

Stage de fin d'études 2021

Assimilation de données et traitement des observations

Niveau M2

1 Description du stage

1.1 Titre du stage

Valorisation des observations par des méthodes d'écrémage novatrices facilitant la modélisation des corrélations d'erreurs d'observation.

1.2 Encadrants du stage

Encadrant principal : Oliver Guillet, GMAP/ALGO, oliver.guillet@meteo.fr

L'encadrement se fera en collaboration avec l'équipe GMAP/OBS pour le traitement des données des observations et le CERFACS pour les aspects algorithmiques de la représentation des corrélations d'erreurs d'observation.

1.3 Informations sur le stage

Depuis plusieurs mois, un ensemble de développements techniques menés par l'équipe GMAP/OBS a permis la mise au point d'une nouvelle méthode d'écrémage (de sélection) des observations. Dans ce stage, on se propose de valoriser cette nouvelle méthode en regard des développements actuels en assimilation de données.

Le candidat doit être à l'aise en informatique, en mathématiques numériques, en assimilation de données et en algorithmie. Il est également appréciable que le candidat possède un esprit de synthèse, car il devra faire le lien entre les aspects pratiques de l'assimilation de données et les domaines scientifiques associés.

Bien entendu, on encourage très fortement le candidat à prendre contact avec l'encadrant principal pour en savoir plus sur le déroulement et sur les compétences requises.

1.4 Domaines abordés pendant le stage (mots-clefs)

Erreurs d'observation, écrémage des observations, corrélations des erreurs d'observation, assimilation de données, algorithmes d'assimilation.

1.5 Intérêt scientifique, contexte et état de l'art

L'assimilation de données place les observations au coeur de la prévision météorologique. Le recours à des instruments de plus en plus nombreux et précis permet de représenter des échelles de plus en plus fines dans l'analyse, améliorant ainsi la qualité de la prévision. Ainsi, parvenir à assimiler un nombre croissant de données est un enjeu fondamental de l'assimilation de données.

Toutefois, cette volonté d'expansion se heurte au défi de la représentation des corrélations d'erreurs d'observation. En effet, certaines observations sont entâchées d'erreurs spatialement (et spectralement) dépendantes : plus les mesures sont rapprochées, plus elles sont corrélées. Or, ces corrélations doivent être prises en compte dans les équations de l'assimilation de données au moyen d'un modèle de corrélation adapté, sous peine de dégrader la qualité de l'analyse.

Des études récentes ont mis en évidence d'importantes corrélations d'erreurs d'observation dans les jeux de données télédéteectées telles que les AMV (Bormann, 2002), les réflectivités et les vents radiaux des radars (Waller et al., 2016a) ou les radiances des sondeurs infrarouge hyperspectraux (Bormann et al., 2010). Ces études mettent ainsi l'accent sur l'importance de prendre en compte les corrélations d'erreurs d'observation pour pouvoir améliorer l'utilisation des observations dans les futurs systèmes d'assimilation de données (voir Liu et Rabier, 2003).

En pratique, ces corrélations sont négligées et les erreurs d'observation sont supposées indépendantes. Pour que cette hypothèse soit vérifiée, seule une fraction des observations disponibles est assimilée. Pour ce faire, on a recours à une procédure d'écrémage (*thinning* en anglais) qui consiste à ne retenir qu'un nombre limité d'observations par domaine géographique ou par région spectrale.

Récemment, des travaux menés au GMAP ont permis d'améliorer la procédure d'écrémage en sélectionnant les observations plus régulièrement, évitant ainsi de sélectionner des observations trop proches les unes des autres. Des premiers résultats ont montré une consistance de la qualité de l'analyse et une amélioration des scores sur l'Europe.

D'autre part, une distribution plus régulière des observations offre un cadre favorable à la représentation des corrélations d'erreurs d'observation. En effet, pour modéliser ces corrélations, on passe par l'évaluation d'opérateurs différentiels sur des maillages non structurés, lesquels s'appuient directement sur la distribution des observations (Guillet et al., 2019). Ainsi, l'amélioration de la procédure d'écrémage présente deux intérêts. Lorsqu'on suppose que les erreurs d'observation ne sont pas corrélées, une distance minimale entre les données est respectée. Lorsqu'on suppose que les erreurs d'observation sont corrélées, une distribution assez uniforme des données permet de générer un meilleur maillage pour discrétiser les opérateurs différentiels inhérents au modèle de corrélation.

1.6 Objectif général et questions de recherche traitées

L'objectif de ce stage est de valoriser le travail effectué sur l'amélioration de l'écrémage en quantifiant son impact sur les algorithmes d'assimilation. On s'intéresse en particulier à l'impact de l'écrémage sur les propriétés de la matrice de corrélation d'erreurs d'observations \mathbf{R} . On souhaite notamment étudier son nombre de conditionnement, qui influe sur la convergence des algorithmes de minimisation. La relation entre le conditionnement de \mathbf{R} et la convergence du 3DVar / 4DVar fait actuellement l'objet d'une collaboration avec le Cerfacs.

Lorsque la matrice \mathbf{R} est non diagonale, on adopte le modèle basé sur l'équation de diffusion pour représenter les corrélations horizontales (Guillet et al., 2019). Dans ce cas, le contenu de \mathbf{R} est déduit de la discrétisation d'un opérateur différentiel sur le maillage des observations. La qualité du maillage et de la distribution des observations influence donc directement les propriétés mathématiques de l'opérateur de corrélation. On se propose donc dans ce cas de mettre en relation la nouvelle méthode d'écrémage avec les propriétés du maillage et de l'équation de diffusion.

Pour cette étude, on pourra considérer des données de radars sur le domaine Arôme-France, de diffusiomètres (radars embarqués sur satellites pour la mesure du vent à la surface des océans) ou d'AMV dans Arpège.

Enfin, en considérant l'écrémage comme une procédure de sélection des données dans un cadre de génération de maillage, on pourrait le substituer naturellement par une méthode de raffinement de maillage. Les méthodes de raffinement de maillage étant bien connues dans le cadre de la méthode des éléments finis, il peut être intéressant de considérer leur apport dans le prétraitement des données météorologiques.

1.7 Plan de recherche, méthodologie et calendrier de réalisation

L'étude commencera par l'étude de deux schémas d'écrémage présents dans le code d'Arpege/IFS. Le candidat sera amené à entrer en contact avec l'équipe OBS pour comprendre chaque schéma et les traitements spécifiques aux types d'observation considérés. On pourra s'intéresser dans un premier temps aux vents mesurés par diffusiomètres (réfèrent : Christophe Payan), puis aux données radar (réfèrent : Maud Martet) ou encore aux données SEVIRI du satellite géostationnaire MSG (réfèrent : Olivier Audouin). L'identification de situations météorologiques critiques mettant en défaut ces méthodes d'écrémage permettra de mieux cibler les critères objectifs permettant de les valider.

A partir d'un type de données, on examinera les propriétés mathématiques de la matrice \mathbf{R} dans le cas diagonal et dans le cas corrélé. On s'intéressera notamment à son nombre de conditionnement et son rôle théorique dans la convergence de la minimisation. Pour ce faire, le candidat pourra récupérer les données d'intérêt depuis les bases de données d'observations pour l'assimilation pour ensuite les étudier dans un code simplifié, élaboré en Python par exemple.

Ensuite, on traitera le cas spécifique où la matrice \mathbf{R} est modélisée par un opérateur de diffusion. Il s'agira alors de s'approprier la méthode des éléments finis d'ordre 1 et la triangulation de Delaunay pour bien comprendre le rôle du maillage et la distribution des observations sous-jacente dans la structure de la matrice \mathbf{R} . Pour cela, on s'appuiera sur l'étude très complète de Shewchuk (2002).

Enfin, à la lumière des résultats précédents et si le temps le permet, on implémentera une nouvelle méthode d'écrémage dans le code des modèles Arpege et Arome, basée sur des critères objectifs de qualité du maillage, et on commentera ses performances par rapport aux approches actuelles. Plusieurs points de départ sont envisageables, comme l'effondrement d'arêtes (*edge collapsing*, Hoppe, 1996) ou l'inflation de sphères (Miller et al. 1999).

1.8 Résultats envisageables

- Une étude théorique de l'écrémage et de son impact sur le contenu de la matrice \mathbf{R} dans les cas corrélé et non corrélé ;
- Un nouveau schéma d'écrémage incluant des critères géométriques de génération de maillage pour préparer la modélisation des corrélations d'erreurs d'observation.

1.9 Références bibliographiques sélectionnées

- N. Bormann, S. Saarinen, G. Kelly, and J.-N. Thépaut. The Spatial Structure of Observation Errors in Atmospheric Motion Vectors from Geostationary Satellite Data. *Monthly Weather Review*, pages 706-718, 2002.
- N. Bormann and P. Bauer. Estimates of spatial and interchannel observation error characteristics for current sounder radiances for numerical weather prediction. I : Methods and application to ATOVS data. *Quarterly Journal of the Royal Meteorological Society*, 136 :1036–1050, 2010.
- N. Bormann, A. Collard, and P. Bauer. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II : Application to AIRS and IASI data. *Quarterly Journal of the Royal Meteorological Society*, 136 :1051–1063, 2010.
- O. Guillet, A. Weaver, X. Vasseur, Y. Michel, S. Gratton, and S. Gurol. Modelling spatially correlated observation errors in variational data assimilation using a diffusion operator on an unstructured mesh. *Quarterly Journal of the Royal Meteorological Society*, 2019.
- H. Hoppe. Progressive meshes. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '96*, pages 99–108, 1996.
- Z.-Q. Liu and F. Rabier. The interaction between model resolution, observation resolution and observation density in data assimilation : A one-dimensional study. *Quarterly Journal of the Royal Meteorological Society*, 128 :1367–1386, 2002.
- G. L. Miller, D. Talmor, and S. Teng. Optimal Coarsening of Unstructured Meshes. *Journal of Algorithms*, 1999.
- J. R. Shewchuk. What is a good linear finite element ? - interpolation, conditioning, anisotropy, and quality measures. Technical report, In *Proc. of the 11th International Meshing Roundtable*, 2002.

- J. A. Waller, S. Ballard, S. L. Dance, G. Kelly, N. K. Nichols, and D. Simonin. Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sens.*, 8 :581, 2016a.
- J. A. Waller, S. L. Dance, and N. K. Nichols. Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis residuals. *Quarterly Journal of the Royal Meteorological Society*, 142 :418–431, 2016b.
- J. A. Waller, D. Simonin, S. L. Dance, N. K. Nichols, and S. Ballard. Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*, 144 :3533–3551, 2016c.