bien adaptée au problème étudié (sous réserve que la matrice des covariances d'erreur d'observation est elle aussi bien spécifiée). Les résultats montrent que le  $\chi^2$  est inférieur à un pour la version 1 (Figure 9). Cela montre que les matrices **B** de 2008 donnent des résultats corrects mais qu'elles ne sont pas bien adaptées aux réanalyses de 2009. Pour les deux autres versions, les  $\chi^2$  obtenus sont proches de 1. Il semble donc important de prendre des valeurs correspondantes à la période assimilée et non celles de l'année d'avant. De plus, la variabilité du  $\chi^2$  est réduite avec les versions 2 et 3, surtout pendant les mois d'été.

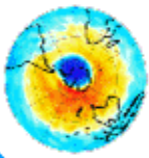
Les diagrammes de Taylor permettent de visualiser simultanément les statistiques des trois versions des réanalyses 2009 (Figure 10). Pour les stations assimilées, les corrélations avec les observations sont de l'ordre de 0.9 et

les meilleures corrélations sont obtenues avec la version 3. La variabilité des réanalyses est du même ordre que celle des observations, car les points des réanalyses et des observations sont sur le même arc de cercle dans le diagramme de Taylor.

On a vu que le  $\chi^2$  de la version 1 est inférieur à 1, ce qui impliquerait une matrice **B** trop grande. L'analyse devrait alors être plus près des observations assimilées mais ce n'est pas ce que l'on observe sur les diagrammes de Taylor. On peut en déduire qu'utiliser une matrice mensuelle n'est pas la bonne approche dans l'étude de l'ozone de surface. Cela peut aussi vouloir dire que la variabilité spatiale des erreurs d'ébauche n'est pas la même pour les années 2008 et 2009, mais cela est peu probable.

Dans le cas des stations de validation, les corrélations avec les observations sont de l'ordre de 0.8. La variabilité des réanalyses est plus proche de celle des observations dans les versions 2 et 3. Cependant comme les trois analyses ont des statistiques presque identiques, il en ressort que l'information apportée par les stations assimilées n'est pas suffisamment bien répartie sur le reste du





domaine. Nous pouvons en déduire que les longueurs de portées choisies ne sont pas adaptées à notre problème. Il serait préférable de prendre des longueurs de portée variables.

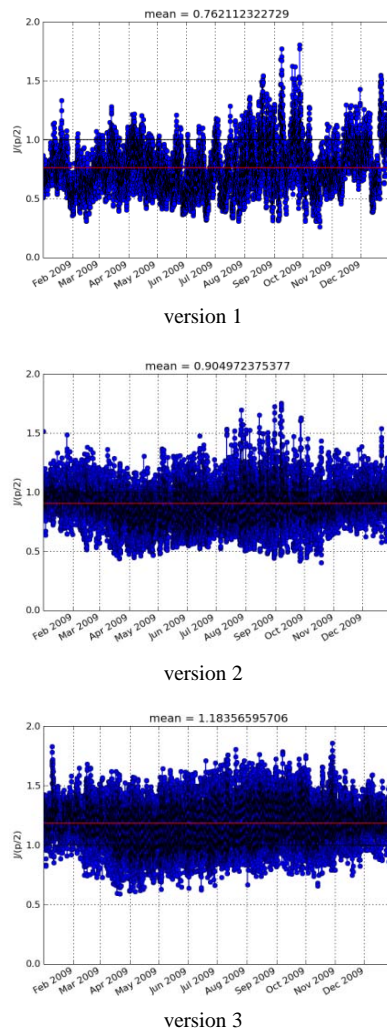


Figure 9:  $\chi^2$  pour les différentes configurations de la matrice  $B$  utilisées pour les réanalyses sur l'année 2009.

Ces résultats sont identiques pour les 3 types de stations, avec des corrélations plus faibles pour les stations rurales.

Ainsi, l'étude des analyses provenant de réanalyses générées avec des matrices de covariance d'erreur d'ébauche différentes permet de conclure qu'une matrice  $B$  journalière est mieux adaptée à l'analyse de l'ozone de surface. Cependant un travail sur les longueurs de portées est nécessaire.

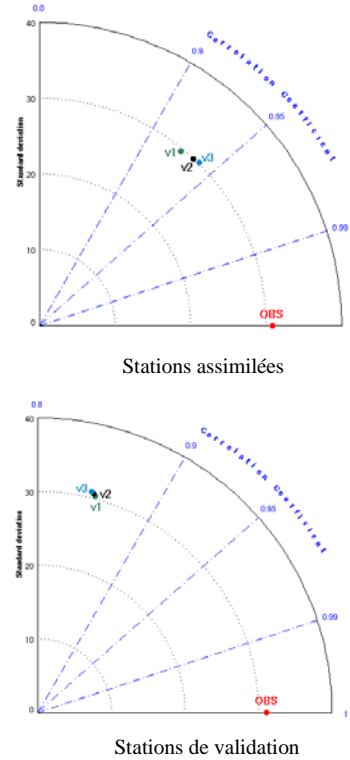


Figure 10: Diagramme de Taylor pour les stations urbaines; statistiques pour les stations assimilées (en haut) et pour les stations de validation (en bas) pour les réanalyses de l'année 2009.

#### Références:

E. Jaumouillé, S. Massart, A. Piacentini, V.-H. Peuch, D. Cariolle, Impact of a time-dependent background error covariance matrix on air quality analysis, GMD, 2012, à soumettre

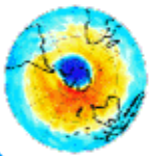
G. Desroziers, L. Berre, B. Chapnik, P. Poli, Diagnosis of observation, background and analysis-error statistics in observation space, Quart. J. Roy. Meteor. Soc., 131, 3385-3396, 2007

## Aspect logiciel

### Le futur de la chaîne d'assimilation VALENTINA

A. Piacentini<sup>5</sup> et les développeurs de VALENTINA

La release de la version 6 de la chaîne d'assimilation VALENTINA a été annoncée lors de l'atelier ADOMOCA-2 du mois de novembre 2011. Ce qui caractérise cette sixième release est l'absence de nouveautés de taille par rapport à la version précédente. Les quelques apports de la V6 concernent surtout la gestion de la configuration régionale avec conditions au bord imposées utilisée pour le projet MACC, la manipulation de gros fichiers de données d'observation au format NetCDF, tels ceux qui contiennent les observations simulées pour POGEQA, et quelques



améliorations de l'ergonomie de l'interface graphique PrePALM.

Cette stabilité est, d'un côté, le signe que le code est assez mûr pour traiter tous les cas à l'étude dans les projets et les activités en cours, mais, sous un autre angle, est aussi le symptôme de la fin des possibilités d'évolution progressive de la chaîne actuelle.

Nous pouvons reconnaître les raisons principales qui ont déterminé les limites de développement de Valentina dans le cadre présent :

1. environnement d'assimilation multi-modèle, multi-observations, multi-configuration;
2. utilisation de l'opérateur d'observation et de son adjoint codés par NOVELTIS au début du projet;
3. choix de parallélisation des opérateurs d'assimilation conçue de façon compatible avec l'environnement de calcul vectoriel de Météo-France et peu adaptée au calcul massivement parallèle;
4. impasse technique pour le codage d'un modèle adjoint de MOCAGE efficace sur machine vectorielle.

Le point 1. a été dicté par le contexte de collaboration scientifique, de comparaison de modèles et de méthodes d'assimilation qui était au coeur du projet ADOMOCA original. Depuis, l'approche avec assimilation a fait ses preuves et on peut envisager de passer progressivement à des chaînes spécialisées et plus performantes. Dans un premier temps on continuera de maintenir une chaîne d'assimilation avec les caractéristiques de flexibilité que Valentina a aujourd'hui, mais on pourra commencer à en dériver une série de versions où les opérateurs génériques migreront progressivement de la partie assimilation directement dans le code du modèle. L'avantage de l'approche sous PALM est que ces opérateurs pourront continuer à être identifiés comme des unités fonctionnelles indépendantes, tout en faisant partie du code de calcul du modèle, ce qui permet de diminuer les transmissions de données et d'exploiter les efforts d'optimisation fournis pour le modèle direct.

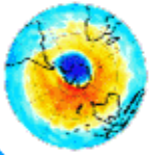
Cette considération se refait directement au point 2., car l'opérateur d'observation externalisé que l'on utilise actuellement dans Valentina comporte, nécessairement, une duplication d'information qui est pénalisante pour l'efficacité, autant en termes de performances que d'occupation mémoire. De plus, il n'a pas été conçu, au début du projet

ADOMOCA, pour les types et les volumes de données d'observation que l'on est et l'on sera censés traiter dans les activités de recherche à venir. La refonte de l'opérateur d'observation est un chantier majeur qui permettra d'assimiler plusieurs espèces à la fois, avec haute résolution spatiale et avec tout angle de visée. Afin que cette effort soit optimal, il est envisageable de spécialiser cet opérateur et son opérateur adjoint pour chaque modèle, de façon à pouvoir exploiter les fonctions optimisées de repérage spatial, d'interpolation et de cumul de quantités intégrées éventuellement présentes dans le modèle. Ce choix, toutefois, ne pourra être entériné qu'avec l'accord des équipes qui développent les modèles, qui devront prendre en charge le codage de l'opérateur d'observation et de son adjoint.

La même réflexion s'applique aux autres opérateurs qui impactent de façon déterminante les performances de la chaîne, dont, en particulier, les approximations de la matrice de covariance d'erreur d'ébauche et les modèles linéaire tangent et adjoint de propagation des incréments. À présent, les opérateurs qui approchent la matrice B sont calculés sur une grille indépendante. Ceci permet de changer facilement de modèle, mais impose des importants transferts de données et une grande occupation de mémoire qui, actuellement, rend rédhibitoire une approche d'assimilation multi-espèces ou multi-variée. De plus, pour obtenir de bonnes performances sur machine vectorielle, la parallélisation pour les opérateurs horizontaux est faite sur les niveaux, tandis que celle pour les opérateurs verticaux est faite sur les colonnes. Cette approche est optimale pour un calculateur vectoriel avec un très bon réseau de communication interne et relativement peu de processeurs, mais absolument non adaptée aux machines massivement parallèles. En vue du passage à la nouvelle plate-forme de calcul de Météo-France, un travail de réécriture partielle du code MOCAGE vient de débuter. L'équipe assimilation devrait s'organiser pour suivre ces développements et les transférer vers les parties du code d'assimilation orientées au calcul vectoriel.

Du même teneur est la dernière remarque, qui a freiné jusqu'à présent l'implémentation de méthodes 4D pour les schémas chimiques complets et pour les configurations régionales : l'optimisation du code de transport dans MOCAGE est absolument incompatible avec son adjointisation, quitte à faire tourner la modèle adjoint sur un seul processeur, ce qui n'est pas envisageable. Cette contrainte sera





très probablement levée avec la réécriture partielle de MOCAGE. Il faut prévoir donc un chantier pour l'implémentation, la validation et l'exploitation de méthodes d'assimilation 4D.

Dans ce contexte il est important de poursuivre la réflexion entamée lors de l'atelier de novembre afin de fixer des priorités scientifiques, de recenser les forces disponibles pour ce travail, de répartir les tâches et d'en organiser la coordination.

**Lien ADOMOCA sur ETHER :**

<http://ether.ipsl.jussieu.fr>

**Contact**

elodie.jaumouille@cerfacs.fr

**Prochaine lettre : Septembre 2012**