

Soutenance de thèse

Olivier GOUX soutiendra sa thèse de doctorat, préparée au sein de l'équipe d'accueil doctoral CECI et intitulée « *Prise en compte des corrélations d'erreur d'observation en assimilation variationnelle de données océaniques : application aux données altimétriques* »

**Le 20 février 2025 à 14h00,
Salle JCA CERFACS, 42 avenue Gaspard Coriolis, 31100 Toulouse**

devant le jury composé de

| | | |
|---------------------|-------------------------------------------------------|----------------------|
| M. Anthony WEAVER | CERFACS | Directeur de thèse |
| M. Oliver GUILLET | CNRM | Codirecteur de thèse |
| Mme Sarah DANCE | University of Reading | Rapporteuse |
| M. Andrew MOORE | University of California, Santa Cruz | Rapporteur |
| M. Massimo BONAVITA | European Centre for Medium-Range Weather Forecasts | Examineur |
| Mme Nadia FOURRIÉ | CNRM | Examinatrice |
| M. Arthur VIDARD | INRIA | Examineur |

Résumé : L'assimilation de données consiste en un ensemble de méthodes qui peuvent être utilisées pour corriger une estimation initiale de l'état d'un système, appelé l'ébauche, à l'aide d'observations de ce système. En géophysique, le système en question peut être par exemple l'océan, dont l'état est défini comme les valeurs prises par plusieurs variables physiques (telles que la température ou la salinité) en chaque point d'une grille tridimensionnelle. Pour estimer l'état de l'océan un jour donné, l'ébauche pourrait être par exemple une prévision réalisée la veille, qui aurait besoin d'être corrigée avec des observations plus récentes. Les observations peuvent provenir de nombreuses sources hétérogènes, telles que des satellites ou des instruments in-situ. L'état corrigé du système qui en résulte est appelé analyse, et peut être utilisé par exemple comme condition initiale par le modèle utilisé pour établir les prévisions du lendemain. L'ébauche et des observations sont affectés par des erreurs, que nous supposons décrites statistiquement par leurs matrices de covariance : B pour l'erreur d'ébauche et R pour l'erreur d'observation. L'assimilation de données variationnelle est une catégorie de méthodes d'assimilation des données dans laquelle l'analyse est approchée progressivement en minimisant itérativement une fonction de coût mesurant simultanément l'adéquation d'un état du système à l'ébauche et aux observations, avec des poids définis par B^{-1} et R^{-1} . Dans les prévisions météorologiques opérationnelles, l'ébauche contient souvent des milliards de degrés de libertés, et sera combinée avec des millions d'observations tous les jours. Cela rend notamment la construction explicite des matrices B et R trop coûteuse pour être réalisée. Les approches itératives utilisées en assimilation de données sont souvent utilisées pour contourner le problème car elles ne nécessitent des opérateurs modélisant des produits matrice-vecteur sans accès explicite aux matrices elles-mêmes. Les erreurs d'observation sont souvent supposées non corrélées (et donc R diagonal) pour simplifier l'accès à l'opérateur de corrélation inverse, R^{-1} , qui apparaît dans de nombreuses formulations. Cependant, cette hypothèse n'est pas réaliste pour certains types d'observations, en particulier les données satellitaires à haute résolution. Négliger les corrélations des erreurs d'observation pendant l'assimilation mène généralement à des analyses sous-optimales, trop proches des observations aux grandes échelles spatiales, et trop distantes des observations aux petites échelles spatiales. Pour résoudre ce problème, nous avons conçu un opérateur de corrélation associé aux erreurs d'observation basé sur un opérateur de diffusion pour le système d'assimilation de données

océaniques NEMOVAR. Les opérateurs de diffusion permettent de modéliser de manière efficace et flexible l'inverse de l'opérateur de corrélation de l'erreur d'observation, et ce même avec des données non structurées. La prise en compte des corrélations affectant les erreurs d'observation a pour but d'améliorer la qualité de l'analyse, mais elle risque également d'affecter le taux de convergence des algorithmes de minimisation utilisés pour approximer cette analyse. Dans les applications opérationnelles, le processus de minimisation est généralement tronqué avant d'atteindre la convergence totale. En conséquence, même si les corrélations d'erreurs d'observation sont correctement prises en compte, elles pourraient potentiellement compromettre la qualité de la solution obtenue. Sur la base de résultats analytiques et numériques, nous explorons l'influence des corrélations d'erreur d'observation sur la sensibilité et le taux de convergence des algorithmes d'assimilation variationnelle de données. En particulier, nous cherchons à comprendre comment choisir un modèle de corrélation des erreurs d'observation pour refléter un équilibre entre l'efficacité du calcul et la précision de la solution.

Mots-clés : assimilation de données, diffusion, corrélations, altimétrie

Summary: Data assimilation consists of an array of methods which can be used to correct an a priori estimate of the state of a system, called the background state, with observations of this system. In the context of geophysical applications, the system in question can be, for example, the ocean, and its state would be defined as the values taken by several physical variables (such as temperature or salinity) at every point on a three-dimensional grid. If we were trying to estimate the state of the ocean on a given day, the background state could be, for example, a forecast initiated the day before, which needs to be corrected with the latest observations. Observations of the system can come from many heterogeneous sources, such as satellites and in situ networks. The resulting corrected state of the system is called the analysis, and is typically used as the model initial condition for producing the forecast for the next day. In order for the analysis to be an accurate representation of the state of the system, we first need to know how accurate the background state and observations are. Both are affected by errors, which we assume can be described statistically by their covariance matrices: B for the background error, and R for the observation error. Variational data assimilation is a class of data assimilation methods where the analysis is obtained by minimising iteratively a cost function measuring the fit of a candidate state of the system to both the background and observations, with weights defined by B^{-1} and R^{-1} . However, most of the complexity of data assimilation pertains to adapting these basic methods to the constraints of the real system, in order to obtain methods that are practical for real applications. In Numerical Weather Prediction, the background state of the system can often represent billions of degrees of freedom, which need to be combined with millions of observations every day. In such a system, the explicit construction of the matrices B and R is too expensive to be done in practice. The construction of these matrices can be avoided using the iterative approaches of variational data assimilation, which only require operators that represent matrix-vectors products. The assumption of uncorrelated observation errors (diagonal R) is commonly made to simplify the inverse covariance operator R^{-1} , which is required by many minimization algorithms. However, this assumption becomes problematic when dealing with certain observation types, notably high-resolution satellite data. Neglecting observation error correlations during assimilation often results in suboptimal analyses, where observations tend to be overfitted at larger spatial scales and underfitted at smaller scales. To address this issue, we have developed an observation error correlation operator based on a diffusion operator for the ocean data assimilation system NEMOVAR. Diffusion operators — initially designed for modelling correlations in background error — offer a cost-effective and flexible framework for modelling the inverse observation error correlation operator (and the operator itself) with unstructured data. While accounting for observation error correlations should improve the accuracy of the solution, it also affects the convergence rate of the minimisation algorithms used to approximate the solution. In operational applications, where the minimisation

process is usually truncated before achieving full convergence, even correctly accounted for observation error correlations might therefore compromise the accuracy of the analysis. Through analytical and numerical results, we explore the influence of the observation-error correlations on both the sensitivity and convergence rate of variational data assimilation algorithms. In particular, we provide insights into how the choice of an observation error correlation model must reflect a balance between computational efficiency and solution accuracy.

Keywords: data assimilation, diffusion, correlations, altimetry