HAL open science

# Quelques contributions à l'assimilation de données : des moindres carrés non-linéaires pondérés en grande dimension, applications en océanographie

Ehouarn Simon

**To cite this version:**

Ehouarn Simon. Quelques contributions à l'assimilation de données : des moindres carrés non-linéaires pondérés en grande dimension, applications en océanographie. 2023. hal-04084111v1

**HAL Id: hal-04084111**

**https://hal.science/hal-04084111v1**

Preprint submitted on 27 Apr 2023 (v1), last revised 12 May 2023 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

**TOULOUSE INP**

# Mémoire

Pour obtenir
**l'Habilitation à diriger des recherches - Toulouse INP**

Discipline : **Mathématiques Appliquées**

Présentée par

**Ehouarn SIMON**

---

# Quelques contributions à l'assimilation de données : des moindres carrés non-linéaires pondérés en grande dimension, applications en océanographie

---

# Table des matières


**1 Introduction** **4**

**2 Extension des filtres de Kalman d'ensemble pour la prise en compte de contraintes** **7**
2.1 Filtres de Kalman d'ensemble et non-gaussiannité . . . . . 7
2.1.1 Anamorphose gaussienne . . . . . 8
2.1.2 Vers la prise en compte de contraintes de somme et de positivité . . . . . 13
2.2 Application à la prévision quasi-opérationnelle en océanographie . . . . . 21
2.2.1 Des systèmes d'assimilation en constante mutation . . . . . 21
2.2.2 Des données produites massivement et à valoriser . . . . . 23

**3 Des problèmes d'estimation en très grande dimension : de la possibilité de réduire le volume des données** **29**
3.1 Au niveau des observations . . . . . 29
3.1.1 Une approche multi-niveaux dans l'espace des observations . . . . . 30
3.1.2 De la sensibilité de l'incrément d'analyse pour la sélection des observations à assimiler . . . . . 36
3.2 Au niveau du modèle . . . . . 39
3.2.1 Approche par réduction de modèles ou apprentissage profond . . . . . 39
3.2.2 Methodes d'ensemble . . . . . 42

**4 Des problèmes aux moindres carrés non-linéaires pondérés** **49**
4.1 Des normes $\|\|_{\mathbf{D}^{-1}}$ . . . . . 49
4.1.1 Un préconditionnement non trivial . . . . . 50
4.1.2 Formulation point-selle pour l'algorithme 4D-VAR à contraintes faibles : quel critère d'arrêt ? . . . . . 53
4.2 Des calculs inexacts pour une convergence garantie . . . . . 57
4.2.1 Un algorithme du gradient conjugué inexact pour la minimisation de quadratique convexe . . . . . 57
4.2.2 Une variante de GMRES avec produits scalaires inexacts . . . . . 61
4.3 Des normes non-standards . . . . . 66
4.3.1 Régularisation en norme $\|\|_p$ en assimilation variationnelle de données . . 66
4.3.2 Vers l'introduction de normes non différentiables . . . . . 69

**5 Quelques perspectives** **76**
5.1 Vers une assimilation de données ensembliste multi-fidèle . . . . . 76
5.2 Assimilation de données et non-linéarité : les méthodes à noyaux . . . . . 81

5.3 Des problèmes issus de l'océanographie et plus généralement de la modélisation du système climatique terrestre . . . 85

**A Curriculum Vitae** **91**
A.1 Etudes et expériences professionnelles . . . 91
A.1.1 Parcours universitaire . . . 91
A.1.2 Déroulement de carrière . . . 91
A.2 Liste des communications . . . 92
A.2.1 Revues internationales . . . 92
A.2.2 Actes de conférences, chapitres de livres, vulgarisation scientifique . . . 93
A.2.3 Développements technologiques . . . 94
A.2.4 Communications orales et posters . . . 95
A.3 Encadrements d'étudiants et chercheurs . . . 97
A.3.1 Post-doctorant . . . 97
A.3.2 Doctorants . . . 97
A.3.3 Stage de Master 2 . . . 100
A.4 Coordination de projets . . . 100
A.5 Responsabilités collectives . . . 101
A.5.1 Enseignement . . . 101
A.5.2 Recherche . . . 101

# Chapitre 1

# Introduction

J'ai commencé mes activités de recherche en m'intéressant à l'assimilation de données. Ces classes de méthodes visent à combiner de manière "optimale", dans un sens à définir, les informations mathématiques issues des modèles, et les informations "physiques" issues des observations, en tenant compte de leurs incertitudes, en vue de reconstituer l'état d'un système. Ces approches sont naturellement utilisées en météorologie, océanographie, science du climat, géosciences pour l'analyse passée et la prédiction de l'évolution du système. C'est un champs disciplinaire riche de par les constants progrès de la modélisation numérique, avec notamment le couplage de systèmes complexes, de l'apparition de nouveaux types d'observations, de l'augmentation de la puissance de calcul, et de l'essor récent des méthodes issues de l'apprentissage machine dans ce domaine. De nombreux livres [48, 117, 4, 49], ainsi que les références associées, décrivent le large spectre de ces méthodes, leur application à des problèmes réalistes, les difficultés rencontrées ainsi que les perspectives méthodologiques.

Me concernant, suite à ma thèse dans l'équipe MOISE du Laboratoire Jean Kuntzmann (LJK), pendant laquelle je m'étais intéressé à des développements algorithmiques en assimilation variationnelle de données, avec application à des modèles numériques d'océans présentant des raffinements de maillages locaux, ainsi que l'introduction des méthodes multi-grilles, je suis allé étudier les approches stochastiques en assimilation de données via les méthodes de filtrage d'ensemble dans le cadre de mon séjour au Nansen Environmental and Remote Sensing Center (NERSC, Norvège). Ces travaux ont porté sur le développement d'une extension non-gaussienne des filtres de Kalman d'ensemble par changements de variables et leurs applications dans le cadre de l'estimation conjointe de variables d'état et paramètres dans les modèles couplés océan-écosystème. Par la suite, j'ai été recruté comme maître de conférences dans l'équipe Algorithmes Parallèles et Optimisation de l'Institut de Recherche en Informatique de Toulouse, (IRIT) dans laquelle j'ai pu continuer et continue mes activités de recherche en assimilation de données, tout en ayant évolué vers des développements méthodologiques en algèbre linéaire numérique et en optimisation.

Dans ce manuscrit, je m'attache à présenter mes activités de recherche depuis celles débutées lors de mon post-doctorat en Norvège, jusqu'à celles plus récentes réalisées à l'IRIT. Dans un premier temps, je présente une application, dite réaliste, de l'océanographie, à laquelle je me suis intéressé, à savoir l'estimation de variables d'état et de paramètres de modèles de biogéochimie

marine, par assimilation de données, ainsi que les développements méthodologiques associés (chapitre 2). Je présente ensuite quelques contributions à la réduction de la dimension de problèmes d'assimilation de données, dans le but de réduire les coûts de calcul associés (chapitre 3). Je m'intéresse également à la modélisation mathématique de l'assimilation de données, vue comme des problèmes aux moindres-carrés non linéaires pondérés, et présente quelques développements méthodologiques associés (chapitre 4). Enfin, je propose quelques perspectives à ces travaux (chapitre 5).

# Chapitre 2

# Extension des filtres de Kalman d'ensemble pour la prise en compte de contraintes

**Sommaire**


**2.1 Filtres de Kalman d'ensemble et non-gaussiannité . . . . . 7**
2.1.1 Anamorphose gaussienne . . . . . 8
2.1.2 Vers la prise en compte de contraintes de somme et de positivité . . . . 13
**2.2 Application à la prévision quasi-opérationnelle en océanographie . . 21**
2.2.1 Des systèmes d'assimilation en constante mutation . . . . . 21
2.2.2 Des données produites massivement et à valoriser . . . . . 23


*Ce chapitre porte sur des axes de recherche étudiés à la sortie de ma thèse, essentiellement lors de mes années au NERSC, en tant que post-doctorant, puis scientifique sur projets. Ces travaux ont porté sur le développement d'extensions non-gaussiennes des filtres de Kalman d'ensemble, basées sur des changements de variables "judicieux", appelés anamorphoses gaussiennes, ainsi que leur application à l'estimation conjointe état - paramètres dans le cadre de modèles de biogéochimie marine de complexité variée. Ceci m'a naturellement amené à travailler dans le cadre de l'océanographie pré-opérationnelle, à la fois sur le développement de systèmes d'assimilation réalistes, ainsi que sur la production d'une réanalyse de la biogéochimie marine de l'océan Arctique.*

## 2.1 Filtres de Kalman d'ensemble et non-gaussiannité

Le développement des supercalculateurs au cours des dernières décennies a entraîné des progrès considérables dans la prévision des modèles du système terrestre. Néanmoins, ces modèles restent encore incertains et présentent de nombreuses erreurs liées aux approximations théoriques sur lesquelles ils sont construits, à la discrétisation des équations "continues", et au choix des résolutions utilisées. Les observations constituent une seconde source d'information sur ces systèmes. Leur distribution spatio-temporelle est incomplète et hétérogène, leur support varie de la mesure quasi-ponctuelle aux échelles de plusieurs dizaines de kilomètres, et elles peuvent

présenter de larges erreurs. Les méthodes d'assimilation de données permettent de combiner les informations hétérogènes et incertaines fournies à la fois par les modèles numériques, issus de la discrétisation de problèmes d'équations aux dérivées partielles, et les observations afin d'estimer l'état et/ou certains paramètres d'un système. Elles ont été historiquement introduites pour la prévision météorologique, et permettent de traiter des problèmes en grande dimension, tels que ceux rencontrés dans les géosciences [29].

Une classe de méthodes ayant connu un fort essor sont celles dites ensemblistes, et notamment les variantes des filtres et lisseurs de Kalman d'ensemble [47]. Ces méthodes exploitent l'évolution d'un ensemble de trajectoires modèles afin d'approximer certaines quantités présentes dans le processus d'estimation (p.e. les matrices de covariance d'erreur). Sous les hypothèses de linéarité des différents opérateurs introduits dans le processus d'estimation, et de distributions gaussiennes des erreurs et variables, celles-ci s'interprètent comme la résolution numérique par échantillonnage d'un problème d'estimation bayésienne. Néanmoins, ces hypothèses n'étant pas vérifiées pour les fluides géophysiques, l'utilisation de ces méthodes reste suboptimale en terme de réduction de la variance d'erreur. Ceci est notamment vérifié dans le cadre de leur utilisation en biogéochimie marine pour laquelle des contraintes de positivité s'appliquent aux différentes variables d'état et paramètres, pouvant conduire à leur divergence [127].

### 2.1.1 Anamorphose gaussienne

Je me suis intéressé aux extensions non-gaussiennes des filtres de Kalman d'ensemble par anamorphose gaussienne [10]. Le principe consiste à appliquer, préalablement à l'analyse, un changement de variables univarié spécifique pour chacune des variables d'état, des observations et des paramètres à estimer, afin d'obtenir des variables "transformées" suivant chacune une loi Normale. L'algorithme, ainsi proposé [10], se base sur la structure de l'EnKF et se décompose en deux étapes : l'une de prévision, et l'autre d'analyse.

**Prévision** : l'étape de prévision se fait via la propagation d'un ensemble de $N$ simulations, appelées membres de l'ensemble, afin d'approximer la densité de prévision

$$\forall i = 1 : N, \quad \mathbf{x}_n^{f,i} = f_{n-1}(\mathbf{x}_{n-1}^{a,i}, \epsilon_n^{m,i}) \tag{2.1}$$

avec $\mathbf{x}_n$ le vecteur d'état au temps $t_n$, $f_{n-1}$ le modèle non-linéaire et $\epsilon_n^m$ l'erreur modèle.

**Analyse** : l'étape d'analyse conditionne chacun des membres de l'ensemble de prévision aux nouvelles observations $\mathbf{y}_n$ et ce de manière linéaire. Les fonctions anamorphoses sont introduites dans cette étape.

Pour chacune des variables du modèle au temps $t_n$, nous appliquons un changement de variables non-linéaire $\psi_n$, qui va de l'espace de la variable d'état vers un espace d'une variable gaussienne. Chacune des variables est transformée séparément des autres. Afin de simplifier les notations, nous supposons avoir qu'une seule variable dans le modèle (et donc une seule fonction $\psi_n$). Ceci s'écrit

$$\forall i = 1 : N, \quad \tilde{\mathbf{x}}_n^{f,i} = \psi_n(\mathbf{x}_n^{f,i}) \tag{2.2}$$

En pratique, un changement de variables est appliqué à chaque variable en chacun des points de grille du domaine.

De la même manière, un changement de variables $\chi_n$ est introduit pour les observations $\mathbf{y}_n$ au temps $t_n$ :

$$\tilde{\mathbf{y}}_n = \chi_n(\mathbf{y}_n). \tag{2.3}$$

En notant $\mathbf{H}$ l'opérateur d'observation, faisant le lien entre l'espace du vecteur d'état et celui des observations, le nouvel opérateur d'observation $\tilde{\mathbf{H}}_n$ faisant le lien entre les variables d'état et observations transformées est défini par

$$\tilde{\mathbf{H}}_n = \chi_n \circ \mathbf{H} \circ \psi_n^{-1} \tag{2.4}$$

avec $\circ$ la composition de fonctions. Si $\tilde{\mathbf{H}}_n$ est linéaire, ce qui est par exemple le cas lorsque les observations sont des composantes du vecteur d'état, l'étape d'analyse linéaire sur les variables transformées s'écrit formellement comme l'étape d'analyse de l'EnKF :

$$\forall i = 1 : N, \quad \tilde{\mathbf{x}}_n^{a,i} = \tilde{\mathbf{x}}_n^{f,i} + \tilde{\mathbf{K}}_n(\tilde{\mathbf{y}}_n - \tilde{\mathbf{H}}_n \tilde{\mathbf{x}}_n^{f,i} + \epsilon_n^{o,i}) \tag{2.5}$$

avec $\tilde{\mathbf{K}}_n$ la matrice du gain de Kalman associée aux variables transformées et $\epsilon_n^{o,i}$ l'erreur d'observation associée aux observations transformées et suivant une loi Normale. Dans le cas d'un opérateur $\tilde{\mathbf{H}}_n$ non-linéaire, il est également possible d'appliquer la stratégie proposée par [47].

Le retour dans l'espace des variables d'état est obtenu par l'utilisation de l'inverse du changement de variables :

$$\forall i = 1 : N, \quad \mathbf{x}_n^{a,i} = \psi_n^{-1}(\tilde{\mathbf{x}}_n^{a,i}) \tag{2.6}$$

Les moyenne $\mathbf{x}_n^a$ et matrice de covariance d'erreur $\mathbf{P}_n^a$ d'analyse sont naturellement approximées par leurs estimateurs empiriques depuis l'ensemble d'analyse $(\mathbf{x}_n^{a,i})_{i=1:N}$.

Cet algorithme basé sur l'utilisation de fonctions d'anamorphose univariées ne gère pas la non-gaussianité multivariée du vecteur d'état. Même si chaque variable transformée suit une distribution gaussienne, leurs distributions multivariées ne seront pas nécessairement gaussiennes. En pratique, cette propriété est difficile à vérifier en raison de la grande taille des vecteurs, et des coûts de calculs associés. Néanmoins, il est espéré une "amélioration" des distributions des variables transformées, résultant en de meilleurs résultats en terme d'estimation, comparativement à ne pas appliquer ces transformations. Dans la suite de ce chapitre, l'emploi d'expression du type "variable transformée gaussienne" renverra à la distribution univariée de la variable transformée.

Une difficulté fondamentale de cette approche réside dans la construction de ces changements de variables, sachant que nous ne connaissons pas a priori la distribution marginale des variables d'intérêt, mais que nous disposons d'échantillons de celles-ci. Notons $\mathbf{Z}(x)$ une telle variable distribuée spatialement et $(z_i)_{i=1:N}$ un échantillon de cette variable.

Le but, dans un premier temps, est de construire une fonction constante par morceaux $\psi$ telle que

$$\mathbf{Z}(x) = \psi(\mathbf{Y}(x)) \tag{2.7}$$

avec $\mathbf{Y}(x)$ suivant une distribution Normale $\mathcal{N}(0,1)$. Celle-ci est appelée anamorphose empirique par la suite. Sa construction se fait de la manière suivante :

— Tri des données $(z_i)_{i=1,N}$ par ordre croissant des valeurs prises

$$z_1 < z_2 < ... < z_{N-1} < z_N, \tag{2.8}$$

— Calcul d'un échantillon $(y_i)_{i=1:N}$ de la variable Normale $\mathbf{Y}(x)$.

$$\forall i = 1 : N, \quad y_i = G^{-1}(\frac{i}{N}) \tag{2.9}$$

avec $G$ la fonction de répartition de la variable $\mathbf{Y}(x)$ :

$$G(t) \quad = P(\mathbf{Y}(x) < t) \quad = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy, \tag{2.10}$$

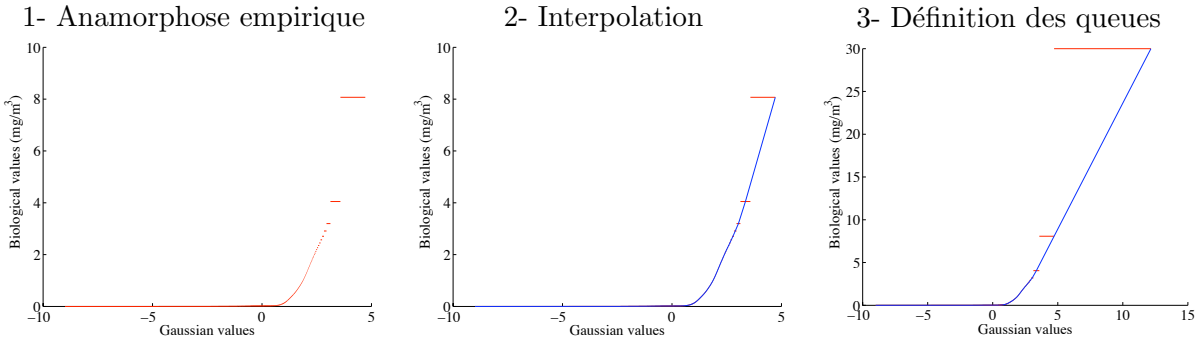



FIGURE 2.1 – Chlorophylle de surface (NORWECOM) : illustration des trois étapes de la construction d'une anamorphose univariée. Figure tirée de Simon et Bertino (2009) [125].

— Définition de l'anamorphose empirique $\psi$ :

$$\psi(y) = \sum_{i=1}^{N} z_i 1_{[y_{i-1}, y_i[}(y). \tag{2.11}$$

Cette fonction n'étant pas bijective, il est nécessaire de construire un nouveau changement de variables par interpolation. De surcroît, selon l'échantillon disponible, il peut être nécessaire de modéliser des queues, afin de tenir compte de l'information a priori dont nous disposons, en dehors des valeurs prises par l'échantillon. Ces trois étapes de la construction de la fonction anamorphose sont illustrées sur la Figure 2.1, pour le cas de la concentration de chlorophylle, telle qe simulée depuis le modèle HYCOM-NORWECOM dans les océans Atlantique Nord et Arctique. Il est à noter que d'autres stratégies existent pour la construction des fonctions anamorphoses. Elles se basent notamment sur la spécification a priori de quantiles cibles, permettant la construction de la fonction anamorphose depuis les échantillons des variables. Plus de détails sont disponibles dans [26, 43, 23].

La Figure 2.2 illustre l'impact de telles transformations sur la distribution de la concentration de Diatomées dans l'Atlantique Nord et l'Arctique. L'échantillon utilisé provient des résultats d'une simulation historique de quatre ans, pour laquelle une fenêtre glissante de trois mois autour de la date d'analyse est appliquée avant extraction des valeurs. L'échantillon ne provient donc pas directement depuis les valeurs présentes dans l'ensemble au moment de l'analyse, ce qui fait que la distribution marginale obtenue n'est pas gaussienne, même si nous observons visuellement une meilleure adéquation comparativement à ne pas appliquer le changement de variables. Nous pouvons voir sur cet exemple que l'échantillon utilisé pour construire la fonction anamorphose joue un rôle crucial sur la pertinence de celle-ci vis-à-vis de l'ensemble de prévision. Plus de détails concernant ces travaux se trouvent dans [125].

Nous nous sommes donc intéressés à différentes stratégies, pour le choix de l'échantillon à utiliser pour la construction des fonctions anamorphoses dans le cas de l'estimation jointe du vecteur d'état et de certains paramètres du modèle : simulation d'ensemble sans assimilation, ensemble de prévision, et hybridation selon la nature des variables (observations ou variable d'état) [127]. Ces stratégies sont définies ainsi.

1. La première approche, dite statique, propose d'utiliser des échantillons issus d'une simulation d'ensemble avec un faible nombre de membres, afin de réduire les temps et coût de calcul, ainsi que le coût de stockage des données simulées. Considérant le cadre particulier

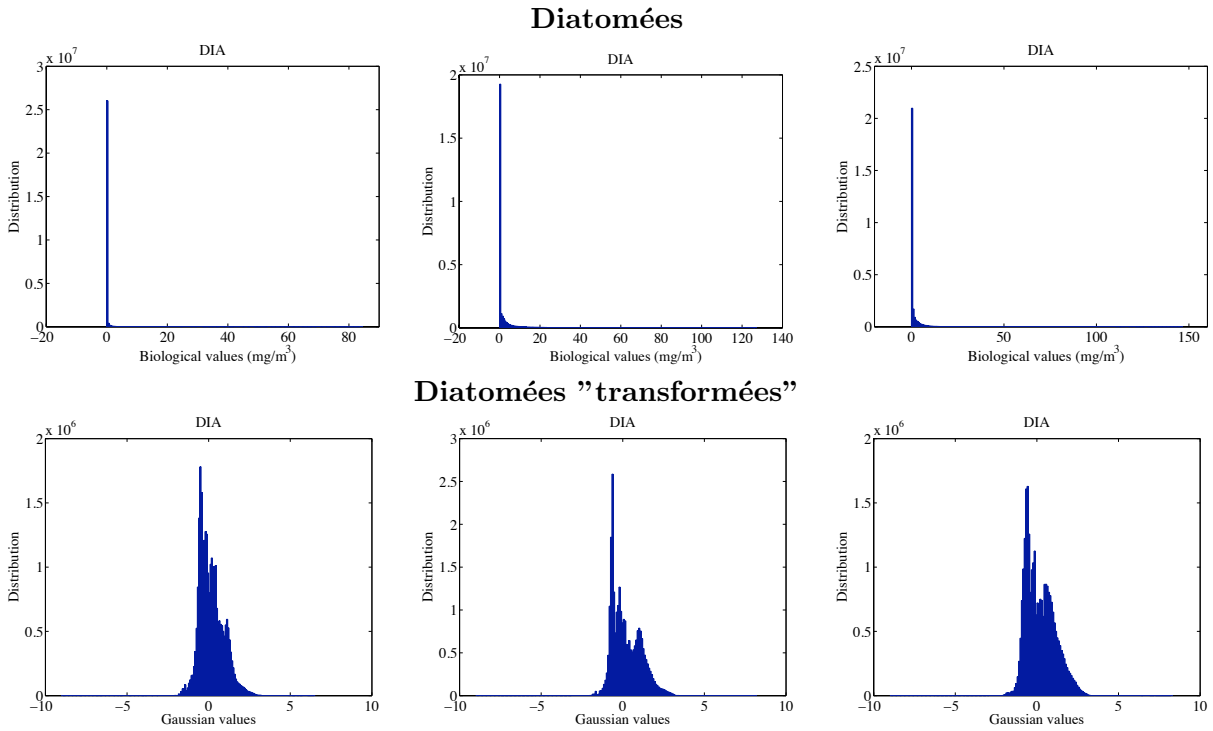



FIGURE 2.2 – Distributions de la concentration de Diatomées (variables biologiques et transformées par anamorphose) pour trois différentes dates (décembre, mai et septembre). Figure tirée de Simon et Bertino (2009) [125].

de l'estimation jointe variables d'état et paramètres, celui-ci rend peu pertinent l'emploi d'une simulation déterministe, pour une réalisation particulière de paramètres. Une fenêtre glissante est ensuite utiliser pour sélectionner les valeurs des variables autour de la date de l'analyse.

2. La seconde approche, dite dynamique, propose d'utiliser directement les valeurs présentes dans l'ensemble de prévision à la date de l'analyse pour constituer l'échantillon sur lequel construire l'anamorphose. Cette approche à le mérite d'obtenir des variables transformées gaussiennes. Néanmoins elle est sensible au biais du modèle, et laisse à la modélisation des queues des anamorphoses la transformation des valeurs en dehors de celles présentes dans l'ensemble. Ceci est problématique, tant pour les observations en dehors de l'intervalle de valeurs présentes dans l'ensemble, que dans le cas de biais de modèle fort, ou lorsque certaines variables voient leur variance s'écrouler au fil des cycles d'analyse.
3. La troisième, appelée hybride, vise à tirer profit des deux précédentes approches, et consiste à appliquer l'approche statique pour les variables observées, et dynamique pour les autres. L'intérêt de l'approche statique réside dans les grands intervalles de valeurs disponibles pour la construction de l'anamorphose empirique, ce qui lui permet d'être plus robuste aux différences possibles entre les distributions des observations et des variables observées du modèle. L'utilisation de l'approche dynamique pour les variables non-observées permet d'obtenir des variables transformées ayant bien une distribution gaussienne.

Il est également possible, pour les variables observées, d'ajouter les valeurs issues des observations, passées ou en cours d'assimilation. Ceci a l'avantage, notamment pour l'approche dynamique, d'éviter l'utilisation des queues d'anamorphoses pour la transformation des observations en dehors de l'intervalle des valeurs présentes dans l'ensemble.

Nous avons comparé les performances du DEnKF [123], avec celles de ses variantes avec anamorphoses gaussiennes, pour les différentes stratégies présentées, ainsi que celle basée sur le changement de variables logarithmique. Les expériences numériques ont été réalisées avec un modèle d'écosystème marin à trois variables, de type NPZ pour Nutriments - Phytoplancton - Zooplancton, dans une colonne d'eau (domaine 1D vertical). Celles-ci ont été répétées 20 fois (différentes valeurs de paramètres, conditions initiales et observations), afin de présenter des résultats plus robustes face à l'aléa d'une unique configuration expérimentale. La Figure 2.3 représente l'évolution temporelle de la moyenne (courbe noire) et de la moyenne plus/moins un écart type (zone grise) des trois paramètres estimés, à savoir les taux de perte métabolique des plantes, de l'efficacité de broutage et de la perte due aux carnivores. Ces courbes correspondent aux moyennes de ces quantités calculées sur les 20 expériences. La valeur "vraie", utilisé par la simulation de référence à l'origine de observations, correspond à la courbe noire pointillée. Comme attendu, les coefficients des paramétrisations linéaires impliquant les variables observées peuvent être estimés efficacement avec les méthodes d'ensemble de base. Le DEnKF sans anamorphose est ainsi capable d'estimer correctement le paramètre du taux de perte métabolique des plantes. Par contre, l'estimation des coefficients des paramétrisations non linéaires ou n'affectant pas directement les variables observées reste un problème difficile. Nous constatons que les corrections sur les paramètres de taux d'efficacité de broutage et de la perte due aux carnivores sont dans la direction opposée lors de la première floraison pour le système d'assimilation sans anamorphose. Cela conduit à des moyennes finales inférieures aux valeurs prescrites a priori lors de l'initialisation des ensembles, et ce alors que la valeur "vraie" est plus élevée. De plus, l'estimation semble sensible à la contribution aléatoire des erreurs d'observation, ainsi qu'aux perturbations, comme le montre le nuage de points, présent sur la Figure 2.4, et correspondant

à l'estimation en fin de simulation de la moyenne de l'ensemble du taux d'efficacité de broutage versus taux de perte due aux carnivores pour les 20 expériences réalisées. Nous notons que l'assimilation dégrade les deux paramètres dans la plupart des expériences : environ 90% des points appartiennent au coin inférieur gauche, indiquant des corrections dans la mauvaise direction pour les deux paramètres.

L'introduction des anamorphoses conduit à une amélioration significative de l'estimation des paramètres des taux d'efficacité de broutage et de la perte due aux carnivores. Le "choc" lors de la première fleuraison est beaucoup mieux absorbé, évitant ainsi un collapse des composantes des ensembles associées à ces deux paramètres, et conduisant ainsi à une amélioration de la moyenne en sortie de cette première phase de floraison, puis dans une grande majorité des analyses tout au cours des simulations. Ceci suggère que l'assimilation a fourni des corrections dans la direction de la vraie valeur des paramètres. Nous observons ainsi sur la Figure 2.4 que la plupart des valeurs obtenues en fin de simulation sont dans la bon quadrant, à savoir celui où se trouvent les valeurs de référence. Seule l'anamorphose statique conduit à une dégradation de la valeurs des deux paramètres pour quelques expériences. Sur la base de ces expériences, la stratégie d'anamorphose hybride nous a semblé la plus pertinente parmi celles construites empiriquement. Il est à noter que la fonction logarithme conduit également à de très bons résultats, et ne doit donc pas être à exclure par défaut, d'autant plus quand des hypothèses de distribution log-Normale sont inhérentes à la modélisation ou aux observations.

Ces travaux ont été poursuivis jusqu'à l'assimilation de données réelles, dans un modèle d'écosystème marin plus complexe et représentatif d'une colonne d'eau dans la mer du Nord (station Mike), dans le cadre d'une collaboration avec M. Gharamti, A. Samuelsen, L. Bertino, U. Dewel et A. Korosov [61]. Ces travaux ont porté sur l'évaluation des performances de différentes stratégies pour estimer conjointement les variables d'état et des paramètres du modèle, dans un modèle de complexité similaire à celle des modèles utilisés opérationnellement. Nous avons constaté que l'application de fonctions d'anamorphoses lors des phases d'analyse résultaient en une représentation des distributions de l'ensemble de prévision ayant des formes gaussiennes appropriées. De plus, les paramètres optimisés ont permis une amélioration des capacités de prédiction, ainsi que de la variabilité saisonnière, du modèle couplé 1D.

### 2.1.2 Vers la prise en compte de contraintes de somme et de positivité

Le développement des modèles numériques de biogéochimie océanique au cours des deux dernières décennies a conduit à des représentations de plus en plus complexes des interactions entre les différents niveaux trophiques, notamment entre les différentes espèces de plancton à la base de la chaîne alimentaire. Alors que le régime alimentaire du zooplancton est représenté de manière relativement simple dans les premiers modèles NPZ, tel que celui utilisé dans [127], à savoir que l'unique groupe de zooplancton (Z) se nourrit uniquement du groupe de phytoplancton (P) lui aussi unique, l'ajout de multiples types fonctionnels de plancton (PFT) pour le phytoplancton et le zooplancton conduit à des régimes alimentaires plus complexes et des préférences de broutage sont ainsi ajoutées. Ces paramètres sont toujours positifs et, bien que non obligatoires, leur somme est généralement égale à un. Les préférences de broutage spécifient la direction de l'alimentation dans l'espace des aliments, et donc, la direction du transfert des PFT (l'aliment) vers les PFT du zooplancton (le nourricier). Par conséquent, leur impact sur la distribution des différents PFTs obtenus à partir d'une simulation de modèle peut être significatif. Nous nous sommes donc intéressés à leur possible estimation par méthodes de filtrage de Kalman d'ensemble [128]. Le problème est alors le suivant, soit $(\pi_i)_{i=1:N}$ des paramètres positifs

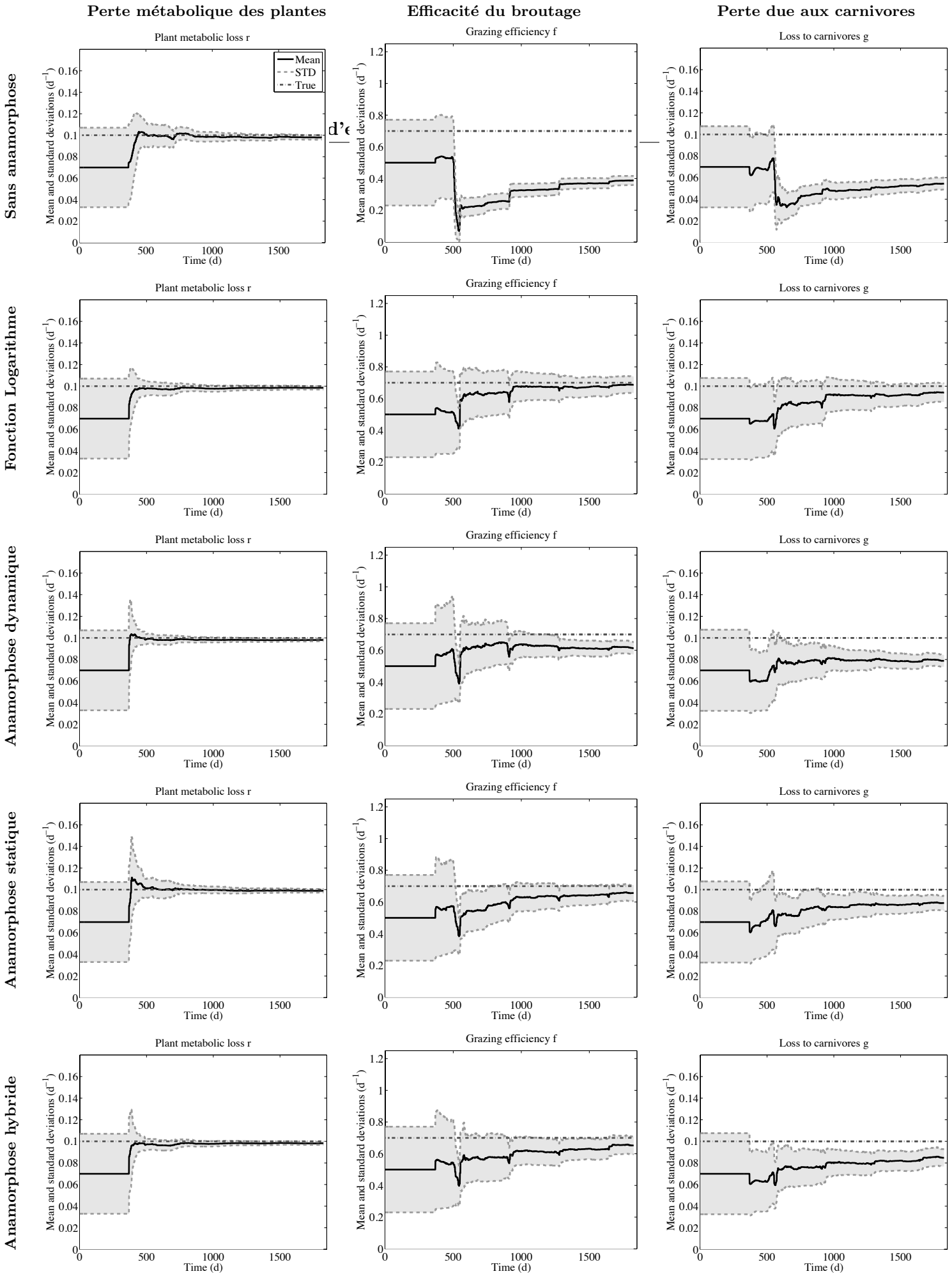



FIGURE 2.3 – Evolution temporelle de la moyenne (courbe noire) et de la moyenne plus/moins un écart type (zone grise) des paramètres estimés. Elles correspondent aux moyennes de ces quantités calculées sur 20 expériences. La valeur "vraie" correspond à la courbe noire pointillée. Figure tirée de Simon et Bertino (2012) [127].

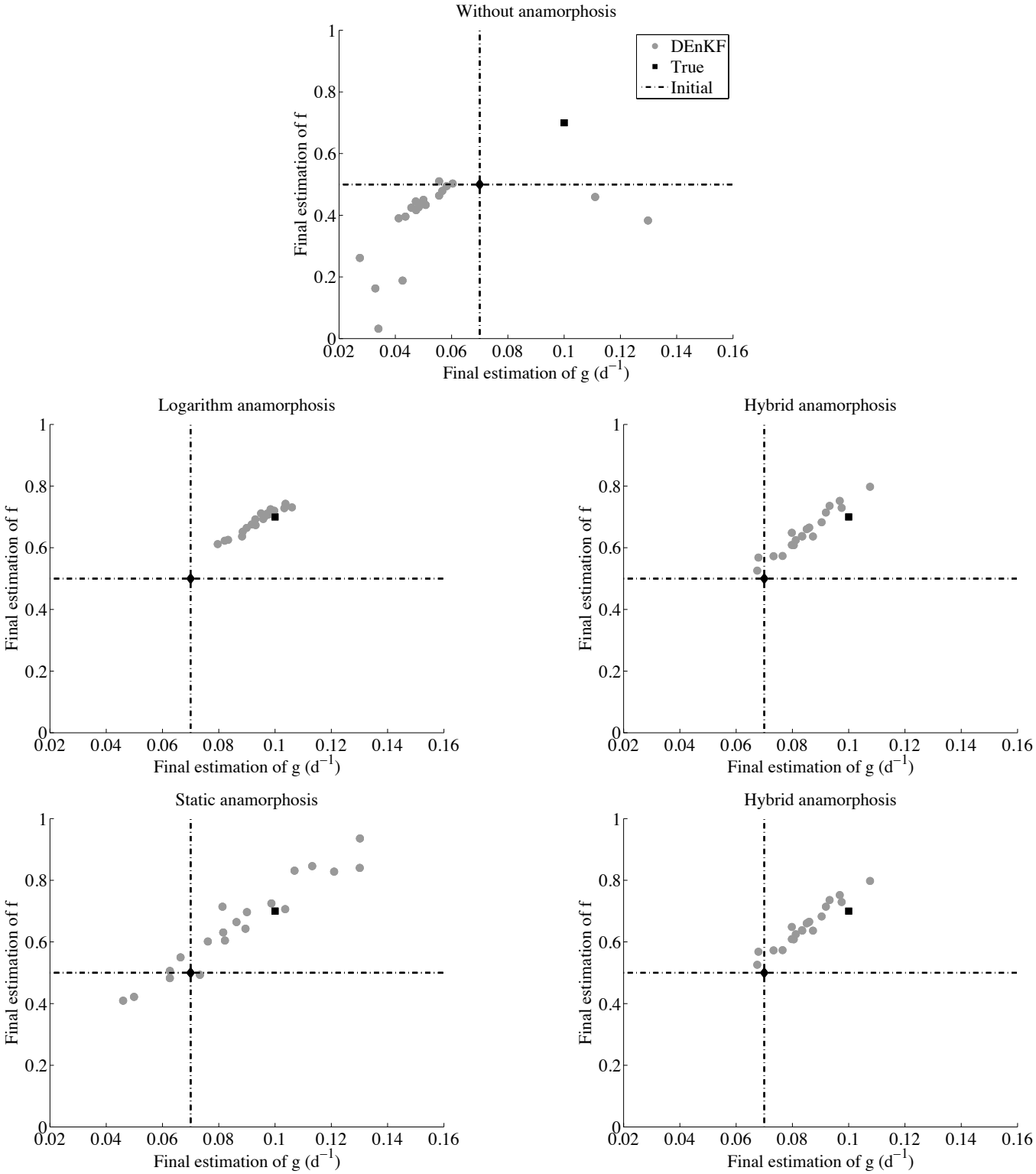



FIGURE 2.4 – Nuage de points de l'estimation en fin de simulation de la moyenne de l'ensemble du taux d'efficacité de broutage (f) versus taux de perte due aux carnivores (g) pour les 20 expériences réalisées. Les valeurs obtenues par assimilation correspondent aux cercles gris, les valeurs "vraies" des paramètres au carré noir, et le moyenne du jeu de paramètre initial à l'intersection des deux droites pointillées. Figure tirée de Simon et Bertino (2012) [127]

de somme égale à 1, à savoir :

$$\forall i = 1 : N, \pi_i \geq 0 \text{ et } \sum_{i=1}^{N} \pi_i = 1. \tag{2.12}$$

Pour ce problème, la conservation des propriétés linéaires intrinsèque au filtre de Kalman d'ensemble [47], et notamment la contrainte de somme, n'est pas garantie pour ces paramètres en raison de la contrainte de positivité qui s'applique à eux. La troncature des valeurs négatives qui résulte de l'analyse de Kalman peut donc conduire à des paramètres estimés ne respectant pas la contrainte de somme. De la même manière, même si les extensions des filtres de Kalman par anamorphoses gaussiennes rendent possible l'estimation de paramètres positifs [127], les transformations non linéaires ne garantissent pas que leur somme soit toujours égale à un. Dans le cadre d'une collaboration avec L. Bertino, A. Samuelsen et D. Dumont [128], nous nous sommes intéressés à de nouvelles reformulations de ce problème, pour n'avoir que des paramètres positifs à estimer.

La première possibilité repose sur l'introduction de la loi de Dirichlet d'ordre $N$ pour les paramètres $(\pi_i)_{i=1:N}$, celle-ci engendrant des variables aléatoires satisfaisant naturellement les contraintes (2.12). Ceux-ci peuvent être obtenus depuis $N$ variables aléatoires indépendantes $(\phi_i)_{i=1:N}$ de loi Gamma de la manière suivante :

$$\forall i = 1 : N, \pi_i = \frac{\phi_i}{\sum_{k=1}^{N} \phi_k} \quad \text{with} \quad \phi_i \sim \Gamma(\theta_i, 1). \tag{2.13}$$

Il reste alors simplement à estimer les paramètres $(\phi_i)_{i=1:N}$ par filtrage de Kalman d'ensemble avec anamorphoses gaussiennes. Une fois ceux-ci obtenus, les paramètres originaux $(\phi_i)_{i=1:N}$ sont calculés depuis (2.13). Une autre approche similaire consiste à remplacer les lois Gamma par des lois log-Normale [60]. Le changement de variables s'écrit alors :

$$\forall i = 1 : N, \pi_i = \frac{\mathrm{e}^{\phi_i}}{\sum_{k=1}^{N} \mathrm{e}^{\phi_k}} \quad \text{with} \quad \phi_i \sim \mathcal{N}(\theta_i, \Sigma_i). \tag{2.14}$$

Pour cette variante, les paramètres $(\phi_i)_{i=1:N}$ suivent bien une loi gaussienne, et ne nécessitent donc pas leur transformation par anamorphose dans l'étape d'analyse du filtre de Kalman d'ensemble. Quelques premières remarques concernant ces deux approches.

1. De part le rôle symétrique jouer par les paramètres $(\phi_i)_{i=1:N}$ dans ces deux formulations, l'estimation des paramètres $(\pi_i)_{i=1:N}$ n'est pas sensible au choix de l'affectation des paramètres $(\phi_i)_{i=1:N}$ aux $(\pi_i)_{i=1:N}$ dans les changements de variables.
2. Ces deux approches ne permettent pas aux paramètres $(\pi_i)_{i=1:N}$ de s'annuler. Dans le cas de l'estimation de paramètres de préférences alimentaires, ceci implique qu'un type de nourriture ne peut jamais disparaître théoriquement du régime alimentaire par assimilation, même si en pratique cela peut-être le cas numériquement.

Afin de réduire le nombre de paramètres à estimer, et pouvoir théoriquement en annuler certains, nous avons proposé une nouvelle formulation du problème basée sur les coordonnées hypersphériques, qui généralisent les coordonnées sphériques à la dimension $N$. Considérant

les paramètres $(\pi_i)_{i=1:N}$ comme les coordonnées cartésiennes d'un vecteur $\pi$ de $\mathbb{R}^N$, une idée naturelle consiste à le représenter dans un autre système de coordonnées. Nous avons suggéré d'introduire $N-1$ angles $(\phi_i)_{i=1:N-1}$ pour représenter le vecteur $\pi$. Cette approche a par ailleurs été également introduite pour rendre implicite des contraintes de sommes pour d'autres applications géométriques [111]. Ceci conduit au système d'équations non-linéaires, liant les paramètres $(\pi_i)_{i=1:N}$ et $(\phi_i)_{i=1:N-1}$ :

$$\left\{\begin{array}{rcl} \pi_1 & = & \cos^2(\frac{\pi}{2}\phi_1) \\ \forall i = 2 : N-1, & & \\ \pi_i & = & \displaystyle\prod_{k=1}^{i-1} \sin^2(\frac{\pi}{2}\phi_k)\cos^2(\frac{\pi}{2}\phi_i) \\ \pi_N & = & \displaystyle\prod_{k=1}^{N-2} \sin^2(\frac{\pi}{2}\phi_k)\sin^2(\frac{\pi}{2}\phi_{N-1}) \end{array}\right. \tag{2.15}$$

avec $(\phi_i)_{i=1:N-1}$ $N-1$ variables aléatoires suivant une loi de probabilité à valeurs dans $[0,1]$. Ainsi définis, il est possible de montrer que les paramètres $(\pi_i)_{i=1:N}$ satisfont les contraintes (2.12). De nouveau, les paramètres $(\phi_i)_{i=1:N-1}$ peuvent être estimés en introduisant les fonctions d'anamorphoses durant l'étape d'analyse. Quelques remarques peuvent être formulées.

1. Cette approche permet de réduire le nombre de paramètres à estimer, passant de $N$ à $N-1$. Même si cela peut paraître modeste pour un jeu de paramètres $(\pi_i)_{i=1:N}$, cela devient intéressant dans le cas de modèle écosystèmes complexes, impliquant différents niveaux trophiques, avec différents types fonctionnels aux régimes alimentaires variés.
2. Contrairement aux deux approches précédentes, les valeurs obtenues pour les paramètres $(\pi_i)_{i=1:N}$ peuvent être sensibles aux choix de l'affection des paramètres $(\phi_i)_{i=1:N-1}$ aux $(\pi_i)_{i=1:N}$ dans le changement de variables.

Une fois défini ce changement de variables, le problème qui se pose est la définition de la loi de probabilité pour les paramètres $(\phi_i)_{i=1:N-1}$. Si la loi de probabilité des paramètres $(\pi_i)_{i=1:N}$ est connue, ou qu'un échantillon est disponible, il est alors possible de fournir une valeur a priori pour les paramètres $(\phi_i)_{i=1:N-1}$ par inversion des coordonnées hypersphériques, partant de $\phi_1 = \dfrac{2}{\pi}\arccos(\sqrt{\pi_1})$, et procédant par récurrence sur les paramètres. Dans le cas où ces informations seraient manquantes, nous avons proposé, non pas de fournir des lois de probabilité ad-hoc pour les paramètres $(\phi_i)_{i=1:N-1}$, mais plutôt d'être en mesure d'estimer les valeurs de leurs hyper-paramètres, une fois celles-ci imposées, afin que chacun des paramètres $(\pi_i)_{i=1:N}$ suive une moyenne et variance cible fixée a priori. Par exemple, dans le cas des préférences alimentaires d'un type fonctionnel de zooplancton, une fois imposée les lois de probabilités pour les paramètres $(\phi_i)_{i=1:N-1}$, nous aimerions pouvoir spécifier a priori les valeurs de leurs hyper-paramètres, afin de commencer le processus d'estimation des $(\pi_i)_{i=1:N}$, avec un ensemble pour lequel les paramètres auraient tous la même moyenne $\frac{1}{N}$ (aucune préférence particulière dans le régime alimentaire).

Le problème est le suivant. Supposons que les paramètres $(\phi_i)_{i=1:N-1}$ soient indépendants et de lois marginales $(\mathcal{D}_i(\mathbf{\Theta}_i))_{i=1:N-1}$. Selon le choix des lois $\mathcal{D}_i$, il est possible de spécifier a priori les valeurs des moyennes et variances des paramètres $(\pi_i)_{i=1:N}$ depuis les hyper-paramètres $(\mathbf{\Theta}_i)_{i=1:N-1}$. Ceci conduit à résoudre des systèmes d'équations non-linéaires décrits dans le Théorème 2.1.1.

**Theorem 2.1.1.** *Notons $(m_i)_{i=1:N}$ et $(\sigma_i^2)_{i=1:N}$ les moyennes et variances cibles des paramètres $(\pi_i)_{i=1:N}$, et $(\Phi_{\phi_i})_{i=1:N-1}$ les fonctions caractéristiques des paramètres $(\phi_i)_{i=1:N-1}$. Nous avons les propriétés suivantes :*

1. *Spécification des moyennes $(m_i)_{i=1:N-1}$*
   *Les hyper-paramètres $(\mathbf{\Theta}_i)_{i=1:N-1}$ sont solutions des $N-1$ équations non-linéaires*

$$\begin{cases} \frac{1}{4}(\Phi_{\phi_1}(\pi)+\Phi_{\phi_1}(-\pi)) &= m_1 - \frac{1}{2} \\ \forall i = 2:N-1, & \\ \frac{1}{4}(\Phi_{\phi_i}(\pi)+\Phi_{\phi_i}(-\pi)) &= \dfrac{m_i}{1-\sum_{k=1}^{i-1} m_k} - \frac{1}{2} \end{cases} \tag{2.16}$$

2. *Spécification des variances $(\sigma_i^2)_{i=1:N-1}$*
   *Les hyper-paramètres $(\mathbf{\Theta}_i)_{i=1:N-1}$ sont solutions des $N-1$ équations non-linéaires :*

$$\begin{cases} \frac{1}{16}(\Phi_{\phi_1}(2\pi)+\Phi_{\phi_1}(-2\pi)) = \\ \qquad -\frac{3}{8}+\sigma_1^2+m_1^2-\frac{1}{4}(\Phi_{\phi_1}(\pi)+\Phi_{\phi_1}(-\pi)) \\ \forall i = 2:N-1, \\ \frac{1}{16}(\Phi_{\phi_i}(2\pi)+\Phi_{\phi_i}(-2\pi)) = -\frac{3}{8}-\frac{1}{4}(\Phi_{\phi_i}(\pi)+\Phi_{\phi_i}(-\pi)) \\ +\dfrac{\sigma_i^2+m_i^2}{\sum_{k=1}^{i}(-2)^{k-1}(\sigma_{i-k}^2+m_{i-k}^2)\prod_{l=1}^{k-1}\frac{1}{4}(\Phi_{\phi_{i-l}}(\pi)+\Phi_{\phi_{i-l}}(-\pi))} \end{cases} \tag{2.17}$$

   *avec les conventions $\sigma_0^2+m_0^2=1$ et $\prod_{l=1}^{0}\frac{1}{4}(\Phi_{\phi_{i-l}}(\pi)+\Phi_{\phi_{i-l}}(-\pi))=1$.*
   *Les valeurs de $(\Phi_{\phi_i}(\pi)+\Phi_{\phi_i}(-\pi))_{i=1:N-1}$ dépendent uniquement des moyennes $(m_i)_{i=1:N-1}$ et sont données par (2.16).*

Selon le choix des lois marginales $(\mathcal{D}_i(\mathbf{\Theta}_i))_{i=1:N-1}$, les systèmes en question peuvent ne pas admettre de solution, ou si existence, nécessiter leur résolution numérique.

En guise d'exemple, supposons vouloir spécifier la même moyenne pour les paramètres $(\pi_i)_{i=1:N}$. La première étape consiste en le choix des lois marginales $(\mathcal{D}_i(\mathbf{\Theta}_i))_{i=1:N-1}$ pour les paramètres $(\phi_i)_{i=1:N-1}$. Nous supposons donc qu'ils suivent des lois triangulaires définies par :

$$\forall i = 1:N-1, \phi_i \sim \mathcal{T}(0,1,\theta_i). \tag{2.18}$$

avec $\theta_i \in [0,1]$ le mode de la loi $\mathcal{D}_i$. La fonction caractéristique $\Phi_{\phi_i}$ est définie par :

$$\forall i = 1:N-1, \quad \forall t \in \mathbb{R}, \quad \Phi_{\phi_i}(t) = -2\frac{(1-\theta_i)-e^{j\theta_i t}+\theta_i e^{jt}}{\pi^2\theta_i(1-\theta_i)}. \tag{2.19}$$

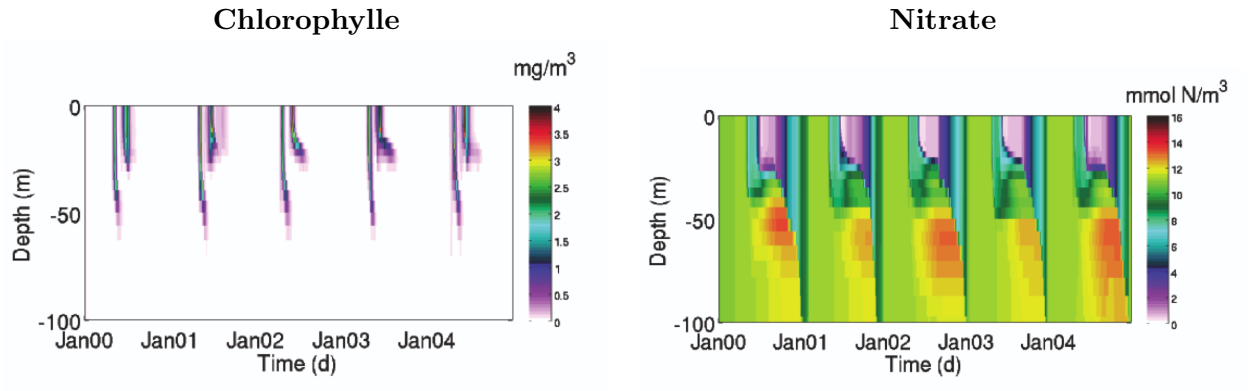



FIGURE 2.5 – Solution de référence : évolution temporelle de la chlorophylle et du nitrate, dans les 100 premiers mètres de la colonne d'eau, du 1er janvier 2000 au 31 décembre 2004. Figure tirée de Simon *et al.* (2012) [128].

avec $j^2 = -1$. Il nous reste alors à choisir les $N-1$ modes $(\theta_i)_{i=1:N-1}$ pour obtenir des moyennes égales :

$$\forall i = 1 : N, \quad m_i = \frac{1}{N}. \tag{2.20}$$

Le système d'équations (2.16) à résoudre s'écrit alors :

$$\begin{aligned} &\forall i = 1 : N-1, \\ &(\mathcal{S}_i) \quad \frac{\cos(\pi\theta_i) + 2\theta_i - 1}{\pi^2\theta_i(1-\theta_i)} + \frac{N-i-1}{2(N-i+1)} = 0. \end{aligned} \tag{2.21}$$

Il est possible de montrer que ce système d'équations admet une solution pour $N \leq 3$. En pratique, il ne sera donc pas possible d'utiliser la loi triangulaire, pour estimer quatre paramètres ou plus.

Nous avons évalué les performances de ces approches dans le cadre d'expériences jumelles dans un modèle d'écosystème marin représentatif d'une colonne d'eau de la Mer du Nord (station Mike). La dynamique de floraison du phytoplancton dans les 100 premiers mètres (simulation de référence) est illustrée sur la Figure 2.5. Le modèle inclue un type fonctionnel de microzooplanton, qui se nourrit de phytoplanctons (diatomées et flagellés) et de détritus, et un type fonctionnel de mésozooplancton, qui se nourrit de phytoplancton (diatomées), de microzooplancton et de détritus. L'objectif fut donc de voir si nos approches permettaient d'estimer correctement les paramètres de préférences alimentaires de ces deux types fonctionnels, et ceux par méthode de filtrage de Kalman d'ensemble. De nouveau, nous avons réalisé 20 expériences (différentes valeurs de paramètres initiaux, conditions initiales et observations), afin de présenter des résultats plus robustes face à l'aléa d'une unique configuration expérimentale.

La Figure 2.6 représente l'évolution temporelle de la moyenne (courbe noire) et de la moyenne plus/moins un écart type (zone grise) des paramètres de préférences alimentaires estimés par assimilation de données, pour la stratégie construite depuis les coordonnées hypersphériques. Nous avons supposé que les paramètres $(\phi_1, \phi_2)$, pour les deux types fonctionnels de zooplanctons, suivaient une loi triangulaire et avons choisi leurs hyper-paramètres pour obtenir des préférences

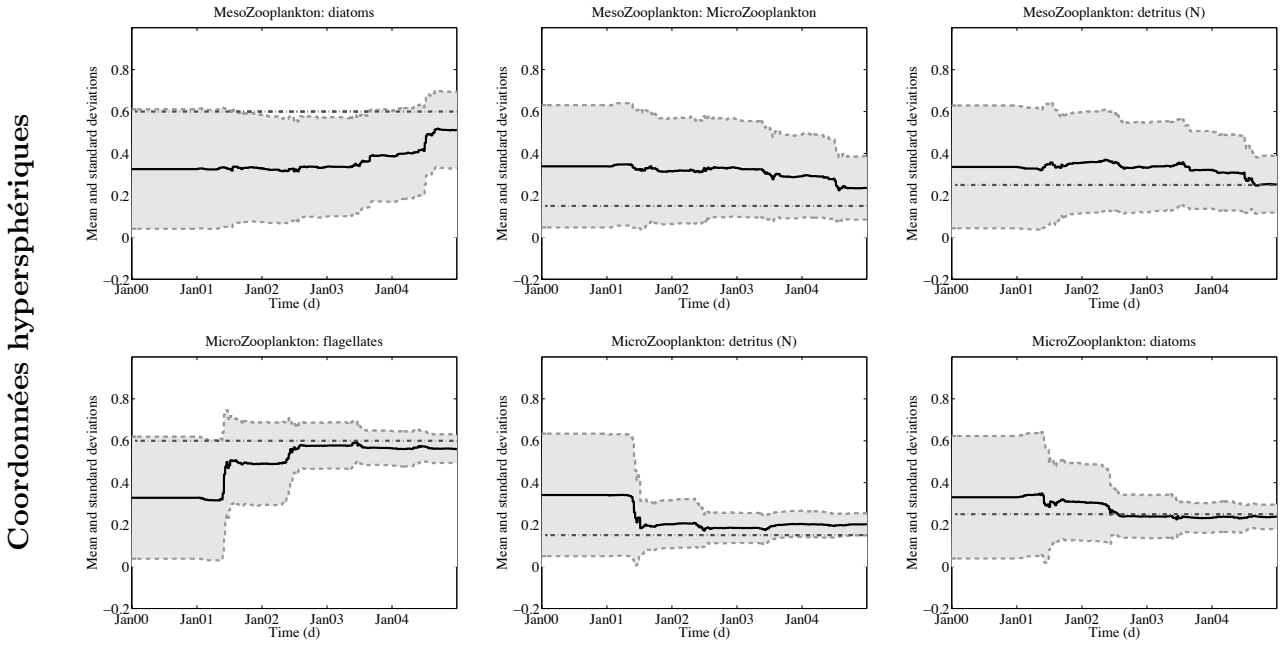



FIGURE 2.6 – Coordonnées hypersphériques : évolution temporelle de la moyenne (ligne noire)et de la moyenne plus/moins un écart type (zone grise) des préférences alimentaires, moyennées sur 20 expériences, pour les types fonctionnels mésozooplancton (haut) et microzooplacton (bas). Les valeurs vraies sont affichées en pointillés (constantes). Figure tirée de Simon *et al.* (2012) [128].

alimentaires égales, conduisant ainsi à la résolution numérique des systèmes (2.21). Nous notons de nouveau que les plus fortes corrections sur les paramètres ont lieu lors des périodes de floraison des phytoplanctons. Ce phénomène est particulièrement prononcé lors des premières années pour les préférences alimentaires du microzooplancton, type fonctionnel qui se nourrit de phytoplanctons, desquels sont déduites les concentrations de chlorophylle de surface qui sont assimilées. Les corrections sur les paramètres de préférences alimentaires du mésozooplancton sont moins marquées, les plus larges étant observée sur la dernière année, une fois obtenue la convergence du filtre sur les paramètres de préférences alimentaires du microzooplancton. En fin de simulation, nous obtenons des paramètres optimisés, qui en moyenne sur les expériences, semblent avoir été améliorés. A noter que ces résultats sont similaires à ceux obtenus pour les approches construites sur la distribution de Dirichlet.

Ceci est confirmé par le diagramme ternaire des estimés en fin de simulation (moyenne de l'ensemble) des préférences alimentaires, pour les 20 expériences réalisées, obtenus par assimilation de données affiché sur la Figure 2.7. Nous observons que les trois préférences alimentaires du microzooplancton ont été améliorées par assimilation de données sur 18 des 20 expériences réalisées, ce qui se traduit par la présence des cercles gris dans la zone de même couleur. Les résultats sont un peu moins bons pour les préférences alimentaires du mésozooplancton, avec une amélioration des trois paramètres pour 14 expériences. Néanmoins, les difficultés rencontrées pour estimer les préférences alimentaires du mésozooplancton nous ont semblé être liées au cadre expérimental, et notamment les variables observées (concentrations de chlorophylle de surface) et leur fréquence d'observation, plutôt qu'au choix des transformations de ces paramètres.

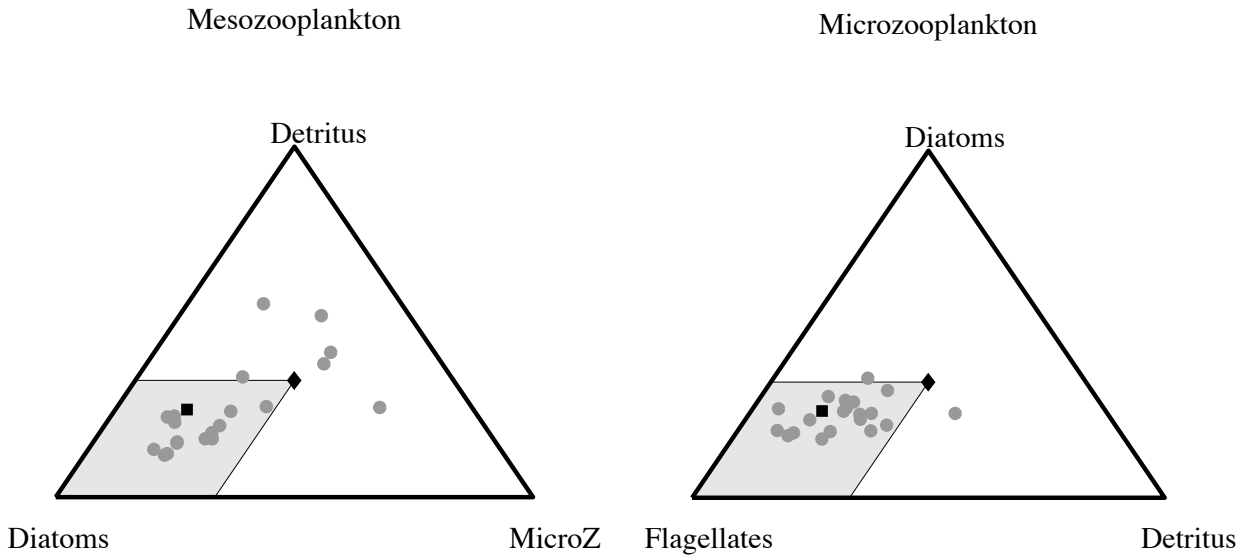



FIGURE 2.7 – Coordonnées hypersphériques : diagramme ternaire des estimés en fin de simulation (moyenne de l'ensemble) des préférences alimentaires pour les 20 expériences réalisées obtenus par assimilation de données (cercles gris). Les valeurs vraies sont affichées avec un carré noir, tandis que la moyenne des paramètres tirés a priori correspond aux diamants noirs. La zone grise signifie une amélioration des trois paramètres vis-à-vis du tirage a priori. Figure tirée de Simon *et al.* (2012) [128].

## 2.2 Application à la prévision quasi-opérationnelle en océanographie

Outre des développements méthodologiques sur les extensions non-gaussiennes des filtres de Kalman d'ensemble par anamorphose gaussienne, je me suis également intéressé à l'application de ces approches sur des configurations réalistes des composantes biogéochimiques des océans Atlantique Nord et Arctique, allant notamment jusqu'à la production de données de réanalyse dans le cadre de projets européens, ainsi que celle d'un océan global dans un modèle climatique du système Terre (cycle du carbone). Ceci m'a donc amené à modifier différents systèmes d'assimilation associés à l'océan physique et sa composante glace de mer, pour y inclure une composante biogéochimique, dans l'optique de réaliser l'estimation de variables d'état ou l'estimation jointe des variables d'état et paramètres du modèle. De surcroît, ces systèmes d'assimilation de données ensemblistes produisant une quantité importante de données au fil des cycles d'assimilation, je me suis donc intéressé à leur exploitation via des techniques d'apprentissage machine, notamment pour identifier des zones géographiques pertinentes (clusters) partageant une dynamique commune dans les données produites par assimilation.

### 2.2.1 Des systèmes d'assimilation en constante mutation

De part la nature même d'un système d'assimilation de données, celui-ci est naturellement amené à évoluer au fil du temps : inclusion de nouvelles d'observations, mise à jour du modèle numérique d'océan suite aux progrès de la modélisation, prise en compte de nouvelles com-

posantes dans le système couplé (par exemple la glace de mer, la biogéochimie, etc..), mise à jour de la méthode d'assimilation, intérêt pour de nouvelles applications, etc.. J'ai donc été amené à travailler sur différents systèmes d'assimilation, avec une part de développement logiciel conséquente, au cours de mes années au NERSC. J'en fais une description succincte dans ce qui suit, et les communications associées.

— **TOPAZ-ECO**
Dans un premier temps, j'ai réalisé un premier système permettant l'estimation conjointe des variables d'état et de certains paramètres du modèle écosystème par assimilation d'observations de surface de chlorophylle-a (satellitaires) dans le modèle couplé océan-écosystème HYCOM-NORWECOM via des filtres de Kalman d'ensemble avec anamorphoses gaussiennes. Il était basé sur la version 3 du système opérationnel d'assimilation de données TOPAZ, qui produit des prévisions à court terme, dans l'Atlantique Nord et dans l'Arctique, disponibles en accès libre et participe à différents projets internationaux relatifs à la surveillance, l'analyse et la prédiction des océans. Le passage à la version 4 du système TOPAZ a entrainé la réalisation d'un second système incluant également l'estimation des variables d'état du modèle physique. Dans le cadre du projet MyOcean, préliminaire à la création du Service Marin de Copernicus, les sources du module d'anamorphoses gaussiennes ont été transférées au Global Marine Forecasting Center (Mercator-Océan) dans l'optique de l'assimilation de données de glace de mer, et les sources du module permettant la conservation des quantités de traceurs biogéochimiques après corrections des variables physiques ont été transférées à l'Institut norvégien de météorologie (met.no). Ces systèmes sont à l'origine des communications [125, 126, 129], ainsi que la production de données de réanalyse de la biogéochimie marine de l'océan Arctique pour la période 2007-2010.

— **GOTM-NORWECOM**
J'ai développé un système d'assimilation de données, par filtrage de Kalman d'ensemble avec anamorphose gaussienne, générique aux différents modèles d'écosystèmes marins 1D (colonne d'eau) embarqués dans le modèle GOTM (General Ocean Turbulence Model, version de D. Dumont). Il correspond à différents modules Fortran90 implémentés dans GOTM, ainsi qu'à quelques modifications du code de base (introduction de l'assimilation de données, parallélisation MPI). Ce système se veut représentatif d'une colonne d'eau (un point de grille) du système TOPAZ-ECO et a pour objectif la validation des développements méthodologiques avant transfert dans le système "opérationnel". Il est à l'origine de la publication [128].

— **MICOM-HAMOCC (NorESM)**
J'ai également été impliqué dans le développement d'un système d'assimilation par méthodes d'ensemble pour la composante biogéochimique océanique (MICOM-HAMOCC) du Norwegian Earth System Model (NorESM). Ce système correspond à différents programmes Fortran90 (EnKF, module d'anamorphose) déjà implémentés par le passé dans TOPAZ-ECO et adaptés à la structure de NorESM, différentes modifications des codes des modèles MICOM et HAMOCC dans l'optique d'une estimation de paramètres, et différents scripts shell. Ce système avait pour but une meilleure représentation du cycle du carbone dans la composante océanique de NorESM, via l'assimilation de données de chlorophylle de surface et de pCO2. Certains résultats de ce système sont présentés dans [59].

### 2.2.2 Des données produites massivement et à valoriser

Dans le cadre des projets européens MyOcean et MyOcean2, prémices au développement du Service Marin de Copernicus, relatifs à la surveillance, l'analyse et la prédiction des océans, j'ai été amené à produire des données de réanalyse (analyse historique d'une phénomène océanique ou climatique) de la biogéochimie de l'océan Arctique sur la période 2007-2010. Ce type de données de réanalyse peut-être intéressant pour étudier l'évolution du système sur une longue période, en tenant compte des informations fournies par le modèle et les observations assimilées. D'un point de vue technique, ces données, au format netcdf, contiennent les moyennes mensuelles de différentes variables et paramètres biogéochimiques du modèle couplé océan-écosystème HYCOM-NORWECOM et furent disponibles en accès libre sur le portail web de Copernicus. Le NERSC était responsable du centre de prévision pour l'océan Arctique ("Arctic Marine Forecasting Center"). Ces travaux ont nécessité la réalisation de la version 4 du système d'assimilation de données TOPAZ-ECO mentionné précédemment.

Dans un premier temps, l'objectif de ses travaux fût d'étudier la capacité d'un système d'assimilation par filtrage de Kalman d'ensemble pour l'estimation jointe variables d'état et paramètres du modèle sur de longues périodes. Le système a donc évolué selon trois phases : l'année 2007 correspond à l'assimilation de données pour les composantes océan physique et glace de mer, la composante biogéochimique s'ajustant aux variations issues de ces corrections, puis les données de chlorophylle de surface furent assimilées à partir du 1er janvier 2008, avec estimation de paramètres du modèle biogéochimique pour les années 2008 et 2009, et l'utilisation de séries temporelles de ces mêmes paramètres optimisés au lieu de leur estimation, pour l'année 2010. Cette dernière année visait à évaluer l'intérêt d'utiliser des paramètres optimisés dans le cadre d'une estimation des variables d'état seulement. La Figure 2.8 montre l'évolution temporelle de l'erreur RMS pour la moyenne mensuelle de la concentration de chlorophylle de surface dans l'océan Arctique (40°-90°N). La courbe bleu correspond à l'erreur d'une simulation déterministe, tandis que la courbe rouge l'erreur de l'ensemble d'analyse. La courbe verte représente le pourcentage de points de grille du domaine couvert par les observations satellitaires. Les zones bleues illustrent les périodes pour lesquelles l'erreur RMS de la solution issue du système d'assimilation est meilleure que celle associée à la solution déterministe, et vice-versa pour les zones rouges.

Un point important à mentionner est qu'une très petite fraction du domaine arctique, pour sa composante biogéochimique, est observée en hiver principalement située près de la frontière sud. Cela signifie que l'impact d'une meilleure représentation de la couverture de glace sur la distribution de la chlorophylle en surface, grâce à l'assimilation de données physiques, ne peut pas être évalué avec les données satellitaires de couleur de l'océan pendant les périodes hivernales. Même si l'assimilation de données et les simulations libres présentent des erreurs spatiales moyennes similaires, de grandes différences dans la localisation de la lisière de glace entre les deux solutions entraînent des différences significatives dans la distribution de chlorophylle dans la partie non observée du domaine. Deuxièmement, nous notons que les deux erreurs évoluent de manière similaire en 2007 quand aucune observation de chlorophylle n'est assimilée. L'assimilation des données physiques entraîne une légère amélioration en juin, tandis que des erreurs plus importantes sont observées de juillet à octobre en raison des concentrations de chlorophylle plus importantes dans la solution issue de l'assimilation de données physiques. Néanmoins, l'assimilation des données de chlorophylle de surface entraîne une forte diminution des erreurs en 2008 et 2009. Le pic d'erreur le plus important (juin) est désormais trois fois moins élevé que celui de la simulation libre. Cependant, nous notons que l'erreur augmente fortement en 2010 lorsque les paramètres ne sont plus estimés. Ceci est à lier avec une moindre qualité des ob-

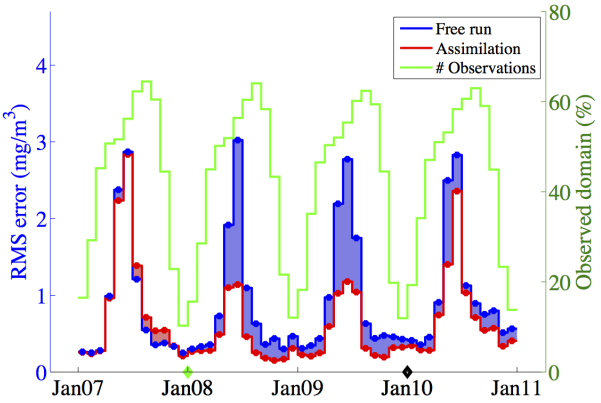



FIGURE 2.8 – Moyenne mensuelle de la concentration de chlorophylle de surface : évolution temporelle de l'erreur RMS dans l'océan Arctique. La courbe bleu correspond à l'erreur d'une simulation déterministe, tandis que la courbe rouge l'erreur de l'ensemble d'analyse. La courbe verte représente le pourcentage de points de grille du domaine couvert par les observations satellitaires. En abscisse, le diamant vert indique la date de début de l'assimilation des données biogéochimiques, et le diamant noir la date de fin d'estimation de certains paramètres du modèle biogéochimique. Figure tirée de Simon *et al.* (2015) [129]

servations assimilées sur cette période (entraînant une sous-estimation de l'erreur d'observation dans le filtre), des événements climatiques exceptionnels (forte oscillation nord-atlantique cette année là), et les changements de stratégie pour améliorer la croissance de l'erreur lors des étapes de prévision (pas de paramètres biogéochimiques aléatoires). En l'état actuel des modèles de biogéochimie de l'océan et des observations dont nous disposons, l'obtention de paramètres pertinents, pour l'océanographie opérationnelle, par assimilation de données reste un problème difficile. En effet, contrairement aux cadres expérimentaux classiques, pour lesquels l'estimation de paramètres du modèle convergent bien vers des constantes, nous observons dans le cas d'applications réalistes une variation souvent saisonnière de ces mêmes paramètres, comme illustrée sur la Figure 2.9 pour le taux de mortalité du microzooplancton. Ceci suggère que les corrections sur les paramètres tendent à être associées à des erreurs multiples dans le système (erreur modèle non liée aux paramètres, biais, etc..), rendant le processus d'estimation plus complexe. Dans ces conditions, il semble nécessaire de maintenir le processus d'estimation des paramètres, de manière jointe avec l'estimation des variables d'état, plutôt que d'utiliser un jeu de paramètres optimisés.

Quand bien même les paramètres obtenus après assimilation pourraient ne pas avoir de grand intérêt pour un usage prédictif, les variations de ceux-ci sur une fenêtre suffisamment longue peuvent présenter des intérêts quant à l'analyse de leur dynamique. A mon arrivée à l'IRIT, nous nous sommes intéressés avec S. Mouysset, à la définition de provinces géographiques depuis l'évolution spatio-temporelle des paramètres estimés lors de la réanalyse, et nous les avons comparées à celles définies dans [95], se basant sur des observations uniquement. Devant tenir compte d'informations à la fois géographiques et biogéochimiques, l'analyse en cluster a été menée en deux étapes successives. La première consiste en l'application d'une méthode de K-means [99] sur les paramètres optimisés pour générer une partition de ceux-ci pour les océans

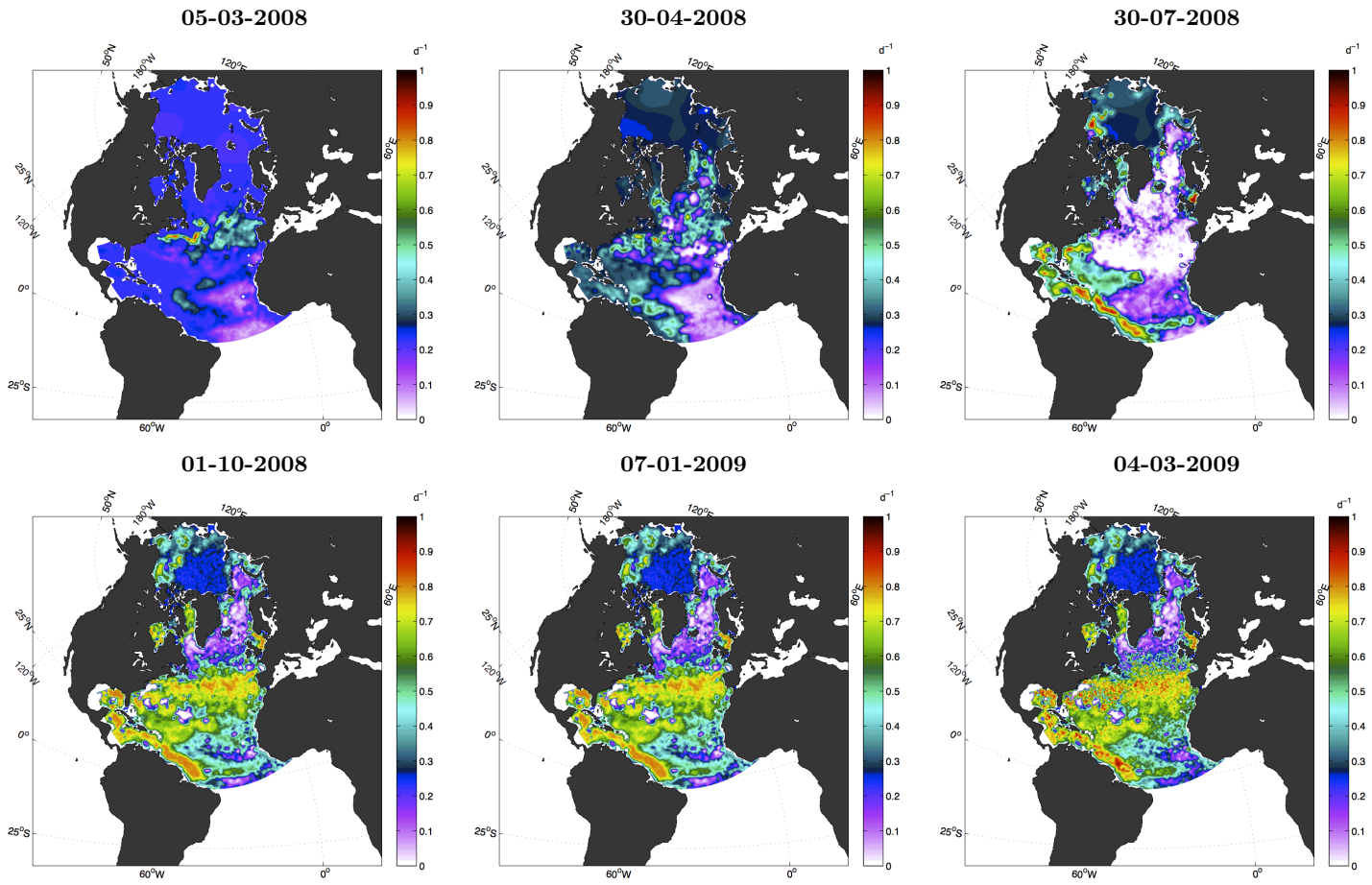



FIGURE 2.9 – Estimation de paramètres : distribution spatiale du taux de mortalité du micro-zooplancton (moyenne de l'ensemble) à différentes dates. Figure tirée de Simon *et al.* (2015) [129].

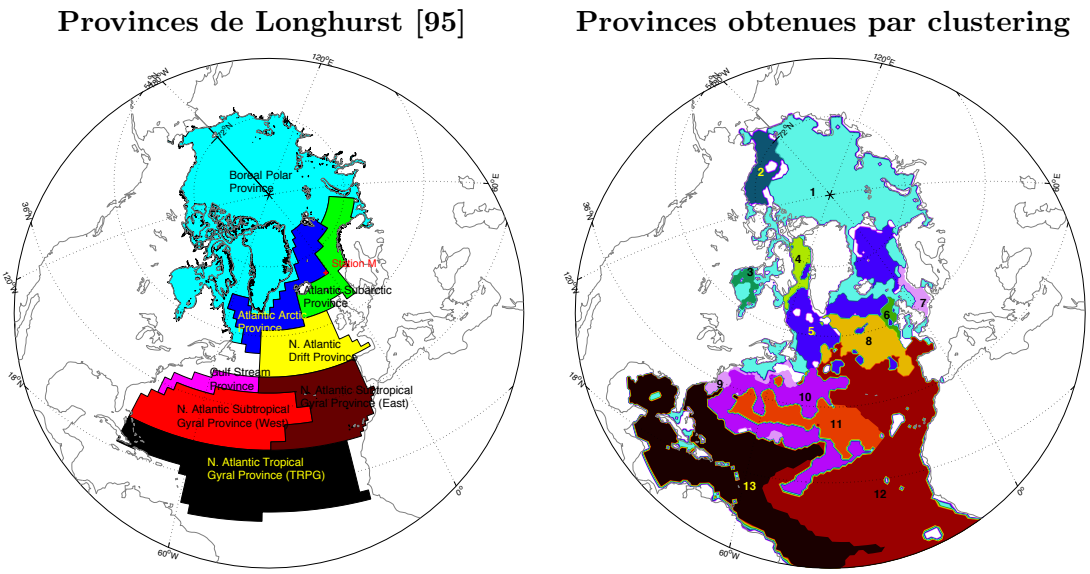



FIGURE 2.10 – Définition de provinces d'intérêt depuis Longhurst [95] et par analyse en cluster depuis l'évolution des paramètres du modèle au cours du processus d'estimation. Chaque province se voit affectée une couleur particulière et un nom / numéro. Figure tirée de Simon *et al.* (2015) [129].

Atlantique Nord et Arctique. L'idée principale est de définir $k$ centroïdes - un pour chaque cluster qui représente la moyenne des paramètres estimés assignés dans le même cluster, le nombre de clusters $k$ étant défini a priori par une méthode ad-hoc. Une classification spectrale [108] est ensuite appliquée afin de définir des clusters géométriquement séparés. L'approche consiste à sélectionner les vecteurs propres dominants d'une matrice d'affinité afin de construire un espace de données de faible dimension dans lequel les données géométriques sont regroupées en clusters [101]. Ceci conduit à une partition des océans en clusters représentatifs d'une même classe d'évolution des paramètres. Les résultats de cette approche, à savoir 13 zones géographiques, ainsi que huit provinces définies depuis Longhurst sont présentées sur la Figure 2.10. Le cluster 1, représenté en cyan dans l'analyse en cluster, correspond aux zones géographiques pour lesquelles les paramètres n'ont pas ou peu subi de corrections (absence d'observation sous la glace de mer ou proche des côtes).

Tout d'abord, nous pouvons noter que les deux analyses régionales sont consistantes dans la partie Nord du domaine. Mêmes si les provinces ne coïncident pas exactement en terme de couverture géographique, nous retrouvons visuellement un certain nombre de provinces de Longhurst depuis l'analyse en cluster. Néanmoins, des différences importantes apparaissent également. Tout d'abord la province, dite Boréale, est maintenant divisée en quatre clusters : la mer de Beaufort (cluster 2), les baies d'Hudson (cluster 3) et de Baffin (cluster 4) et la partie ne présentant pas ou peu d'observations (cluster 1). Ces différences peuvent être attribuées à plusieurs sources d'erreur modèle (par exemple à des forçages en nutriments erronés au Détroit de Bering et pour la rivière Hudson), mais peuvent également suggérer que l'océan Arctique présente des écosystèmes variés, justifiant une division de la province Boréale. Néanmoins, des travaux supplémentaires sont nécessaires, notamment via l'analyse de données in-situ, afin de mieux cerner la pertinence de ces nouvelles provinces. Enfin, certaines différences visibles dans le Sud du domaine sont liées

à l'erreur modèle, celui-ci ayant été développé pour les hautes latitudes, se révèle peu pertinent proche des Tropiques. Il est à noter que ceci est un exemple de valorisation des données massives produites par des systèmes d'assimilation de données opérationnels. L'essor des approches par apprentissage machine ouvre un large spectre pour leur exploitation.

## Publications associées

[125] E. Simon, L. Bertino : Application of the Gaussian anamorphosis to assimilation in a 3D coupled physical-ecosystem model of the North Atlantic with the EnKF : a twin experiment, *Ocean Sci.*, 5, 495-510, 2009 ;

[126] E. Simon, L. Bertino : Joint state-parameter estimation in a 3D coupled physical-ecosystem model of the North Atlantic : assimilation of SeaWiFS data with a non-Gaussian extension of an ESRF, *Proceedings of ESA Living Planet Symposium*, Bergen, 2010 ;

[127] E. Simon, L. Bertino : Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation : application to a 1D ocean ecosystem model, *J. Mar. Sys.*, 89, 1-18, 2012 ;

[128] E. Simon, A. Samuelsen, L. Bertino, D. Dumont : Estimation of positive sum-to-one constrained zooplankton grazing preferences with the DEnKF : a twin experiment, *Ocean Sci.*, 8, 587-602, 2012 ;

[59] M. Gehlen, R. Barciela, L. Bertino, P. Brasseur, M. Butenschöon, F. Chai, A. Crise, Y. Drillet, D. Ford, D. Lavoie, P. Lehodey, C. Perruche, A. Samuelsen, E. Simon : Building the capacity for forecasting marine biogeochemistry and ecosystems : recent advances and future developments, *J. Operat. Oceanogr.*, 8 (S1), s168-s187, 2015 ;

[129] E. Simon, A. Samuelsen, L. Bertino, S. Mouysset : Experiences in multiyear combined state-parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using the Ensemble Kalman Filter, *J. Mar. Sys.*, 152, 1-17, 2015 ;

[61] M.E. Gharamti, A. Samuelsen, L. Bertino, E. Simon, A. Korosov, U. Daewel : Online Tuning of Ocean Biogeochemical Parameters using Ensemble Estimation Techniques : Application to a one-dimensional Model in the North Atlantic, *J. Mar. Sys.*, 168, 1-16, 2017.

[109] T.H. Nguyen, S. Ricci, A. Piacentini, E. Simon, R. Rodriguez-Suquet, S. Peña-Luque : Gaussian anamorphosis for ensemble Kalman filter analysis of SAR-derived wet surface ratio observations, soumis.

# Chapitre 3

# Des problèmes d'estimation en très grande dimension : de la possibilité de réduire le volume des données

**Sommaire**


**3.1 Au niveau des observations . . . . . 29**
3.1.1 Une approche multi-niveaux dans l'espace des observations . . . . . 30
3.1.2 De la sensibilité de l'incrément d'analyse pour la sélection des observations à assimiler . . . . . 36
**3.2 Au niveau du modèle . . . . . 39**
3.2.1 Approche par réduction de modèles ou apprentissage profond . . . . . 39
3.2.2 Methodes d'ensemble . . . . . 42


*Ce chapitre porte sur des axes de recherche plus récents, et renvoie à des problématiques liées au volume de données, et donc les coûts et temps de calcul associés, que nous sommes amenés à rencontrer dans le cadre d'applications réalistes issues de l'industrie ou des géosciences. Ces travaux ont été réalisés, soit dans le cadre de collaborations internationales avec des chercheurs expérimentés, soit dans le cadre de deux thèses de doctorat en lien avec des partenaires industriels (convention CIFRE avec EDF, financement de l'ONERA sur des problématiques industrielles). Il est notamment présenté quelques travaux n'étant pas en lien direct avec les problématiques d'assimilation de données, mais offrant quelques applications/perspectives à celles-ci.*

## 3.1 Au niveau des observations

Une première source d'information qu'exploitent les méthodes d'assimilation de données réside dans les observations du système dont nous disposons. Celles-ci, de nature variée et incertaine, peuvent présenter une distribution spatio-temporelle hétérogène selon l'application considérée. Dans le cadre de l'océanographie, nous disposons d'un volume de données satellite et in-situ conséquent. Ainsi, rien que pour le système TOPAZ4 dédié aux océans Atlantique Nord et Arctique [121], de l'ordre de $10^6$ observations étaient disponibles pour la composante "physique" du système en 2012, sans tenir compte des observations de la biogéochimie marine, si le système

d'assimilation bénéficie d'une telle composante. Même si ceci constitue une volume de données conséquent, certaines zones restent très peu observées. Ainsi, le déploiement de données in-situ permettant une couverture fine de l'intérieur de l'océan reste difficile, certains instruments embarqués sur satellite sont inopérants de nuit (données de biogéochimie) et donc l'hiver dans les hautes latitudes, etc.. Quand bien même le système reste sous-observé, au sens où la dimension du problème est beaucoup plus grande que le nombre d'observations disponibles, il n'est pas rare que certaines zones géographiques et/ou variables présentent une forte densité d'observations quand d'autres restent très peu observées. Ainsi, disposant d'un réseau d'observations tel qu'il a été construit, est-il possible de réduire les coûts et temps de calcul des étapes d'analyse du processus d'assimilation, via la définition de stratégies de sélection des observations à assimiler, visant à réduire leur nombre sans compromettre la qualité de l'analyse ?

### 3.1.1 Une approche multi-niveaux dans l'espace des observations

A mon arrivée dans l'équipe APO, je me suis intéressé à des travaux entamés par S. Gratton, M. Rincon-Camacho et Ph. Toint sur la possibilité de faire décroître les coûts de calcul de l'algorithme 4D-Var incrémental, en incorporant une stratégie de sélection des observations à assimiler, directement dans le processus d'optimisation. L'approche proposée exploite la formulation duale de l'algorithme 4D-Var incrémental depuis la méthode *Restricted Preconditioned Conjugate Gradient* (RPCG ; [75]). Pour chaque itération externe de l'algorithme 4D-var incrémental, celle-ci opère dans l'espace des observations, afin d'exploiter le faible nombre d'observations comparativement à la dimension du vecteur d'état, tout en garantissant une décroissance monotone de la quadratique (définie dans l'espace d'état) au cours de ses itérations. L'objectif des travaux [69] fut de proposer une stratégie adaptative de sélection des observations au cours des itérations externes de l'algorithme 4D-var incrémental, exprimé dans sa formulation duale. A l'instar des approches multi-niveaux, celle-ci exploite la construction d'une hiérarchie emboîtée d'ensembles d'observations, construite depuis l'ensemble de toutes les observations, et adaptée au problème d'optimisation.

Supposons ainsi disposer d'un ensemble de $m$ observations, noté $\mathcal{O}$, décomposable en une hiérarchie de sous-ensembles d'observations $\{\mathcal{O}_i\}_{i=0}^r$, de cardinal $m_i$ pour $i$ fixé, et tel que

$$\mathcal{O}_i \subset \mathcal{O}_{i+1} \text{ et } m_i < m_{i+1} \quad (i = 0, \ldots, r-1)$$

avec par convention, $\mathcal{O}_r = \mathcal{O}$ et $m_r = m$. Pour chaque sous-ensemble $\mathcal{O}_i$, nous obtenons un vecteur d'observations $\mathbf{y}_i \in \mathbb{R}^{m_i}$, ainsi qu'une matrice de covariance d'erreur d'observation $\mathbf{R}_i$. Il est alors possible de définir une hiérarchie de problèmes d'optimisation depuis $\{\mathcal{O}_i\}_{i=0}^r$ :

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{x}_b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathcal{H}_i(\mathbf{x})-\mathbf{y}_i\|^2_{\mathbf{R}_i^{-1}}, \quad i = 0, \ldots, r, \tag{3.1}$$

avec $\mathcal{H}_i$ l'opérateur d'observation associé à l'ensemble $\mathcal{O}_i$.

La construction de cette hiérarchie d'ensembles d'observations est basée sur la définition d'une borne a posteriori sur les erreurs entre les solutions des problèmes (3.1) pour deux niveaux consécutifs d'ensembles d'observations $\mathcal{O}_i$ et $\mathcal{O}_{i+1}$. Pour simplifier les notations, nous notons $\mathcal{O}_c$ un sous-ensemble contenant $m_c$ observations et $\mathcal{O}_f$ un second sous-ensemble contenant $m_f$ observations et tel que $m_c < m_f$ et $\mathcal{O}_c \subset \mathcal{O}_f$. Le problème d'optimisation, dit "fin", avec pour point de départ $\mathbf{x}$, s'écrit :

$$\min_{\delta\mathbf{x}_f\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}+\delta\mathbf{x}_f-\mathbf{x}_b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathbf{H}_f\delta\mathbf{x}-\mathbf{d}_f\|^2_{\mathbf{R}_f^{-1}} \tag{3.2}$$

avec $\mathbf{d}_f = \mathbf{y}_f - \mathcal{H}_f(\mathbf{x})$ et $\mathbf{H}_f$ le modèle linéaire tangent de $\mathcal{H}_f$ en $\mathbf{x}$ et associé au sous-ensemble d'observations $\mathcal{O}_f$. En reformulant ce problème comme un problème d'optimisation quadratique convexe, avec contraintes d'égalité linéaires, la condition d'optimalité conduit au système

$$\begin{array}{rcl} (\mathbf{H}_f\mathbf{B}\mathbf{H}_f^T + \mathbf{R}_f)\lambda_f & = & \mathbf{d}_f - \mathbf{H}_f(\mathbf{x}_b - \mathbf{x}) \\ \delta\mathbf{x}_f & = & \mathbf{B}\mathbf{H}_f^T\lambda_f + (\mathbf{x}_b - \mathbf{x}) \end{array} \tag{3.3}$$

Le problème d'optimisation, dit grossier, s'écrit

$$\min_{\delta\mathbf{x}_c \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{x} + \delta\mathbf{x}_c - \mathbf{x}_b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathbf{\Gamma}_f(\mathbf{H}_f\delta\mathbf{x}_c - \mathbf{d}_f)\|^2_{\mathbf{R}_c^{-1}}, \tag{3.4}$$

avec $\mathbf{\Gamma}_f$ un opérateur de restriction $\mathbf{\Gamma}_f : \mathbb{R}^{m_f} \to \mathbb{R}^{m_c}$ de l'ensemble fin vers l'ensemble grossier. Notons $\mathbf{\Pi}_c$ un opérateur de prolongation de l'ensemble grossier vers l'ensemble fin satisfaisant

$$\mathbf{\Pi}_c \stackrel{\text{def}}{=} \sigma_f\mathbf{\Gamma}_f^T \tag{3.5}$$

avec $\sigma_f > 0$. Il est alors possible d'écrire un problème d'optimisation grossier de la manière suivante :

$$\min_{\delta\mathbf{x}_c \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{x} + \delta\mathbf{x}_c - \mathbf{x}_b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathbf{\Pi}_c^T(\mathbf{H}_f\delta\mathbf{x}_c - \mathbf{d}_f)\|^2_{\bar{\mathbf{R}}_c^{-1}}, \tag{3.6}$$

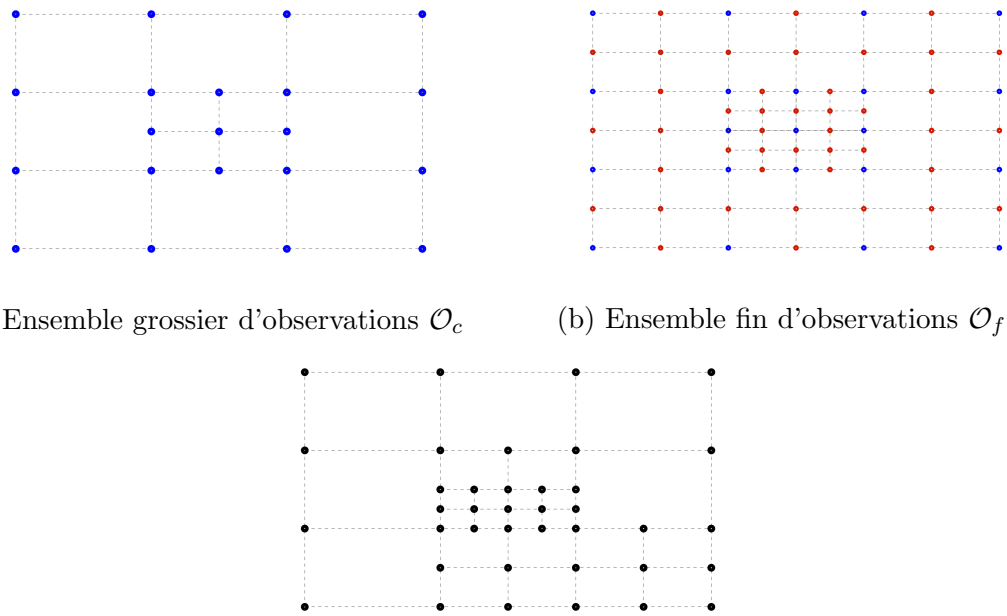
avec $\bar{\mathbf{R}}_c^{-1} = (\frac{1}{\sigma_f})^2\mathbf{R}_c^{-1}$. En reformulant ce problème comme un problème quadratique convexe avec des contraintes d'égalité, il vient

$$\begin{array}{rcl} (\bar{\mathbf{R}}_c^{-1}\mathbf{\Pi}_c^T\mathbf{H}_f\mathbf{B}\mathbf{H}_f^T\mathbf{\Pi}_c + \mathbf{I}_{m_c})\lambda_c & = & \bar{\mathbf{R}}_c^{-1}\mathbf{\Pi}_c^T(\mathbf{d}_f - \mathbf{H}_f(\mathbf{x}_b - \mathbf{x})) \\ \delta\mathbf{x}_c & = & \mathbf{B}\mathbf{H}_f^T\mathbf{\Pi}_c\lambda_c + (\mathbf{x}_b - \mathbf{x}) \end{array} \tag{3.7}$$

Réciproquement, ayant obtenu $(\delta\mathbf{x}_c, \lambda_c)$ solution de (3.7), est-il possible de choisir "judicieusement" des observations dans $\mathcal{O}_f$ pour construire un ensemble d'observations $\tilde{\mathcal{O}}_f$, avec $\mathcal{O}_c \subset \tilde{\mathcal{O}}_f \subset \mathcal{O}_f$, et tel que la solution issue de l'assimilation de ces données soit pertinente pour le problème (3.3), et ce de manière a priori ? En illustration de ce problème, supposons disposer d'observations "vivant" dans le plan, et que l'ensemble grossier des observations $\mathcal{O}_c$ renvoie à la grille composite présentée à la Figure 3.1 (a), et celui de l'ensemble fin $\mathcal{O}_f$ à la Figure 3.1 (b). L'objectif est alors de construire l'ensemble auxiliaire $\tilde{\mathcal{O}}_f$, présenté à la Figure 3.1 (c), en sélectionnant des observations dans $\mathcal{O}_f$, depuis les résultats de l'assimilation des données de $\mathcal{O}_c$. Ces observations seraient alors assimilées en vu de fournir à la fois un nouvel incrément, et les informations nécessaires pour la sélection d'un nouvel ensemble d'observations, qui seront assimilées à l'itération suivante.

Nous avons proposé pour cela une stratégie basée sur la norme de la différence entre l'hypothétique $\lambda_f$, qui est inconnu, et la quantité $\mathbf{\Pi}_c\lambda_c$. Le choix de construire un critère de sélection des observations depuis les différences entre les multiplicateurs de Lagrange associées aux problèmes (3.3) et (3.7) est motivé par les résultats suivants.

1. D'une part, les multiplicateurs de Lagrange fournissent de l'information sur la variation de la fonctionnelle à minimiser lorsqu'une perturbation $\epsilon$ intervient dans le terme de droite de la contrainte, à savoir dans notre cas, des changements dans les valeurs des observations ou du réseau d'observations. Pour une petite perturbation $\epsilon$, nous avons l'égalité suivante [110, Chapitre 12] :
$$J(\mathbf{x}_\epsilon) = J(\mathbf{x}) - \lambda^T\epsilon + \mathcal{O}(\|\epsilon\|^2).$$



(a) Ensemble grossier d'observations $\mathcal{O}_c$ (b) Ensemble fin d'observations $\mathcal{O}_f$

(c) Ensemble d'observations auxiliaires $\tilde{\mathcal{O}}_f$

FIGURE 3.1 – Définition d'un ensemble fin d'observations depuis un ensemble plus grossier. Figure tirée de Gratton *et al.* (2015). [69]

2. D'autre part, les multiplicateurs de Lagrange sont directement liés aux solutions du problème primal, via l'opérateur $\mathbf{BH}^T$ :

$$\delta\mathbf{x} = \mathbf{x}^b - \mathbf{x} + \mathbf{BH}^T\lambda.$$

De fait, nous obtenons

$$\|\delta\mathbf{x} - \delta\tilde{\mathbf{x}}\|_{\mathbf{M}} = \|\lambda - \tilde{\lambda}\|_{\mathbf{HBMBH}^T},$$

avec respectivement $\delta\mathbf{x}$ et $\delta\tilde{\mathbf{x}}$ la solution du problème primal et une approximation, $\lambda$ et $\tilde{\lambda}$ les multiplicateurs solutions du problème primal et une approximation, et $\mathbf{M}$ une matrice symétrique positive définie. Comparer les multiplicateurs de Lagrange dans une norme est donc équivalent à comparer les incréments d'analyse dans une autre norme.

Nous avons démontré le Théorème 3.1.1, qui vise à borner la norme énergie de l'erreur sur les multiplicateurs de Lagrange obtenus en résolvant le problème (3.7) au lieu de (3.3), et ce par des quantités explicitement disponibles ou calculables lors de la résolution de (3.7).

**Theorem 3.1.1.** *Soient* $(\delta\mathbf{x}_f, \lambda_f)$ *la solution de (3.3) et* $(\delta\mathbf{x}_c, \lambda_c)$ *celle de (3.7). Une borne d'erreur a posteriori sur les multiplicateurs de Lagrange de ces deux problèmes s'écrit*

$$\|\lambda_f - \mathbf{\Pi}_c\lambda_c\|_{\mathbf{R}_f+\mathbf{H}_f\mathbf{BH}_f^T} \quad \leq \quad \|\mathbf{d}_f - \mathbf{H}_f\delta\mathbf{x}_c - \mathbf{R}_f\mathbf{\Pi}_c\lambda_c\|_{\mathbf{R}_f^{-1}} \tag{3.8}$$

Nous retrouvons bien que la borne (3.8) ainsi proposée ne dépend pas des quantités $\lambda_f$ et $\delta x_f$ inconnues lors de la résolution de (3.7).

Afin de sélectionner les observations, nous en avons décliné un indicateur local pour chacune des observations :

$$\eta_j \stackrel{\text{def}}{=} w_j \langle (\tilde{\mathbf{d}}_f - \tilde{\mathbf{H}}_f \delta \mathbf{x}_c - \tilde{\mathbf{R}}_f \tilde{\mathbf{\Pi}}_c \lambda_c)|_j, (\tilde{\mathbf{R}}_f^{-1}(\tilde{\mathbf{d}}_f - \tilde{\mathbf{H}}_f \delta \mathbf{x}_c - \tilde{\mathbf{R}}_f \tilde{\Pi}_c \lambda_c))|_j \rangle,$$

avec $\tilde{\mathbf{d}}_f$, $\tilde{\mathbf{H}}_f$, $\tilde{\mathbf{R}}_f$ associés aux observations $\tilde{O}_f$, et avec la notation $|_j$ pour indiquer la composante $j$ du vecteur. L'indicateur $\eta$ ainsi défini correspond au terme de droite de l'inégalité (3.8). Une fois cet indicateur calculé pour chaque observation, leur sélection s'opère selon le marquage Döfler [42]. Soit $\theta_1 \in (0, 1)$, un ensemble $\mathcal{S}_\eta$ est construit de telle manière que

$$\theta_1 \left( \sum_{j=1}^{p} \eta_j \right) \leq \sum_{k \in \mathcal{S}_\eta} \eta_k \tag{3.9}$$

avec $p$ le nombre d'observations dans $\tilde{\mathcal{O}}_f$. En pratique, $\mathcal{S}_\eta$ est construit depuis les observations non présentes dans $\mathcal{O}_c$, en incluant celle ayant la valeur de l'indicateur la plus élevée, et ce de manière récursive, jusqu'à satisfaire (3.9). L'ensemble fin des observations est ensuite obtenu en ajoutant les observations de $\mathcal{S}_\eta$ à l'ensemble grossier des observations :

$$\mathcal{O}_f \stackrel{\text{def}}{=} \mathcal{O}_c \cup \left( \bigcup_{k \in \mathcal{S}_\eta} o_k \right). \tag{3.10}$$

Partant d'un ensemble d'observations $\mathcal{O}_0$ de cardinal petit, les observations sont progressivement sélectionnées pour assimilation. Ceci conduit à l'algorithme 4D-Var incrémental adaptatif Algorithm 1.

L'utilisation de $\tilde{\lambda}_i$ au lieu de $\lambda_i$ à l'étape 4 est motivée par le fait que ce dernier est obtenu en résolvant le problème (3.11), qui est associé au problème grossier (3.4) et non pas (3.6). A convergence de l'algorithme RPCG, on a $\tilde{\lambda}_i = \frac{1}{\sigma_f} \lambda_i$. Une simple mise à l'échelle du multiplicateur de Lagrange $\lambda_i$ doit être réalisée avant de calculer l'estimateur a posteriori d'erreur. Dans le cas où l'algorithme RPCG serait stoppé prématurément, pour réduire les coûts et temps de calcul par exemple, il est possible d'introduire cette mise à l'échelle dans l'expression de la fonction coût.

Nous avons évalué les performances de cette approche sur des modèles jouets, et notamment le système de Lorenz [97]. Nous les avons comparées à celles de deux autres approches. La première, appelée "uniform" sur la Figure 3.2, consiste à utiliser un réseau d'observations uniforme, dont la résolution augmente à chaque itération de l'algorithme, en utilisant toujours la même ébauche pour chaque niveau de grille d'observations. Cette première variante vise à mesurer les précision et coût de calcul associés à la résolution du problème par RPCG sur un réseau d'observations uniforme et ce pour différentes résolutions de celui-ci. La seconde approche, appelée "benchmark", vise à exploiter la hiérarchie d'ensembles d'observations, construit depuis des grilles uniformes. Ceci revient à utiliser l'Algorithme 3.1.1 sans les étapes de sélection des observations (étapes 4-6). Ces algorithmes n'assimilant pas les mêmes observations, la comparaison se fait sur la base de la décroissance de la fonction coût définie depuis le niveau le plus fin du réseau d'observations. Nous comparons l'évolution de celle-ci en fonction du nombre total d'observations assimilées - Figure 3.2 (a) - et du nombre total d'opérations en virgule flottante réalisées dans les appels de RPCG - Figure 3.2 (b). Nous observons ainsi que l'approche adaptative proposée conduit à une décroissance plus rapide et plus forte de la fonction coût, le tout pour

**Algorithm 1** 4D-Var incrémental avec observations adaptatives

1. Initialisation : $i = 0$, $\mathbf{x}$ et le réseau d'observations le plus grossier $\mathcal{O}_0$.
2. Calcul de $(\delta\mathbf{x}_i, \lambda_i)$ solution du problème

$$\min_{\delta\mathbf{x}_i \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{x}_i + \delta\mathbf{x}_i - \mathbf{x}_b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathbf{H}_i\delta\mathbf{x}_i - \mathbf{d}_i\|^2_{\mathbf{R}_i^{-1}}, \tag{3.11}$$

   en résolvant approximativement le système

$$(\mathbf{R}_i^{-1}\mathbf{H}_i\mathbf{B}\mathbf{H}_i^T + \mathbf{I}_{m_i})\lambda_i = \mathbf{R}_i^{-1}(\mathbf{d}_i - \mathbf{H}_i(\mathbf{x}_b - \mathbf{x}_i)) \tag{3.12}$$

   avec l'algorithme RPCG dans un premier temps, puis en calculant $\delta\mathbf{x}_i = \mathbf{x}_b - \mathbf{x}_i + \mathbf{B}\mathbf{H}_i^T\lambda_i$.
3. Construction de l'ensemble auxiliaire d'observations $\tilde{\mathcal{O}}_{i+1}$ depuis l'ensemble d'observations $\mathcal{O}_i$ utilisé à l'étape 2.
4. Depuis les observations de $\tilde{\mathcal{O}}_{i+1}$, calcul de l'indicateur d'erreur

$$\eta_j \quad = \quad w_j \left\langle (\tilde{\mathbf{d}}_{i+1} - \tilde{\mathbf{H}}_{i+1}\delta\mathbf{x}_i - \tilde{\mathbf{R}}_{i+1}\tilde{\mathbf{\Pi}}_i\tilde{\lambda}_i)|_j, (\tilde{\mathbf{R}}_{i+1}^{-1}(\tilde{\mathbf{d}}_{i+1} - \tilde{\mathbf{H}}_{i+1}\delta\mathbf{x}_i - \tilde{\mathbf{R}}_{i+1}\tilde{\mathbf{\Pi}}_i\tilde{\lambda}_i))|_j \right\rangle$$

   avec $\tilde{\lambda}_i = \frac{1}{\sigma_f}\lambda_i$.
5. Sélection des observations pertinentes $\mathcal{S}_\eta$ selon

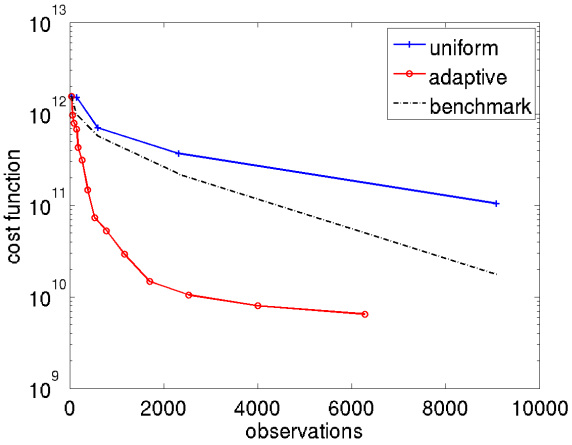
$$\theta_1 \left( \sum_{j=1}^{p_{i+1}} \eta_j \right) \leq \sum_{k \in \mathcal{S}_\eta} \eta_k.$$

6. Construction de l'ensemble d'observations $O_{i+1}$

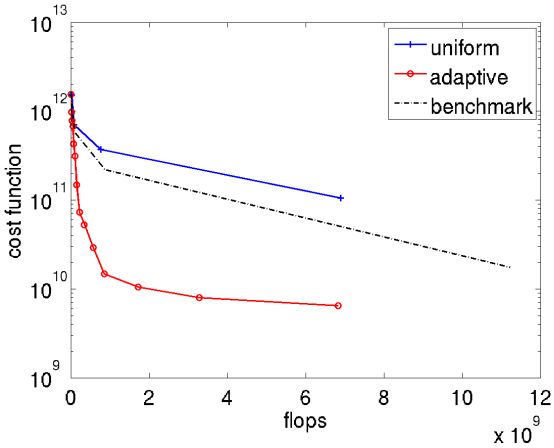
$$O_{i+1} := O_i \cup \left( \bigcup_{k \in \mathcal{S}_\eta} o_k \right).$$

7. Mise-à-jour de $\mathbf{x} \leftarrow \mathbf{x} + \delta\mathbf{x}_i$.
8. Incrémentation de $i$ et retour à l'étape 2.

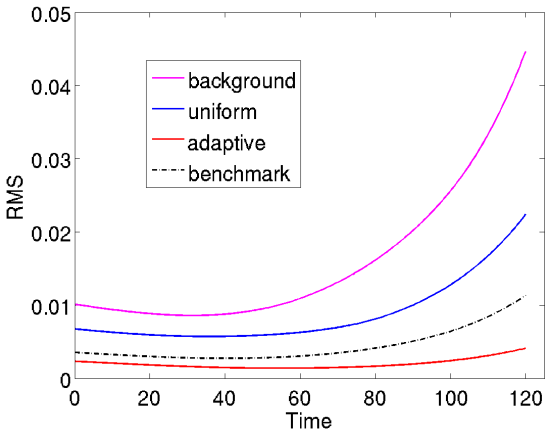
un nombre d'observations assimilées et un coût de calcul plus faibles. On note également que la solution obtenue présente une erreur RMS plus faible, tout au long de la fenêtre d'assimilation et particulièrement sur la fin de celle-ci - Figure 3.2 (c). Ceci s'explique par la densification temporelle du réseau d'observations sur la seconde partie de la fenêtre temporelle.




(a) Fonction coût versus nombre d'observations assimilées




(b) Fonction coût versus nombre d'opérations en virgule flottante




(c) Solution optimisée : erreur RMS versus temps

FIGURE 3.2 – Performances de l'algorithme avec le système Lorenz [97]. Figure tirée de Gratton *et al.* (2015). [69]

Il semble ainsi possible de sélectionner les observations à assimiler directement dans le processus d'optimisation associé à l'étape d'analyse des variantes incrémentales de l'algorithme 4D-VAR. L'extension de cette approche aux filtres/lisseurs itératifs de Kalman d'ensemble [122, 16, 17] reste à étudier. Outre l'impact des erreurs sur le gradient de la fonction issues de l'approximation des dérivées depuis l'ensemble de prévision, se pose également la question des conséquences des stratégies de localisation, notamment 4D [14, 38], nécessaires au "bon" fonctionnement de ces approches sur des applications réalistes en grande dimension, sur l'interprétation de la borne d'erreur construite depuis les multiplicateurs de Lagrange sur les différents niveaux de grilles.

### 3.1.2 De la sensibilité de l'incrément d'analyse pour la sélection des observations à assimiler

Ces travaux ont été menés dans le cadre de la thèse CIFRE de D. Mottet [100] au sein d'EDF R&D, dirigée par S. Gratton, et co-encadrée par J.-P. Argaud et moi-même, de novembre 2017 à janvier 2021. De nouveau, l'objectif fut de construire des indicateurs visant à quantifier l'impact des observations lors de l'étape d'analyse, afin d'en sélectionner un nombre plus faible à assimiler. Un des indicateurs proposés a porté sur la sensibilité de l'incrément d'analyse à des perturbations des variances des erreurs d'observation. L'idée est basée sur le fait qu'une observation avec une forte variance d'erreur (relativement à la variance de l'erreur d'ébauche) devrait peu influer sur l'analyse : il est ainsi possible de simuler la suppression d'une observation simplement en faisant tendre sa variance d'erreur vers l'infini.

En supposant la matrice de covariance d'erreur d'observation diagonale, il est possible d'écrire la matrice de covariance d'erreur d'observation perturbée sous la forme

$$\mathbf{R}_\epsilon = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_p^2) \circ \left[\mathbf{I}_p + \left(\frac{1}{\epsilon} - 1\right)\mathbf{E}_{ii}\right]. \tag{3.13}$$

avec $\circ$ le produit de Schur, $(\sigma_1^2, \cdots, \sigma_p^2)$ les variances d'erreur d'observation, $\epsilon > 0$ le facteur de perturbation, $i$ l'indice de l'observation dont la variance d'erreur est perturbéee, et $\mathbf{E}_{ii}$ la matrice de la base canonique de $\mathcal{M}_p(\mathbb{R})$ associée au couple $(i, i)$.

Dans le cas où la matrice de covariance d'erreur d'ébauche est inversible, l'incrément d'analyse $\delta\mathbf{x}(\epsilon)$ s'écrit alors :

$$\begin{aligned}\delta\mathbf{x}(\epsilon) &= \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}_\epsilon^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} \\ &= (\mathbf{B}^{-1} + \mathbf{H}^T\left[\sum_{j=1}^{p}\frac{\mathbf{E}_{jj}}{\sigma_j^2} + \frac{1}{\sigma_i^2}(\epsilon - 1)\mathbf{E}_{ii}\right]\mathbf{H})^{-1}\mathbf{H}^T\left[\sum_{j=1}^{p}\frac{\mathbf{E}_{jj}}{\sigma_j^2} + \frac{1}{\sigma_i^2}(\epsilon - 1)\mathbf{E}_{ii}\right]\mathbf{d}\end{aligned} \tag{3.14}$$

avec $\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}^b$. En définissant

$$\tilde{\mathbf{g}}_i(\epsilon) = \left(\mathbf{B}^{-1} + \mathbf{H}^T\left[\sum_{j=1}^{p}\frac{E_{jj}}{\sigma_j^2} + \frac{1}{\sigma_i^2}(\epsilon - 1)E_{ii}\right]\mathbf{H}\right)^{-1}\mathbf{h}_i \tag{3.15}$$

avec $\mathbf{h}_i$ la $i$-ème ligne de $\mathbf{H}$, nous obtenons alors la dérivée de l'incrément d'analyse :

$$\delta\mathbf{x}'(\epsilon) = -\frac{\tilde{\mathbf{g}}_i(\epsilon)^T}{\sigma_i^2}\left[\sum_{j=1}^{p}\frac{d_j}{\sigma_j^2}\mathbf{h}_j + \left(\frac{\epsilon - 1}{\sigma_i^2}\right)d_i\mathbf{h}_i\right]\tilde{\mathbf{g}}_i(\epsilon) + \frac{d_i}{\sigma_i^2}\tilde{\mathbf{g}}_i(\epsilon) \tag{3.16}$$

Dans le cas où la matrice est de rang faible, ce qui est souvent le cas pour les méthodes d'ensemble, il est alors possible de reformuler le problème d'optimisation pour travailler sur un sous-espace de rang faible, à l'instar de la formulation ETKF [11, 86] du filtre de Kalman d'ensemble. La fonction coût associée s'écrit alors

$$J(\mathbf{w}) = \left(\frac{N-1}{2}\right)\mathbf{w}^T\mathbf{w} + \frac{1}{2}[\mathbf{y} - \mathbf{H}(\mathbf{x}^f + \mathbf{A}^f\mathbf{w})]^T\mathbf{R}^{-1}[\mathbf{y} - \mathbf{H}(\mathbf{x}^f + \mathbf{A}^f\mathbf{w})], \tag{3.17}$$

avec $\mathbf{A}^f \in \mathcal{M}_N(\mathbb{R})$ la matrice des anomalies de prévisions, associée à la moyenne de l'ensemble $\mathbf{x}^f$. L'étape d'analyse conduit alors à :

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{A}^f \mathbf{w}^a. \tag{3.18}$$

La condition nécessaire d'optimalité du premier ordre donne

$$\mathbf{w}^a = [(N-1)\mathbf{I}_N + \mathbf{S}^T\mathbf{R}^{-1}\mathbf{S}]^{-1}\mathbf{S}^T\mathbf{R}^{-1}\mathbf{d}, \tag{3.19}$$

avec $\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}^f$ et $\mathbf{S} = \mathbf{H}\mathbf{A}^f$.
En définissant

$$\tilde{\mathbf{g}}_i^e(\epsilon) = \left[(N-1)\mathbf{I}_N + \mathbf{S}^T\left[\sum_{j=1}^{p}\frac{E_{jj}}{\sigma_j^2} + \frac{1}{\sigma_i^2}(\epsilon-1)E_{ii}\right]\mathbf{S}\right]^{-1}\mathbf{s}_i. \tag{3.20}$$

avec $\mathbf{s}_i$ la $i$-ème ligne de $\mathbf{S}$, il vient :

$$\delta\mathbf{x}'(\epsilon) = \mathbf{A}^f\left(-\frac{1}{\sigma_i^2}\tilde{\mathbf{g}}_i^e(\epsilon)^T\tilde{\mathbf{g}}_i^e(\epsilon)\left[\sum_{j=1}^{p}\frac{d_j}{\sigma_j^2}\mathbf{s}_j + \left(\frac{\epsilon-1}{\sigma_i^2}\right)d_i\mathbf{s}_i\right] + \frac{d_i}{\sigma_i^2}\tilde{\mathbf{g}}_i^e(\epsilon)\right). \tag{3.21}$$

Il est ainsi possible de quantifier l'impact de la suppression d'une observation sur l'incrément d'analyse, que ce soit pour certaines composantes d'intérêt (variables du vecteur d'état considérées comme d'importance, distance à l'observation, etc..) ou de manière globale (norme euclidienne du vecteur par exemple). Néanmoins, sans information a priori sur l'influence des observations, cette approche nécessite de calculer ces quantités sur l'ensemble des observations disponibles, engendrant un coût de calcul important. Disposant d'un réseau d'observations relativement dense, cette approche semble envisageable pour des ensembles de petite taille - résolution numérique des systèmes linéaires (3.20) - pour le design a priori de la composante observation du système d'assimilation.

Un exemple de résultat numérique se trouve Figure 3.3 dans le cadre du système de Lorenz-63 [96]. Dans cette expérience, les trois composantes du vecteur d'état sont observées : les courbes bleu, verte et orange représentent l'évolution de la moyenne de l'ensemble sur la fenêtre temporelle, tandis que les points représentent les valeurs des observations. La courbe rouge représente l'indice de l'observation conduisant à la plus faible valeur de l'indicateur indiquant ainsi la moins influente, pour chacune des étapes d'analyse. Dans cette configuration, il semble souhaitable de supprimer l'observation de la seconde ou de la troisième composante du vecteur selon les étapes d'analyse. A contrario, l'observation de la première composante semble primordiale, la valeur de l'indicateur étant la plus faible seulement lors d'une analyse.

En terme de qualité de la solution, la Figure 3.4 présente l'évolution de l'erreur RMS relative de la solution après analyse selon que toutes les observations sont assimilées (figure (A) courbe bleu), où qu'une observation est supprimée selon le diagnostique construit depuis la matrice d'influence [27, 28] (courbe orange) ou l'indicateur défini par Eq. (3.16) (courbe verte).

Nous pouvons noter que la suppression d'une observation (une composante du vecteur d'état) lors de l'analyse ne dégrade pas la qualité de la solution analysée et que l'indicateur proposé durant la thèse de Dimitri conduit à des performances similaires à celles obtenues depuis [27, 28].

La suite de ces travaux portent naturellement sur leur application aux problèmes industriels d'EDF, et notamment dans le domaine de la neutronique.

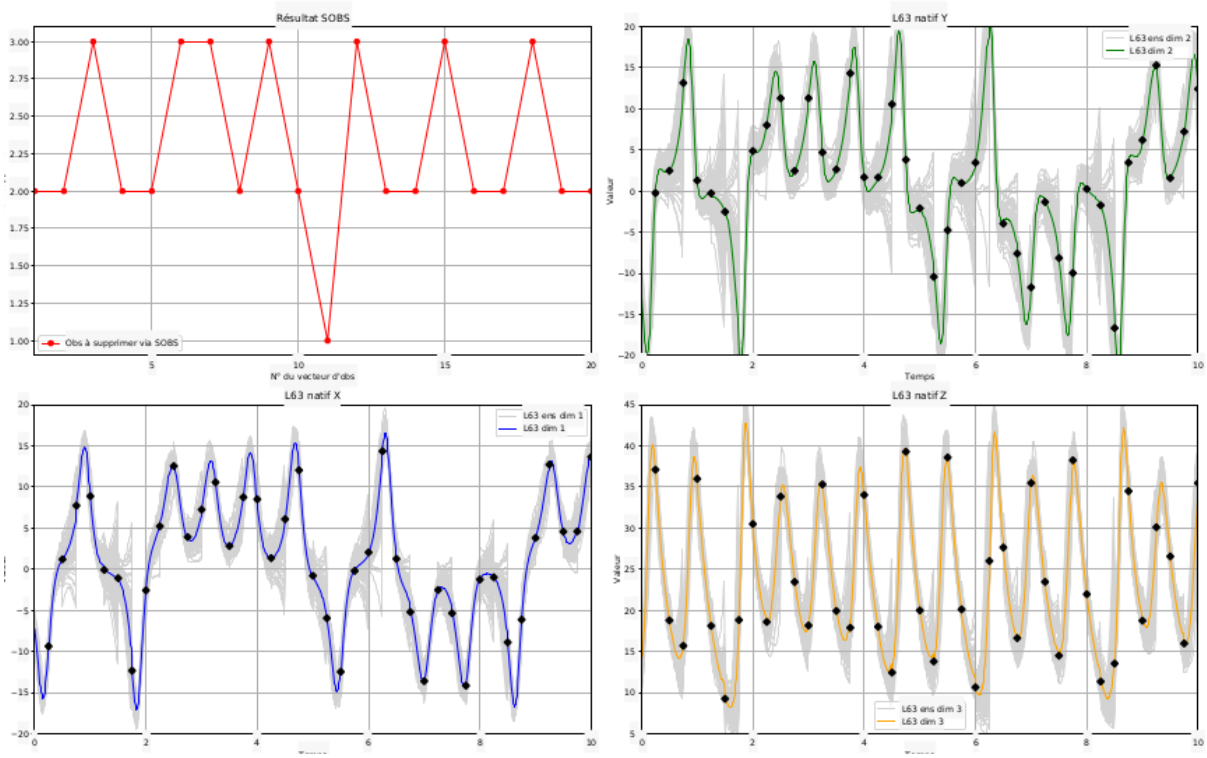
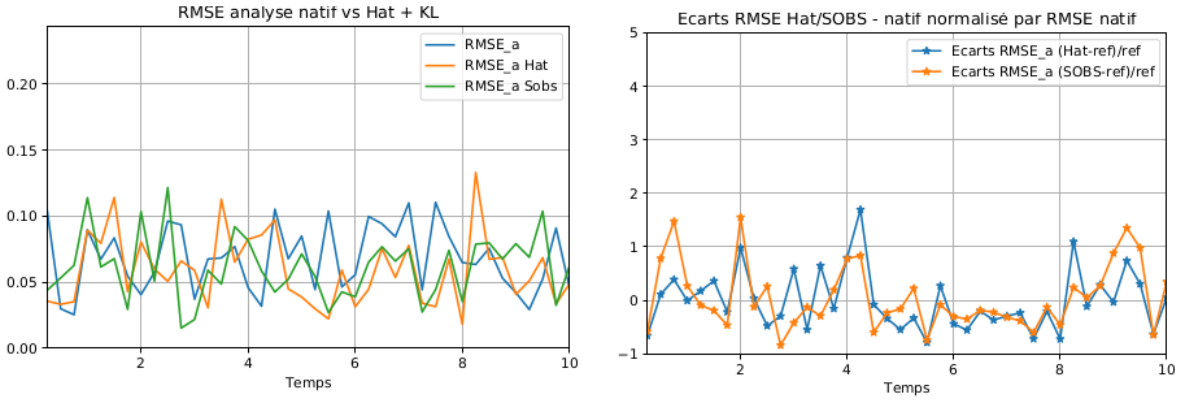



FIGURE 3.3 – Système de Lorenz-63 : évolution de l'ensemble pour les composantes du vecteur d'état et valeurs des observations, ainsi que l'indice de l'observation à supprimer pour chaque analyse (Eq. 3.16). Figure tirée de Mottet (2021) [100]




(A) RMSEs d'analyse.

(B) Ecarts relatifs.

FIGURE 3.4 – Système de Lorenz-63 : erreur RMS d'analyse pour les solution obtenues après sélection des observations depuis la matrice d'influence proposée par [27, 28] ("Hat") et l'indicateur défini par Eq. (3.16) ("SOBS"). Figure tirée de Mottet (2021) [100]

## 3.2 Au niveau du modèle

La seconde source d'informations sur laquelle les méthodes d'assimilation se basent réside dans les modèles, basés sur la résolution numérique bien souvent d'équations aux dérivées partielles, potentiellement stochastiques. Celles-ci sont souvent coûteuses, que ce soit en ressources de calcul et en temps, et ce d'autant plus que la ou les grilles spatiales sont fines, et les conséquences pour le choix des pas de temps que cela implique. De surcroît, l'utilisation d'un modèle adjoint pour les approches variationnelles 4D-Var et l'essor des méthodes d'ensemble ont naturellement conduit à augmenter ces coûts et temps de calcul. Se pose naturellement la question de l'emploi de modèles approchés, potentiellement moins précis mais peu coûteux, tant pour les simulations directes que adjointes. Ces modèles peuvent être construits depuis des approches dites par réduction ou par de l'apprentissage machine, notamment les réseaux de neurones. Une autre conséquence de l'augmentation de la dimension du vecteur d'état du modèle, directement dépendante de la finesse de la grille spatiale, est la taille des fichiers de sortie, ceux-ci pouvant être requis pour des points de contrôle (arrêt prématuré de la simulation), au redémarrage du modèle à la fin de la fenêtre de simulation, et aux diagnostics intermédiaires en cours de simulation. Il est naturellement possible de compresser ces données, ou d'utiliser une précision dégradée, afin de réduire l'espace disque nécessaire au stockage de celles-ci. Néanmoins, ceci nécessite d'être en mesure de pouvoir exploiter a posteriori ces données dégradées.

### 3.2.1 Approche par réduction de modèles ou apprentissage profond

Durant la visite scientifique de J. Pelc au NERSC à l'automne 2010, dans le cadre de sa thèse à TU Delft (Pays-Bas), j'ai eu l'opportunité de m'intéresser aux techniques de réduction de modèle, avec application à l'assimilation variationnelle de données pour l'estimation jointe du vecteur d'état et de certains paramètres d'un modèle d'écosystème marin simplifié [112]. L'objectif de ces travaux fut d'évaluer les performances d'un algorithme de 4D-Var avec modèle réduit [140] sur un modèle simplifié, avant application dans un cadre réaliste. L'intérêt de cette approche réside dans sa capacité à approximer les dérivées partielles du modèle, vis-à-vis du vecteur d'état ou des paramètres, sans modèle linéaire tangent, et permet ainsi l'approximation du gradient de la fonction coût de l'algorithme 4D-Var sans recourir au modèle adjoint. Ceci était crucial dans la thèse de J. Pelc, celle-ci ne disposant pas des sources du code de la composante biogéochimique de son modèle marin (logiciel propriétaire). Pour ce faire, l'approche consiste en la projection du vecteur d'état, sur un sous-espace de dimension faible, afin d'y réaliser une approximation des dérivées partielles par différences finies. Ce sous-espace est construit depuis la décomposition en valeurs singulières de la matrice d'anomalie associée à une simulation d'ensemble préalable, pour laquelle seulement sont conservés les vecteurs singuliers à gauche associés au plus grande valeurs singulières (cf l'analyse en composantes principales).

Dans le cadre d'expériences jumelles dans un modèle simplifié de la biogéochimie d'une colonne d'eau (1D), il a été illustré la capacité de cette approche à estimer conjointement certains paramètres du modèle et le vecteur d'état. La figure 3.5 représente les valeurs des trois paramètres associés à taux d'efficacité du broutage des herbivores (f), de leur perte due aux carnivores (g) et de la perte métabolique des plantes (r), pour 15 expériences différentes. Nous pouvons noter à chaque fois une très nette amélioration de la valeur des paramètres après assimilation. Les détails de cette étude sont disponibles dans [112].

Plus récemment, dans le cadre de la thèse d'A. Rouvière à l'ONERA, co-encadrée par F.

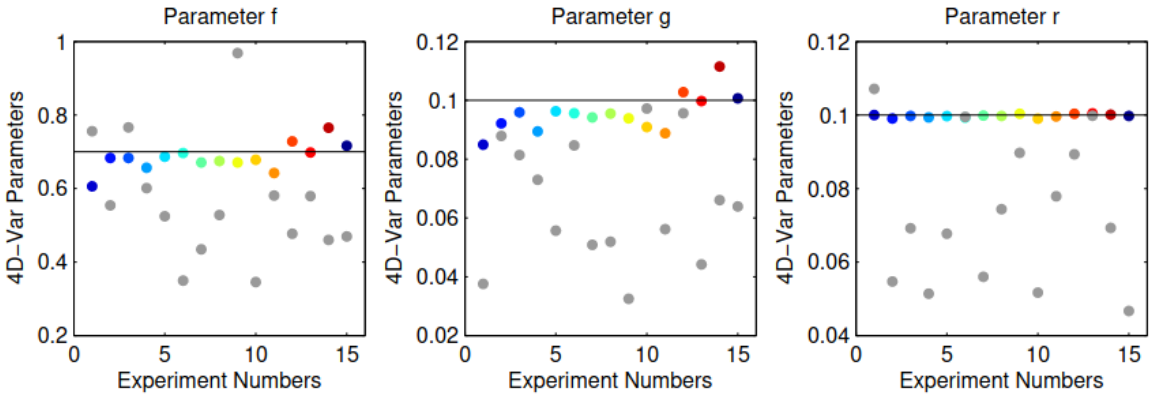



FIGURE 3.5 – Estimation de paramètres biogéochimiques : résultats de l'algorithme 4D-Var avec réduction de modèle pour 15 expériences. Les cercles gris correspondent à l'ébauche, ceux en couleurs aux résultats de l'optimisation. Le trait gris correspond à la valeur de référence. Figure tirée de Pelc *et al.* (2012) [112]

Méry et L. Pascal (ONERA), et S. Gratton et moi-même, nous nous sommes intéressés au développement de modèles de tolérances de surfaces pour les couches limites dans le cas de défauts (2D). Le problème industriel associé est la conception d'ailes, avec un contrôle sur l'amplitude des défauts de surface, permettant une réduction de la consommation de carburant. Pour cela, une solution consiste à réduire la traînée de frottement due à la surface rugueuse des ailes. Une couche limite laminaire induisant un coefficient de frottement plus faible qu'une couche limite turbulente, le fait de retarder l'apparition de la transition laminaire-turbulente entraîne ainsi une réduction de la traînée. Or la trajectoire de la transition est fortement déterminée par l'impact des perturbations externes sur la couche limite, notamment les défauts de surface. Dans le cas de flots 2D, il est possible d'estimer la localisation spatiale de la zone de transition depuis la méthode $e^N$ [132, 139]. Le facteur $N$ représente l'amplification des instabilités de la couche limite, et la transition se déclenche dès lors qu'il dépasse une valeur seuil comprise en 9 et 11. Un des objectifs des travaux d'A. Rouvière fut donc l'estimation de l'augmentation du facteur $N$ associé à un défaut de surface, comparativement à celui associé à une surface lisse, afin de quantifier la tolérance de la couche limite aux défauts.

La figure 3.6 (a) représente un exemple de défaut, ainsi que le sens de l'écoulement. Ce type de défaut paramétré par deux hauteurs distinctes, n'avait pas été étudié à ce jour (seulement des défauts de type marche). Le calcul de la variation du facteur $N$, suite à l'écoulement sur un défaut, est représenté sur la 3.6 (b). Ceci se fait en calculant dans un premier temps, l'évolution de $N$ selon la distance à l'origine (début du domaine), dans le cas d'une surface lisse. Ceci nécessite notamment la résolution numérique d'équations de Navier-Stokes, pour le calcul de l'état stationnaire, puis celles d'équations de Navier-Stokes linéarisées, pour modéliser le dynamique des perturbations autour de cet état, pour une fréquence de perturbation donnée. Les mêmes calculs sont ensuite réalisés dans le cas d'une configuration de défaut particulière. En comparant l'évolution des facteurs $N$ obtenus pour chaque configuration, il est possible d'estimer la variation de $N$ au voisinage du défaut ($\Delta N_{peak}$) et au loin ($\Delta N_{far}$). Le calcul de tels $\Delta N$ est coûteux, puisqu'il faut réaliser un grand nombre de simulations de la dynamique des

perturbations (une par fréquence "test"), pour obtenir une estimation de l'évolution du facteur $N$ associé au défaut. Nous nous sommes donc intéressés à l'utilisation de méthodes d'apprentissage, et notamment les réseaux de neurones, pour la construction d'un prédicteur des variations $(\Delta N_{peak}, \Delta N_{far})$ dès lors connus les paramètres d'un nouveau défaut.

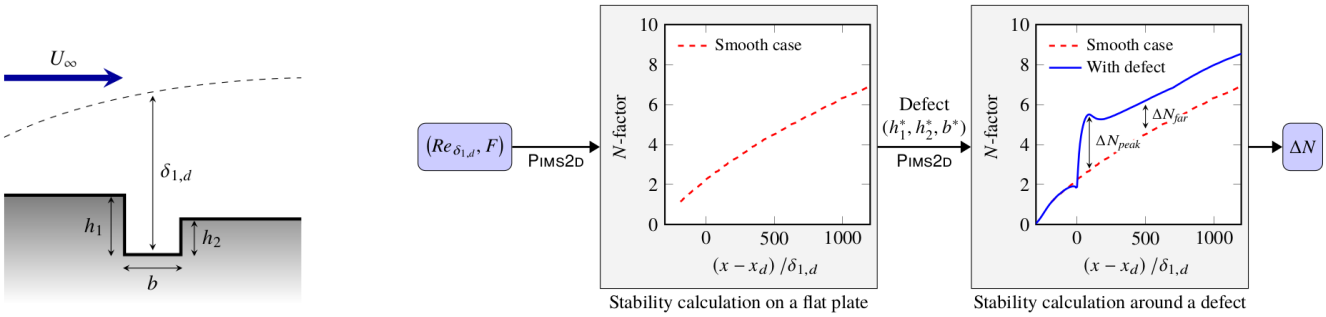



FIGURE 3.6 – Apprentissage profond pour la prédiction de la transition de la couche de limite : (a) Paramètres expérimentaux caractérisant le défaut et l'écoulement, (b) Méthodologie pour la création de la base d'apprentissage. Figures tirées de Rouvière *et al.* (soumis) [119]

Pour trois différentes structures de réseaux, nous constatons que les prédictions du $\Delta N_{far}$ par les réseaux, sur le jeu de données de validation, sont en adéquation avec les valeurs de référence (figure 3.7 à gauche).

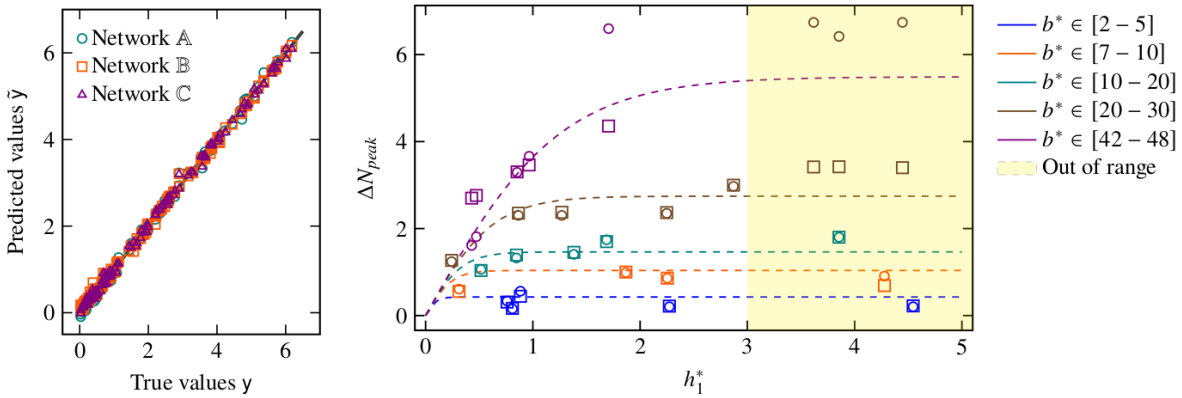



FIGURE 3.7 – Apprentissage profond pour la prédiction de la transition de la couche de limite : à gauche, prédiction de $\Delta N_{far}$ versus la référence, à droite prédiction de $\Delta N_{peak}$ selon les valeurs de $h_1$ (hauteur amont de la cavité). Les couleurs représentent différentes valeurs de $b$ (longueur de la cavité) : les cercles correspondent à des données expérimentales et les lignes pointillées à des modèles d'"interpolation" de ces données [35], et les carrés les résultats de la prédiction des réseaux entraînés (moyenne des trois réseaux). Figure tirée de Rouvière *et al.* (soumis) [119]

L'étape suivante fût donc la comparaison avec des données expérimentales. Les résultats sont présentés figure 3.7 (droite) et représentent les prédictions de $\Delta N_{peak}$ selon les valeurs de $h_1$ (hauteur amont de la cavité). Les couleurs représentent différentes valeurs de $b$ (longueur de

la cavité) : les cercles correspondent aux données expérimentales et les lignes pointillées à des modèles interpolant ces données [35], et les carrés les résultats de la prédiction des réseaux entraînés (moyenne des trois réseaux). La zone colorée en jaune correspond à des plages de valeurs de $h_1$ absentes des données d'apprentissage. Il en est de même pour les valeurs de $b$ supérieures à 15 (couleurs marron et violet). Nous remarquons que l'adéquation entre les prédictions des réseaux et les données expérimentales est très bonne, dès lors que certaines plages de valeurs des paramètres associées à ces données sont présentes dans la base d'apprentissage. Néanmoins, la généralisation à des données d'entrée en dehors de la base d'apprentissage devient difficile (cf l'inadéquation entre les cercles et carrés marron et violet dans la zone jaune). Nous avons vécu la même expérience, dans des travaux menés par R. Benshila et R. Almar sur l'estimation de profils de plage depuis des images de houles par des réseaux de neurones [7]. Là où l'estimation de profils de plage, obtenus depuis des données simulées (résolution numérique d'une EDP), se montre performant sur le jeu de données (simulées) test (cf Figure 3.8), celle sur des données réelles se révèle plus difficile (cf Figure 3.9). La généralisation à des plages de valeurs complètement inconnues de la base d'apprentissage reste encore un problème difficile, et générique à ces approches.

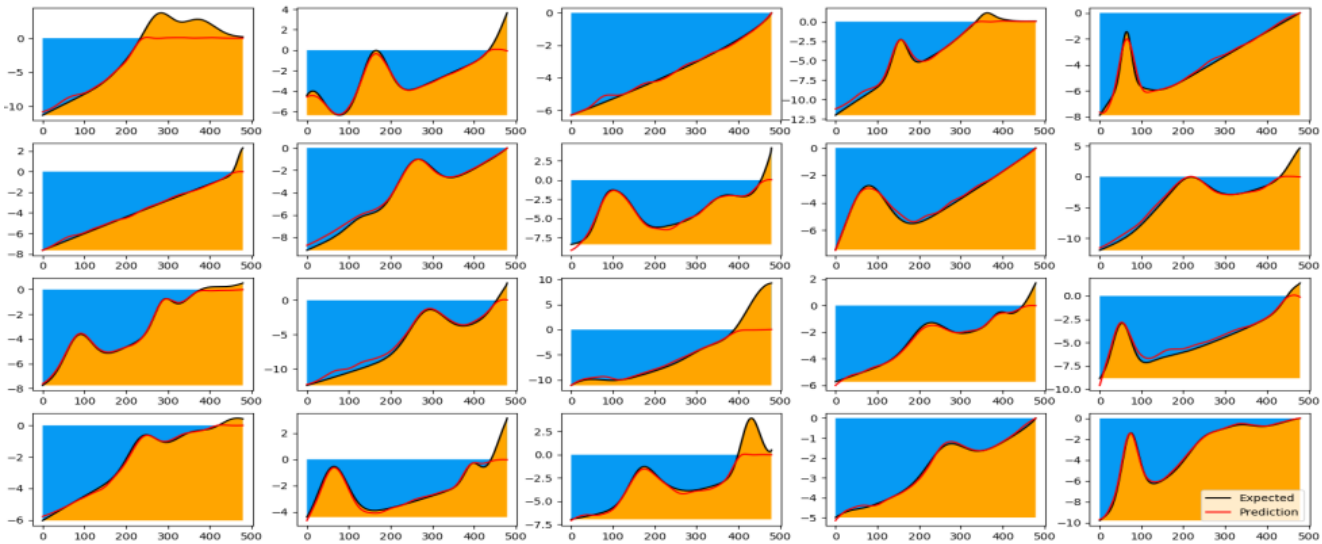



FIGURE 3.8 – Estimation de profils de plage par réseaux de neurones - profils synthétiques : prédictions de la bathymétrie (courbe rouge) versus profils simulés. Figure tirée de Benshila *et al.* (2020) [7]

### 3.2.2 Methodes d'ensemble

Nous nous sommes intéressés avec S. Gratton et D. Titley-Peloquin [70] à l'estimation de matrices de covariance d'erreur, telles que celles introduites pour modéliser les incertitudes dans les variantes ensemblistes du filtre de Kalman, et ce depuis des échantillons bruités. En pratique, ceux-ci peuvent être issus de simulations dégradées (modèles simplifiés, grille plus grossière, calcul en précision multiple) pour lesquelles une réduction des coûts de calcul est envisagée, ou alors d'un stockage de l'ensemble en précision dégradée, afin de réduire la taille des fichiers. Dans un cas simplifié, nous avons proposé une approche par maximum de vraisemblance sur

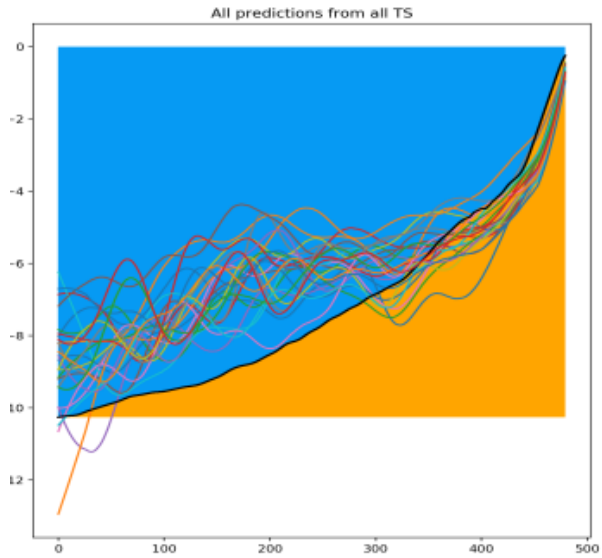



FIGURE 3.9 – Estimation de profils de plage par réseaux de neurones - cas réaliste (plage Gran Popo, Bénin) : prédictions de la bathymétrie à différentes dates et pour différentes conditions de vagues (courbes de couleur) vs profil moyen. Figure tirée de Benshila *et al.* (2020) [7]

des sous-espaces de dimension adaptative, permettant l'estimation de telles matrices, et leur utilisation dans les filtres les plus utilisés opérationnellement.

Nous cherchons ainsi à estimer la matrice de covariance $\mathbf{\Sigma} \in \mathbb{R}^{m\times m}$ associée au vecteur $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ depuis des échantillons bruités de cette variable. Nous supposons disposer de $n_x$ réalisations $\mathbf{x}_i$, ainsi que $n_y$ réalisations bruitées $\mathbf{y}_i$, de la variable aléatoire $\mathbf{x}$. Les vecteurs $\mathbf{y}_i$ sont des réalisation de la variable aléatoire $\mathbf{y}$ définie par :

$$\mathbf{y} = \mathbf{x} + \mathbf{v}, \quad \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}) \text{ et } \mathbf{x} \text{ et } \mathbf{v} \text{ indépendants}, \quad \text{i.e.,} \quad \mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma} + \mathbf{C}),$$

$\mathbf{C}$ une matrice symétrique définie positive, que nous supposons connue ou pour laquelle nous disposons d'une estimation. Cette hypothèse est forte, tant dans la distribution de la variable $\mathbf{y}$ que la connaissance de la matrice de covariance du bruit. Néanmoins, cette approche constitue un premier pas vers une modélisation plus complexe. Toutefois, nous faisons l'hypothèse que les matrices $\mathbf{C}$ et $\mathbf{\Sigma}$ n'ont pas nécessairement la même structure.

Soient $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{n_x}] \in \mathbb{R}^{m\times n_x}$ l'ensemble des réalisations non-bruitées, et $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_{n_y}] \in \mathbb{R}^{m\times n_y}$ l'ensemble des réalisations bruitées, avec $n = n_x + n_y < m$. Nous notons $\bar{\mathbf{x}} = \frac{1}{n}[\mathbf{X}, \mathbf{Y}]\mathbf{e}$ la moyenne empirique de l'échantillon $[\mathbf{X}, \mathbf{Y}]$, et

$$\mathbf{A}_\mathbf{x} = \mathbf{X} - \bar{\mathbf{x}}\mathbf{e}^T, \qquad \mathbf{A}_\mathbf{y} = \mathbf{Y} - \bar{\mathbf{x}}\mathbf{e}^T \tag{3.22}$$

les matrices d'anomalies de $\mathbf{X}$ et $\mathbf{Y}$, avec $\mathbf{e} \in \mathbb{R}^n$ le vecteur dont toutes les composantes sont égalent à 1.

Dans le but de fournir l'estimateur du maximum de vraisemblance de $\mathbf{\Sigma}$ depuis $[\mathbf{X}, \mathbf{Y}]$, nous

nous intéressons à la minimisation de la fonction suivante :

$$\begin{aligned} -\log(\mathcal{L}(\hat{\mathbf{\Sigma}}|\mathbf{X},\mathbf{Y})) = \frac{mn}{2}\log(2\pi) + \frac{n_x}{2}\log\det{}^*(\hat{\mathbf{\Sigma}}) + \frac{n_y}{2}\log\det{}^*(\hat{\mathbf{\Sigma}}+\mathbf{C}) \\ + \frac{1}{2}\mathrm{tr}\big(\mathbf{A_x}\mathbf{A_x}^T\hat{\mathbf{\Sigma}}^\dagger\big) + \frac{1}{2}\mathrm{tr}\big(\mathbf{A_y}\mathbf{A_y}^T(\hat{\mathbf{\Sigma}}+\mathbf{C})^\dagger\big). \end{aligned} \tag{3.23}$$

avec $\det^*$ le pseudo-déterminant (produit des valeurs propres non nulles) et $\hat{\mathbf{\Sigma}}^\dagger$ la matrice pseudo-inverse de $\hat{\mathbf{\Sigma}}$. Nous supposons de plus que $\hat{\mathbf{\Sigma}}$ se décompose sous la forme :

$$\hat{\mathbf{\Sigma}} = \mathbf{Q}\mathbf{L}\mathbf{L}^T\mathbf{Q}^T,$$

avec $\mathbf{Q} \in \mathbb{R}^{m\times p}$ une matrice à définir et dont les colonnes sont orthonormales et $\mathbf{L} \in \mathbb{R}^{p\times p}$ une matrice triangulaire inférieure inversible à estimer. Pour cette dernière, nous utilisation la représentation $\mathbf{L} = \Pi(\boldsymbol{\ell})$ pour le stockage des $p(p+1)/2$ entrées de sa partie triangulaire inférieur dans un vecteur $\boldsymbol{\ell}$. Ceci conduit à définir l'opérateur $\Pi$ de la manière suivante

$$\Pi \ : \ \mathbb{R}^{p(p+1)/2} \to \mathbb{R}^{p\times p}, \qquad \Pi(\boldsymbol{\ell}) = \begin{bmatrix} \ell_1 & & & \\ \ell_2 & \ell_{p+1} & & \\ \vdots & \vdots & \ddots & \\ \ell_p & \ell_{2p-1} & \dots & \ell_{p(p+1)/2} \end{bmatrix}.$$

On vérifie alors que

$$\mathrm{vec}\big(\Pi(\boldsymbol{\ell})\big) = \mathbf{P}\boldsymbol{\ell}, \qquad \mathbf{P} = \begin{bmatrix} \mathbf{I}_p & & & & & \\ & 0 & & & & \\ & & \mathbf{I}_{p-1} & & & \\ & & & 0 & & \\ & & & 0 & & \\ & & & & \mathbf{I}_{p-2} & \\ & & & & & \ddots \end{bmatrix} \in \mathbb{R}^{p^2\times p(p+1)/2}. \tag{3.24}$$

Le problème se réécrit alors comme la minimisation de la fonctionnelle

$$\begin{aligned} -\log(\mathcal{L}(\boldsymbol{\ell}|\mathbf{X},\mathbf{Y})) = \frac{mn}{2}\log(2\pi) + n_x\log|\det(\mathbf{L})| + \frac{n_y}{2}\log\det{}^*(\mathbf{Q}\mathbf{L}\mathbf{L}^T\mathbf{Q}^T+\mathbf{C}) \\ + \frac{1}{2}\mathrm{tr}\big(\mathbf{A_x}\mathbf{A_x}^T\mathbf{Q}(\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{Q}^T\big) + \frac{1}{2}\mathrm{tr}\big(\mathbf{A_y}\mathbf{A_y}^T(\mathbf{Q}\mathbf{L}\mathbf{L}^T\mathbf{Q}^T+\mathbf{C})^\dagger\big), \end{aligned} \tag{3.25}$$

avec $\mathbf{L} = \Pi(\boldsymbol{\ell})$. Néanmoins, la minimisation de cette fonction peut être difficile numériquement quand la norme de $\mathbf{C}$ est petite. En effet, la matrice $\mathbf{Q}\mathbf{L}\mathbf{L}^T\mathbf{Q}^T + \mathbf{C}$ peut être vue comme une petite perturbation d'une matrice de rang faible. La résolution numérique du calcul de sa pseudo-inverse peut conduire à des erreurs conséquentes. Pour circonvenir à ce problème, nous avons proposé l'approximation suivante :

$$\hat{\mathbf{\Sigma}} + \mathbf{C} = \mathbf{Q}\mathbf{L}\mathbf{L}^T\mathbf{Q}^T + \mathbf{C} \approx \mathbf{Q}(\mathbf{L}\mathbf{L}^T + \mathbf{Q}^T\mathbf{C}\mathbf{Q})\mathbf{Q}^T.$$

Cette approche est équivalente à remplacer la matrice de covariance de $\mathbf{y}$ par celle de $\mathbf{Q}\mathbf{Q}^T\mathbf{y}$, la projection de $\mathbf{y}$ sur le sous-espace engendrée par les colonnes de $\mathbf{Q}$. Finalement, nous sommes confrontés à la minimisation de la fonction $h$ définie par :

$$\begin{aligned} h(\boldsymbol{\ell}) = n_x\log|\det(\mathbf{L})| + \frac{n_y}{2}\log\det(\mathbf{L}\mathbf{L}^T + \mathbf{Q}^T\mathbf{C}\mathbf{Q}) \\ + \frac{1}{2}\mathrm{tr}\left(\mathbf{A_x}^T\mathbf{Q}(\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{Q}^T\mathbf{A_x}\right) + \frac{1}{2}\mathrm{tr}\left(\mathbf{A_y}^T\mathbf{Q}(\mathbf{L}\mathbf{L}^T + \mathbf{Q}^T\mathbf{C}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{A_y}\right), \end{aligned} \tag{3.26}$$

avec $\mathbf{L} = \Pi(\boldsymbol{\ell})$.

La condition nécessaire d'optimalité du premier ordre nous amène à chercher les points critiques de $h$, à savoir les points annulant son gradient. Celui-ci est défini par :

$$\nabla h(\boldsymbol{\ell}) = \mathbf{P}^T \mathrm{vec}\big(n_x \mathbf{L}^{-T} + n_y(\mathbf{L}\mathbf{L}^T + \mathbf{Q}^T\mathbf{C}\mathbf{Q})^{-1}\mathbf{L} - \mathbf{L}^{-T}\mathbf{N_1}\mathbf{N_1}^T - \mathbf{N_2}\mathbf{N_2}^T\mathbf{L}\big), \tag{3.27}$$

avec

$$\mathbf{N_1} = \mathbf{L}^{-1}\mathbf{Q}^T\mathbf{A_x}, \qquad \mathbf{N_2} = (\mathbf{L}\mathbf{L}^T + \mathbf{Q}^T\mathbf{C}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{A_y}. \tag{3.28}$$

Malheureusement, n'étant pas en mesure d'obtenir une expression analytique d'un ou des points critiques de $h$, il nous est nécessaire en pratique d'utiliser des algorithmes numériques d'optimisation tels que les régions de confiance [32]. Il faut cependant noter que la dimension de ce problème d'optimisation est petite, à savoir $p(p+1)/2$, avec $p$ inférieur à la taille de l'ensemble. Nous avons également proposé d'adapter le choix de $p$ durant le processus d'optimisation. Partant du rang de la matrice d'anomalies, la valeur de $p$ est réduit si la minimisation "échoue". Ceci peut arriver typiquement lorsque le conditionnement de la Hessienne de $h$ devient supérieur à un seuil fixé. Dans ce cas, la minimisation est stoppée et relancée avec une valeur de $p$ réduite. Si jamais $p$ devait être trop petit, le processus s'arrête et la matrice de covariance empirique est utilisée.

Une fois avoir choisi la matrice $\mathbf{Q}$ et minimisé $h$ pour obtenir la matrice $\mathbf{L}$, il reste à préciser comment utiliser ces matrices, dans les filtres de Kalman d'ensemble, idéalement sans avoir à modifier les codes sources implémentant l'étape d'analyse. Pour cela, nous avons proposé une stratégie pour construire une pseudo-matrice d'anomalies $\mathbf{A}_Q^f \in \mathbb{R}^{m\times n}$ depuis la matrice $\mathbf{QL} \in \mathbb{R}^{m\times p}$, avec $p \leq n-1$. Celle-ci doit vérifier les conditions suivantes :

$$\mathbf{P}^f = (\mathbf{QL})(\mathbf{QL})^T = \left(\frac{\mathbf{A}_Q^f}{\sqrt{n-1}}\right)\left(\frac{\mathbf{A}_Q^f}{\sqrt{n-1}}\right)^T \quad \text{et} \quad \mathbf{A}_Q^f\mathbf{e} = 0,$$

La seconde équation vise à garantir que la somme des colonnes de $\mathbf{A}_Q^f$ est nulle. Cette propriété est importante, car elle garantie que la somme des colonnes de la matrice d'anomalie issue de l'analyse est bien nulle également. Ceci permet d'obtenir un ensemble d'analyse centré sur la moyenne calculée dans cette même phase.

La matrice $\mathbf{A}_Q^f$ ne peut être simplement choisie égale à $\sqrt{n-1}\mathbf{QL}$ : leurs nombres de colonnes diffèrent et $\mathbf{QLe} \neq 0$. Une première stratégie proposée dans [84] consiste à multiplier à droite la matrice $\sqrt{n-1}\mathbf{QL}$ par une matrice aléatoire orthogonale $\mathbf{V}_p \in \mathbb{R}^{n\times p}$, dont les colonnes sont orthogonales à $\mathbf{e}$ : $\mathbf{A}_Q^f = \sqrt{n-1}\mathbf{QLV}_p^T$. Néanmoins, il est probable que l'ensemble de prévision ainsi obtenu, $\mathbf{E}_Q^f = \mathbf{x}^f\mathbf{e}^T + \mathbf{A}_Q^f$, soit constitué de membres physiquement irréalistes, dégradant la qualité des prévisions lors des cycles d'analyse. Pour le cas $p = n-1$, une seconde stratégie, suggérée récemment par [50], consiste à poser $\mathbf{A}_Q^f = [0, \mathbf{QL}]\mathbf{V}_\epsilon$, avec $\mathbf{V}_\epsilon \in \mathbb{R}^{n\times n}$ dépendant d'un paramètre $\epsilon \in (0,1)$ et telle que $\mathbf{V}_\epsilon\mathbf{V}_\epsilon^T = \mathbf{I}_n$ et $\mathbf{V}_\epsilon\mathbf{e} = \mathbf{e}_1$, avec $\mathbf{e}_1$ le premier vecteur de la base canonique de $\mathbb{R}^n$.

Nous avons plutôt proposé d'exploiter le choix qui nous est offert pour définir la matrice $\mathbf{Q}$, afin de reconstruire cette pseudo-matrice d'anomalies $\mathbf{A}_Q^f$. Dans notre cas, la matrice $\mathbf{Q}$ est obtenue depuis la décomposition en valeurs singulières de la matrice $\mathbf{A}^f = [\mathbf{A_x}^f\mathbf{A_y}^f]$, en choisissant les $p$ vecteurs singuliers à gauche dominants :

$$\mathbf{A}^f = \mathbf{Q}\Sigma_p\mathbf{V}_p^T + \mathbf{U}\bar{\Sigma}_p\mathbf{V}_{n-p}^T,$$

avec $\mathbf{Q} \in \mathbb{R}^{m\times p}$, $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_p$ ; $\mathbf{V}_p \in \mathbb{R}^{n\times p}$, $\mathbf{V}_p^T\mathbf{V}_p = \mathbf{I}_p$ ; $\Sigma_p \in \mathbb{R}^{p\times p}$ diagonale ; $\mathbf{U} \in \mathbb{R}^{m\times(m-p)}$, $\mathbf{U}^T\mathbf{U} = \mathbf{I}_{m-p}$ ; $\mathbf{V}_{n-p} \in \mathbb{R}^{n\times(n-p)}$, $\mathbf{V}_{n-p}^T\mathbf{V}_{n-p} = \mathbf{I}_{n-p}$ ; $\bar{\Sigma}_p \in \mathbb{R}^{(m-p)\times(n-p)}$ diagonale ; $\mathbf{Q}^T\mathbf{U} = 0$ et $\mathbf{V}_p^T\mathbf{V}_{n-p} = 0$.

Ayant obtenue cette décomposition, la matrice de pseudo-anomalies est définie par

$$\mathbf{A}_Q^f = \sqrt{n-1}\mathbf{Q}\mathbf{L}\mathbf{V}_p^T. \tag{3.29}$$

Nous pouvons faire les remarques suivantes.

1. Le choix de $p = \mathrm{rg}(\mathbf{A}^f)$ conduit à $\mathbf{A}^f = \mathbf{Q}\Sigma_p\mathbf{V}_p^T$.
2. Si $\mathrm{rg}(\mathbf{A}^f) \geq p$, il vient $\mathbf{A}^f\mathbf{e} = 0 \Rightarrow \mathbf{V}_p^T\mathbf{e} = 0 \Rightarrow \mathbf{A}_Q^f\mathbf{e} = 0$. Nous obtenons ainsi que la somme des colonnes de $\mathbf{A}_Q^f$ est nulle.
3. Avec $p = \mathrm{rank}(\mathbf{A}^f)$, il vient

$$\|\mathbf{A}^f - \mathbf{A}_Q^f\|_F = \|\Sigma_p - (\sqrt{n-1})\mathbf{L}\|_F.$$

   De nouveau, il n'y a aucune garantie que l'ensemble $\mathbf{E}_Q^f$ soit constitué d'états physiquement réalisables. Néanmoins, nous pouvons envisager que les bruits de faible variance vont conduire à de faibles modifications de la matrice $\mathbf{L}$ relativement à la matrice diagonale $\Sigma_p$. Dans ce cas, les déséquilibres engendrés par ces modifications pourraient être légers.

Nous avons évalué les performances de cette approche dans le cadre d'expériences jumelles réalisées avec un modèle quasi-géostrophique. La Figure 3.10 présente la solution de référence à un instant donné, ainsi que le réseau d'observations utilisé. Un exemple de résultats est donné dans la Table 3.1. Nous comparons les meilleurs valeurs d'erreur RMS et de dispersion, moyennées sur les cycles 52-301, pour différentes valeurs d'inflation, obtenues avec un ETKF utilisant les anomalies bruitées $\mathbf{A}^f$ (Basic ETKF) et celles obtenues après optimisation $\mathbf{A}_\mathbf{Q}^f$ (**QL**-based ETKF). Une analyse plus exhaustive des performances est disponible dans [70].

| | Basic ETKF | Basic ETKF | Basic ETKF | **QL**-based ETKF | **QL**-based ETKF | **QL**-based ETKF |
|---|---|---|---|---|---|---|
| Inflation | 1.03 | 1.04 | 1.05 | 1.08 | 1.09 | 1.1 |
| RMS | 0.949 | 0.954 | 0.954 | 0.917 | **0.910** | 0.922 |
| Dispersion | 0.996 | 1.065 | 1.135 | 1.019 | 1.041 | 1.098 |

TABLE 3.1 – Expérience avec $n_x = 10$, $n_y = 15$ et $\sigma_c^2 = 1$ : erreur RMS de l'analyse et dispersion de l'ensemble moyennés sur les cycles 52-301 pour différentes valeurs d'inflation. Table tirée de Gratton *et al.* (accepté). [70]

Les résultats obtenus illustrent les avantages à utiliser des matrices de covariance des erreurs de prévision optimisées, dans le cas de bruit sur une partie des membres de l'ensemble. Le calcul de l'estimateur du maximum de vraisemblance de la matrice de covariance des erreurs de prévision conduit à une erreur RMS d'analyse la plus faible dans les expériences réalisées. De plus, l'augmentation de la variance du bruit ou celle de la taille de l'ensemble non bruité par rapport à la taille de l'ensemble bruité engendre des problèmes numériques qui peuvent être associés à une augmentation du conditionnement de la Hessienne de $h$ au cours de la minimisation. De meilleures performances pourraient être obtenues par l'utilisation de stratégies de préconditionnement appropriées.

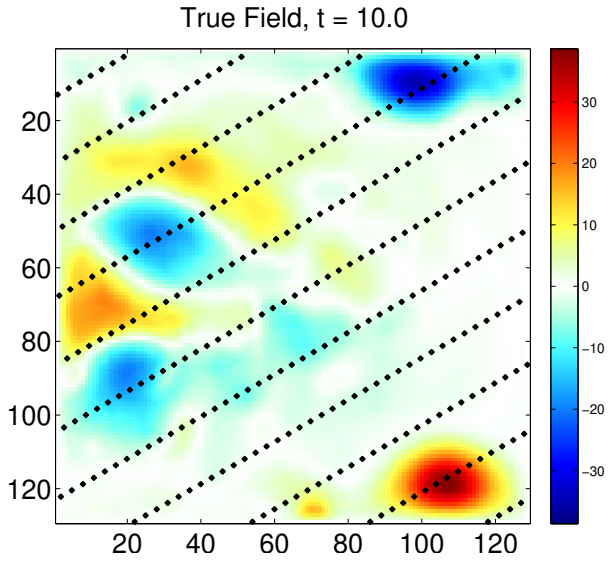



FIGURE 3.10 – Modèle quasi-géostrophique : fonction de courant $\psi$ et réseau d'observation (points noir). Figure tirée de Gratton *et al.* (accepté) [70]

La suite de ces travaux se situe naturellement dans l'enrichissement de la modélisation du problème afin d'attaquer des problèmes réalistes, qui vont au delà du simple stockage en précision dégradée des champs nécessaires à l'étape d'analyse par filtrage de Kalman d'ensemble. La première étape me semble naturellement se tourner vers les approches multi-niveaux. Ceci sera développé plus en détails dans la section 5.1.

## Publications associées

[112] J.S. Pelc, E. Simon, L. Bertino, G. El Serafy, A. Heemink : Application of model reduced 4D-Var to a 1D ecosystem model, *Ocean Model.*, 57-58, 43-58, 2012 ;

[69] S. Gratton, M. Rincon-Camacho, E. Simon, Ph. L. Toint : Observation Thinning in Data Assimilation Computations, *EURO J. Comput. Optim.*, 3, 31-51, 2015 ;

[7] R. Benshila, G. Thoumyre, M. Al Najar, G. Abessolo, R. Almar, E. Bergsma, G. Hugonnard, L. Labracherie, B. Lavie, T. Ragonneau, E. Simon, B. Vieuble, D. Wilson : A deep learning approach for estimation of the nearshore bathymetry, *J. Coastal Research*, 95 (sp1), 1011-1015, 2020 ;

[70] S. Gratton, E. Simon, D. Titley-Peloquin : Covariance matrix estimation for ensemble-based Kalman filters with multiple ensembles, *Math. Geosci.*, accepté ;

[119] A. Rouviere, L. Pascal, F. Méry, E. Simon, S. Gratton : Neural prediction model for transition onset of a boundary layer in presence of 2D surface defect, soumis.

# Chapitre 4

# Des problèmes aux moindres carrés non-linéaires pondérés

**Sommaire**


**4.1 Des normes $\|\|_{\mathbf{D}^{-1}}$ . . . . . 49**
4.1.1 Un préconditionnement non trivial . . . . . 50
4.1.2 Formulation point-selle pour l'algorithme 4D-VAR à contraintes faibles : quel critère d'arrêt ? . . . . . 53
**4.2 Des calculs inexacts pour une convergence garantie . . . . . 57**
4.2.1 Un algorithme du gradient conjugué inexact pour la minimisation de quadratique convexe . . . . . 57
4.2.2 Une variante de GMRES avec produits scalaires inexacts . . . . . 61
**4.3 Des normes non-standards . . . . . 66**
4.3.1 Régularisation en norme $\|\|_p$ en assimilation variationnelle de données . 66
4.3.2 Vers l'introduction de normes non différentiables . . . . . 69


*Ce chapitre porte sur des axes de recherche récents, auxquels je me suis intéressé ces cinq dernières années. Ces travaux ont été principalement réalisés dans le cadre de collaborations internationales avec des chercheurs expérimentés, ainsi que dans le cadre d'une thèse de doctorat. Ils renvoient à des développements méthodologiques associés à la résolution de variantes de l'algorithme 4D-Var, vues comme des problèmes aux moindres carrés non-linéaires pondérés et régularisés.*

## 4.1 Des normes $\|\|_{\mathbf{D}^{-1}}$

Un exemple de modélisation du problème d'assimilation de données est la formulation dite 4D-Var à contraintes faibles [150, 141, 142, 134, 135], qui s'écrit :

$$\min_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\sum_{j=0}^{N_{sw}} \left\|\mathcal{H}_j(\mathbf{x}_j) - \mathbf{y}_j\right\|^2_{\mathbf{R}_j^{-1}} + \frac{1}{2}\sum_{j=1}^{N_{sw}} \|\mathbf{x}_j - \mathcal{M}_j(\mathbf{x}_{j-1})\|^2_{\mathbf{Q}_j^{-1}} \tag{4.1}$$

avec $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N_{sw}})^T \in \mathbb{R}^n$ le vecteur d'état du système à différents instants, $(\mathcal{M}_j)$ le modèle décrivant la dynamique du système sur une fenêtre temporelle (modélisée par des

équations aux dérivées partielles potentiellement stochastiques), $\mathbf{y}_j \in \mathbb{R}^{m_j}$ les observations de l'état du système sur cette même fenêtre, et $(\mathcal{H}_j)_{j=0:N_{sw}}$ les opérateurs d'observations faisant le lien entre les variables d'état et les observations. Les matrices $\mathbf{B}$, $(\mathbf{R_j})_{j=0:N_{sw}}$ et $(\mathbf{Q_j})_{j=1:N_{sw}}$ visent à tenir compte des incertitudes du système via la pondération des différentes équations résiduelles. Pour le cas de la météorologie, de l'océanographie et des sciences du climat, ces problèmes sont facilement de l'ordre de $n \approx 10^9$ et $\sum_{j=0}^{N_{sw}} m_j \approx 10^7$. Les méthodes classiques pour la résolution de ce type de problèmes d'optimisation (algorithmes de Gauss-Newton, de Levenberg-Marquardt, et plus généralement de région de confiance ou de régularisation) conduisent à des séquences de résolution de sous-problèmes quadratiques, se réduisant dans le cas convexe à la résolution de systèmes linéaires de grande dimension. De surcroit, les évaluations de $J$ et de ses dérivées requièrent également la résolution de systèmes linéaires via l'emploi de normes pondérées du type $\|.\|_{\mathbf{D}^{-1}}$. Ceci rend impératif le développement d'algorithmes performants, ainsi que leur preuve de convergence (à minima vers un point critique de $J$).

### 4.1.1 Un préconditionnement non trivial

Dans le cas de la formulation dite "*state*" de l'algorithme 4DVAR à contraintes faibles [134], en notant $\mathbf{M}_j$ (resp. $\mathbf{H}_j$) le modèle linéarisé de $\mathcal{M}_j$ (resp. $\mathcal{H}_j$) au voisinage de l'ébauche (on ne s'intéresse ici qu'à la première itération de l'algorithme de Gauss-Newton), le sous-problème quadratique s'écrit

$$\min_{\delta\mathbf{x}\in\mathbb{R}^n} q_{st}(\delta\mathbf{x}) = \frac{1}{2}\|\mathbf{L}\delta\mathbf{x} - \mathbf{b}\|^2_{\mathbf{D}^{-1}} + \frac{1}{2}\|\mathbf{H}\delta\mathbf{x} - \mathbf{d}\|^2_{\mathbf{R}^{-1}} \tag{4.2}$$

avec $\delta\mathbf{x} = (\delta\mathbf{x_0}, \delta\mathbf{x_1}, \ldots, \delta\mathbf{x_{N_{sw}}})^{\mathbf{T}}$ le vecteur d'incrément.

$$\mathbf{L} = \begin{pmatrix} \mathbf{I}_n & & & & \\ -\mathbf{M}_1 & \mathbf{I}_n & & & \\ & -\mathbf{M}_2 & \mathbf{I}_n & & \\ & & \ddots & \ddots & \\ & & & -\mathbf{M}_{N_{sw}} & \mathbf{I}_n \end{pmatrix} \tag{4.3}$$

et

$$\mathbf{H} = \mathrm{diag}(\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N_{sw}}),\ \mathbf{D} = \mathrm{diag}(\mathbf{B}, \mathbf{Q}_1, \ldots, \mathbf{Q}_{N_{sw}}),\ \mathbf{R} = \mathrm{diag}(\mathbf{R}_0, \mathbf{R}_1, \ldots, \mathbf{R}_{N_{sw}}). \tag{4.4}$$

Les vecteurs $\mathbf{d}$ et $\mathbf{b}$ renvoient aux concaténations des innovations et erreurs d'ébauche et de modèle sur la fenêtre d'assimilation. Ils s'écrivent

$$\begin{aligned} \mathbf{d} &= (\mathbf{y}_0 - \mathcal{H}_0(\mathbf{x}_0), \mathbf{y}_1 - \mathcal{H}_1(\mathbf{x}_1), \ldots, \mathbf{y}_{N_{sw}} - \mathcal{H}_{N_{sw}}(\mathbf{x}_{N_{sw}}))^T, \\ \mathbf{b} &= (\mathbf{x}_0 - \mathbf{x}_b, \mathcal{M}_1(\mathbf{x}_0) - \mathbf{x}_1, \ldots, \mathcal{M}_{N_{sw}}(\mathbf{x}_{N_{sw}-1}) - \mathbf{x}_{N_{sw}})^T, \end{aligned}$$

Par convexité de $q_{st}$, ce problème est équivalent à la résolution du système linéaire

$$(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\delta\mathbf{x} = \mathbf{L}^T\mathbf{D}^{-1}\mathbf{b} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} \tag{4.5}$$

En supposant que le terme $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$ domine devant $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ dans la matrice du système (4.5), ceci conduit à des préconditionneurs de la forme

$$\mathbf{P}^{-1} = \tilde{\mathbf{L}}^{-1}\,\mathbf{D}\,\tilde{\mathbf{L}}^{-T}, \tag{4.6}$$

avec $\tilde{\mathbf{L}}$ une approximation de $\mathbf{L}$ (4.3). Typiquement, cette approximation est construite par approximation des opérateurs $\mathbf{M}_j$ (intégration numérique du modèle linéaire tangent du temps $t_{j-1}$ au temps $t_j$). Le matrice du système préconditionné s'écrit alors

$$(\tilde{\mathbf{L}}^{-1}\mathbf{D}\tilde{\mathbf{L}}^{-T})(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}). \tag{4.7}$$

Dans le cadre d'une collaboration avec S. Gratton, S. Gürol et Ph.L. Toint [67], nous nous sommes intéressés à l'effet du préconditionnement des moindres carrés linéaires pondérés à l'aide d'une approximation de la matrice du modèle, tel que décrit précédemment. En effet, il est observé une certaine inefficacité de cette approche selon les configurations expérimentales.

Dans le cas général d'un problème aux moindres carrés linéaires pondérés, avec une matrice $\mathbf{A}$ inversible,

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2_{\mathbf{W}^{-1}}, \tag{4.8}$$

la solution est caractérisée par les équations normales

$$(\mathbf{A}^T\mathbf{W}^{-1}\mathbf{A})\mathbf{x} = \mathbf{A}^T\mathbf{W}^{-1}\mathbf{b}. \tag{4.9}$$

Notons $\tilde{\mathbf{A}} \in \mathbb{R}^{n\times n}$ une approximation inversible de $\mathbf{A} \in \mathbb{R}^{n\times n}$. En préconditionnant le système (4.9) par

$$\mathbf{P} = \tilde{\mathbf{A}}^T\mathbf{W}^{-1}\tilde{\mathbf{A}} \tag{4.10}$$

il vient

$$\mathbf{P}^{-1}(\mathbf{A}^T\mathbf{W}^{-1}\mathbf{A})\mathbf{x} = \mathbf{P}^{-1}\mathbf{A}^T\mathbf{W}^{-1}\mathbf{b} = (\tilde{\mathbf{A}}^{-1}\mathbf{W}\tilde{\mathbf{A}}^{-T})\mathbf{A}^T\mathbf{W}^{-1}\mathbf{b}. \tag{4.11}$$

Nous avons alors démontré le théorème suivant, qui borne le spectre de la matrice du système préconditionné en fonction de l'erreur d'approximmation sur $\mathbf{A}$ et du conditionnement de la matrice de poids $\mathbf{W}$ :

**Theorem 4.1.1.** *Soit $(\mathbf{A}, \tilde{\mathbf{A}}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$ des matrices inversibles, et soit $\mathbf{W} \in \mathbb{R}^{n\times n}$ symétrique définie positive. On note*

$$\mathbf{A_p} \stackrel{\text{def}}{=} (\tilde{\mathbf{A}}^{-1}\mathbf{W}\tilde{\mathbf{A}}^{-T})(\mathbf{A}^T\mathbf{W}^{-1}\mathbf{A}). \tag{4.12}$$

*Le spectre de la matrice $\mathbf{A_p}$, noté $\sigma(\mathbf{A_p})$, vérifie*

$$\sigma(\mathbf{A_p}) \subset \mathcal{B}\Big(1, (1+\kappa_2(\mathbf{W}))\|\mathbf{E}\|_2 + \kappa_2(\mathbf{W})\|\mathbf{E}\|_2^2\Big) \tag{4.13}$$

*avec $\mathbf{E} \stackrel{\text{def}}{=} \mathbf{A}\tilde{\mathbf{A}}^{-1} - \mathbf{I}_n$ l'erreur d'approximation, $\kappa_2(\mathbf{W}) = \|\mathbf{W}\|_2\|\mathbf{W}^{-1}\|_2$ le conditionnement de $\mathbf{W}$ en norme 2, et $\mathcal{B}(a, r)$ la boule fermée de centre $a$ et de rayon $r$.*

Une conséquence importante de ce théorème porte sur le conditionnement de la matrice du système préconditionné. En effet, vouloir imposer $M > 0$ comme majorant de ce conditionnement résulte en une majoration de l'erreur d'approximation :

$$\|\mathbf{E}\|_2 \leq \frac{-(1+\kappa_2(\mathbf{W})) + \sqrt{(1+\kappa_2(\mathbf{W}))^2 + 4\kappa_2(\mathbf{W})\frac{M-1}{M+1}}}{2\kappa_2(\mathbf{W})} \stackrel{\text{def}}{=} g(\kappa_2(\mathbf{W}), M). \tag{4.14}$$

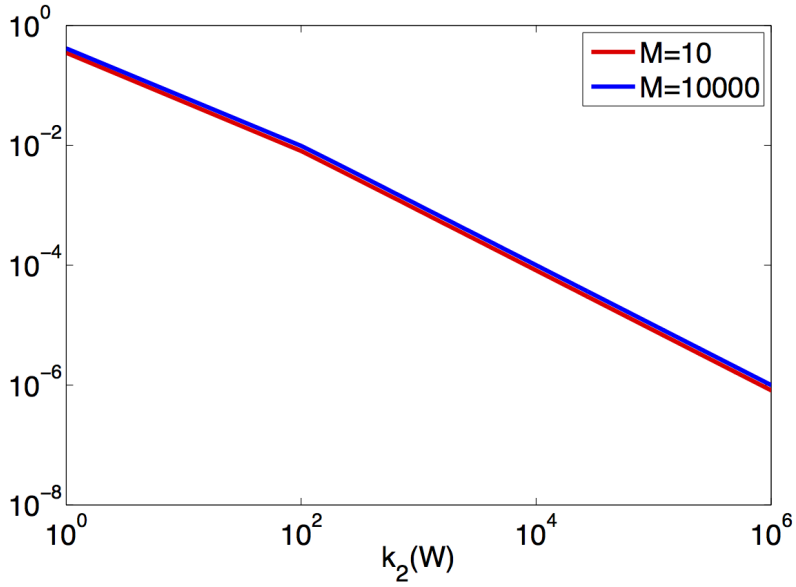



FIGURE 4.1 – Majorant $g$ de l'erreur $\|\mathbf{E}\|_2$ (4.14) en fonction de $\kappa_2(\mathbf{W})$ (échelle logarithmique). Figure tirée de Gratton *et al.* (2018) [67].

La Figure 4.1 représente l'évolution de la borne $g$ (4.14) sur la perturbation de $\mathbf{A}$ en fonction du choix du conditionnement visé ($M$) pour la matrice du système préconditionné $\mathbf{A_p}$ et du conditionnement de $\mathbf{W}$. On note ainsi que, même pour des bornes supérieures élévées sur le conditionnement de $\mathbf{A_p}$, il est nécessaire de spécifier une faible erreur d'approximation sur $\mathbf{A}$, et ce d'autant plus que le conditionnement de $\mathbf{W}$ est élevé.

Appliqués à la formulation "*state*" de l'algorithme 4DVAR contraintes faibles, et en se focalisant uniquement sur la partie $(\tilde{\mathbf{L}}^{-1}\mathbf{D}\tilde{\mathbf{L}}^{-T})(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$ du système (4.7), ces résultats conduisent au corollaire suivant, pour les deux approximations du modèle linéaire tangent $\tilde{\mathbf{M}}_j$ couramment utilisés pour maintenir des propriétés de parallélisme en temps pour $\tilde{\mathbf{L}}^{-1}$ , à savoir $\tilde{\mathbf{M}}_j = \mathbf{0}$ et $\tilde{\mathbf{M}}_j = \mathbf{I}_n$.

**Corollary 4.1.2.** *Soient* $\mathbf{L}$ *et* $\mathbf{M}_j$ *définis par (4.3), et soit* $\tilde{\mathbf{L}}$ *une approximation de* $\mathbf{L}$ *définie depuis* $\tilde{\mathbf{M}}_\mathbf{j} \in \{\mathbf{0}, \mathbf{I}_n\}$ *pour* $j = 1, \ldots, N_{sw}$. *Notons* $\mathbf{A_p} = (\tilde{\mathbf{L}}^{-1}\mathbf{D}\tilde{\mathbf{L}}^{-T})(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$ *la matrice du système préconditionné*[1].

*Il vient*

$$\sigma(\mathbf{A_p}) \subset \mathcal{B}(1, (1 + \kappa_2(\mathbf{D}))\rho + \kappa_2(\mathbf{D})\rho^2) \tag{4.15}$$

*avec*

$$\rho = \begin{cases} \max_{j=1,\ldots,N_{sw}} \sigma_{\max}(\mathbf{M}_j) & si \quad \tilde{\mathbf{M}}_\mathbf{j} = \mathbf{0} \ \ (j = 1 \ldots, N_{sw}), \\ \sqrt{\frac{(N_{sw}+1)(N_{sw}+2)}{2}} \left[\max_{j=1,\ldots,N_{sw}} \sigma_{\max}(\mathbf{I}_n - \mathbf{M}_j)\right] & si \quad \tilde{\mathbf{M}}_j = \mathbf{I}_n \ (j = 1, \ldots, N_{sw}). \end{cases}$$

L'obtention d'une matrice bien conditionnée requiert ainsi des hypothèses spécifiques sur les modèles dynamiques à l'intérieur de chacune des sous-fenêtres. Le choix de $\tilde{\mathbf{M}}_j = \mathbf{0}$ parait

1. le terme d'observation $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ est supposé négligeable devant $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$

donc peu réaliste, sauf pour des cas bien particuliers (modèle linéaire tangent lui-même proche d'être nul). Le second choix $\tilde{M}_j = I_n$ parait plus judicieux, dès lors que la dynamique du modèle linéaire tangent reste limitée, ce qui peut notamment être le cas pour de très courtes sous-fenêtres. On serait alors tenter d'utiliser un grand nombre de sous-fenêtres ($N_{sw}$ grand), même si dans ce cas , la variable $\rho^2$ tendrait à grandir en $N_{sw}^2$, augmentant ainsi l'incertitude sur le conditionnement de $\mathbf{A_p}$. Il apparaît ainsi plus naturel de chercher à améliorer la qualité de $\tilde{\mathbf{M}}_j$ en tant qu'approximation de $\mathbf{M}_j$, mais il reste difficile de sélectionner de bonnes approximations qui préservent un calcul parallèle efficace de $\tilde{\mathbf{L}}^{-1}$ [65].

Enfin, pour nuancer ces conclusions, il faut également tenir compte du fait que les bornes fournies sur le conditionnement sont pessimistes par nature, et que le terme d'observation $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, ignoré dans cette étude, peut influer significativement sur le conditionnement de $\mathbf{A_p}$.

### 4.1.2 Formulation point-selle pour l'algorithme 4D-VAR à contraintes faibles : quel critère d'arrêt ?

En parallèle de ces travaux sur les préconditionneurs utilisés dans la formulation "*state*" de l'algorithme 4DVAR à contraintes faibles, nous nous sommes également intéressés à la formulation dite "*point-selle*" [55, 54]. En reformulant le problème comme un problème d'optimisation avec contraintes, celle-ci conduit à la résolution d'un système linéaire de type point-selle dépendant des matrices $\mathbf{B}$, $(\mathbf{R}_j)_{j=0:N_{sw}}$ et $(\mathbf{Q}_j)_{j=0:N_{sw}}$ et non plus de leurs inverses. En effet, le problème d'optimisation peut s'écrire :

$$\min_{(\boldsymbol{\delta}\mathbf{p},\boldsymbol{\delta}\mathbf{w},\boldsymbol{\delta}\mathbf{x})} \frac{1}{2}\|\boldsymbol{\delta}\mathbf{p}-\mathbf{b}\|^2_{\mathbf{D}^{-1}} + \frac{1}{2}\|\boldsymbol{\delta}\mathbf{w}-\mathbf{d}\|^2_{\mathbf{R}^{-1}} \tag{4.16}$$

$$\text{subject to} \quad \boldsymbol{\delta}\mathbf{p} = \mathbf{L}\boldsymbol{\delta}\mathbf{x} \quad \text{and} \quad \delta\mathbf{w} = \mathbf{H}\boldsymbol{\delta}\mathbf{x} \tag{4.17}$$

Le Lagrangien associé à ce problème s'écrit

$$\mathcal{L}(\delta\mathbf{w},\delta\mathbf{p},\delta\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = \frac{1}{2}\|\boldsymbol{\delta}\mathbf{p}-\mathbf{b}\|^2_{\mathbf{D}^{-1}} + \frac{1}{2}\|\boldsymbol{\delta}\mathbf{w}-\mathbf{d}\|^2_{\mathbf{R}^{-1}} \tag{4.18}$$

$$+ \boldsymbol{\lambda}^T(\delta\mathbf{p}-\mathbf{L}\delta\mathbf{x}) + \boldsymbol{\mu}^T(\delta\mathbf{w}-\mathbf{H}\delta\mathbf{x}). \tag{4.19}$$

De nouveau, par convexité du problème, la solution de ce problème est caractérisée par la condition d'optimalité du premier ordre :

$$\mathbf{D}^{-1}(\mathbf{L}\delta\mathbf{x}-\mathbf{b}) + \boldsymbol{\lambda} = 0; \tag{4.20}$$

$$\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}-\mathbf{d}) + \boldsymbol{\mu} = 0; \tag{4.21}$$

$$\mathbf{L}^T\boldsymbol{\lambda} + \mathbf{H}^T\boldsymbol{\mu} = 0; \tag{4.22}$$

ce qui se reformule sous forme matricielle :

$$\begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{L} \\ \mathbf{0} & \mathbf{R} & \mathbf{H} \\ \mathbf{L}^{\mathrm{T}} & \mathbf{H}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \\ \boldsymbol{\delta}\mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{d} \\ \mathbf{0} \end{pmatrix} \tag{4.23}$$

On est donc amenés à résoudre un système linéaire de type point-selle, ne faisant plus intervenir les inverses de $\mathbf{D}$ et $\mathbf{R}$, contrairement à la formulation "*state*". Ceci conduit à l'algorithme point-selle (Algorithme 2).

**Algorithm 2** Algorithme point-selle [55, 54]

1: **while** Non convergence **do**
2: Calcul de $J(\mathbf{x}_k)$ et $\mathbf{g}_k = \nabla_{\mathbf{x}} J(\mathbf{x}_k)$
3: Calcul de $\delta\mathbf{x}_k$ : application de l'algorithme GMRES [120] pour la résolution de (4.23), avec le critère d'arrêt $\|r(\boldsymbol{\lambda}, \boldsymbol{\mu}, \delta\mathbf{x})\| \leq \varepsilon_r(\|\mathbf{b}\| + \|\mathbf{d}\|)$ or $j = n_{inner}$, avec $r(\boldsymbol{\lambda}, \boldsymbol{\mu}, \delta\mathbf{x})$ le résidu associé à (4.23).
4: $\delta\mathbf{x}_{k+1} = \mathbf{x}_k + \delta\mathbf{x}_k$
5: **end while**

Nous nous sommes intéressés, dans un premier temps, à l'évaluation des performances de cet algorithme via des expériences jumelles réalisées avec un modèle de Burgers 1D, dont la figure 4.2 représente la référence (état vrai) et les observations au début et à la fin de la fenêtre d'assimilation [66].

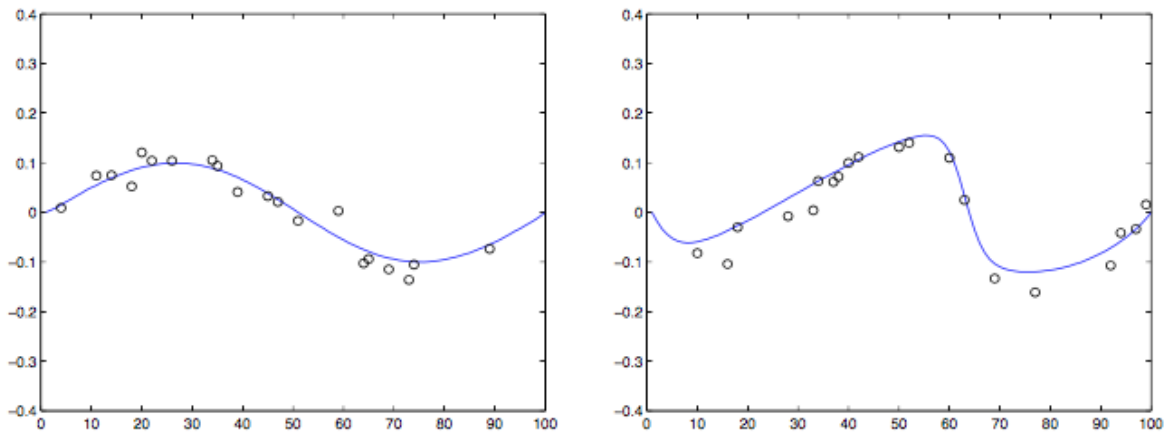



FIGURE 4.2 – Modèle de Burgers 1D : référence (bleu) et observations (cercle noir) en début et fin de fenêtre d'assimilation. Figure tirée de Gratton *et al.* (2018) [66].

La Figure 4.3 représente l'évolution de la fonction coût non-linéaire $J$ (courbes continues) et de la quadratique associée $q_{st}$ (courbes pointillées) lors des itérations internes de Gauss-Newton (résolution du système linéaire par GMRES), et ce pour 10 itérations externes, et un nombre maximum d'itérations internes $n_{inner} = 50$. A noter que ces deux quantités ne sont jamais calculées à l'intérieur d'une boucle interne avec l'algorithme point-selle, afin d'éviter l'évaluation de normes $\|\|_{\mathbf{D}^{-1}}$. Les couleurs bleu et noir traduisent l'utilisation de deux préconditionneurs différents (cf [53, 66] pour plus de détails). Dans un premier temps, nous remarquons une bonne adéquation entre la décroissance de $q_{st}$ et $J$, tant que les pas $\delta\mathbf{x}_k$ restent d'amplitude modérée. Néanmoins, il n'est pas possible de faire décroître significativement la valeur de $J$, qui après 500 itérations internes, reste encore loin de l'optimum (valeur $\sim 63.11$). On note également que les évolutions de $q_{st}$ et $J$ ne sont pas nécessairement monotones au cours des itérations, ce qui rend l'arrêt de l'algorithme point-selle sur la base d'un nombre d'itérations internes maximum incertain (quid de la valeur de ces fonctions lors de l'arrêt ?), et ce d'autant plus que ces quantités ne sont jamais calculées. Au contraire de certaines méthodes de Krylov, telles que l'algorithme du gradient conjugué [80], qui garantissent, au cours des itérations, la décroissance de la quadratique

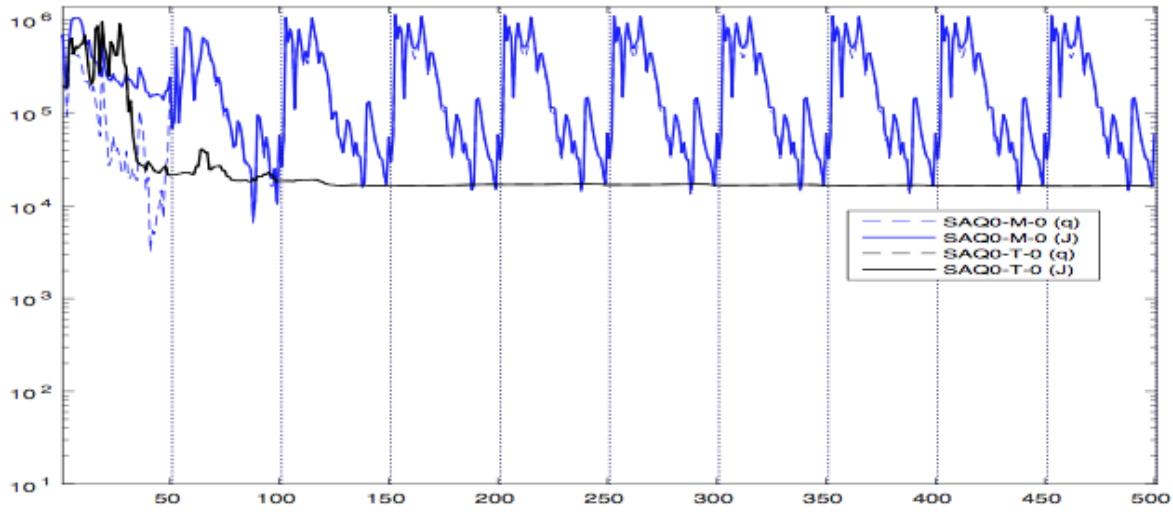



FIGURE 4.3 – Algorithme point-selle (Algorithm 2). Courbes continues : fonction coût non-linéaire $J$ ; courbes pointillées : approximation quadratique $q_{st}$ (algorithme de Gauss-Newton). Figure tirée de Gratton *et al.* (2018) [66].

$q_{st}$ sur des espaces de dimension croissante, les itérations de GMRES [120] portent sur le résidu du système point-selle (Eq. 4.23), qui inclue les équations résiduelles associées aux multiplicateurs de Lagrange. Il n'y a donc *a priori* aucune garantie que les directions $\delta\mathbf{x}_k$, obtenues au cours des itérations, conduisent à une décroissance monotone de $q_{st}$. Deux options s'offrent alors pour arrêter l'algorithme GMRES, sans compromettre la décroissance obtenue sur $q_{st}$.

1. Arrêter l'algorithme GMRES lorsque la précision machine est atteinte : à optimalité (et seulement à optimalité), les directions $\delta\mathbf{x}_k$ obtenues depuis les formulations "*state*" et "*point-selle*" sont les mêmes.
2. Surveiller l'évolution de $q_{st}$ au fil des itérations, et s'arrêter quand une décroissance "suffisante" est atteinte vis-à-vis de $q_{st}(0)$.

Nous en avons donc proposé une variante "globalisée" (Algorithm 3), basée sur cette idée, pour laquelle nous avons démontré la convergence vers un point critique de $J$. Celle-ci repose à la fois sur le critère d'arrêt, garantissant de "bonnes" propriétés pour la direction $\delta\mathbf{x}_k$, et sur l'utilisation d'une recherche linéaire pour le calcul de la longueur de pas (cf [66] pour plus de détails). Quelques remarques concernant cet algorithme.

1. La suite $\{\theta_j\}$ n'intervient pas dans la preuve de convergence : elle vise à assurer que l'algorithme GMRES ne s'arrête pas trop "tôt" ;
2. Le critère d'arrêt nécessite le calcul de la quadratique en la direction courante $q_{st}(\delta\mathbf{x})$ : ceci nécessite de nouveau l'évaluation de normes $\|\|_{\mathbf{D}^{-1}}$ ; et le surcoût associé va notamment dépendre de la fréquence, à laquelle la décroissance de la quadratique est évaluée ;
3. Selon le problème, l'algorithme GMRES peut être amené à réaliser un nombre important d'itérations pour atteindre la décroissance spécifiée sur la quadratique, ou la précision machine, entraînant de nouveau un surcoût par rapport à la version originelle (Algorithme 2).

**Algorithm 3** Algorithme point-selle globalisé [66]

1: **while** Non convergence **do**
2: Calcul de $J(\mathbf{x}_k)$ et $\mathbf{g}_k = \nabla_{\mathbf{x}} J(\mathbf{x}_k)$
3: Calcul de $\delta\mathbf{x}_k$ : application de l'algorithme GMRES [120] pour la résolution de (4.23), avec le critère d'arrêt suivant :

$$\text{mod } (j, \ell) = 0 \quad \text{et} \quad q_{st}(0) - q_{st}(\delta\mathbf{x}) \geq \max\left[\epsilon_q \min[1, \|\mathbf{g}_k\|^2], \theta_j\right],$$

ou si le système est résolu à la précision machine sur l'équation résiduelle.
4: Calcul d'une longueur de pas $\alpha_k$ : recherche linéaire (algorithme du *backtracking*) le long de la direction $\delta\mathbf{x}_k$ ;
5: $\delta\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \delta\mathbf{x}_k$
6: **end while**

avec $\ell \in \mathbb{N}$ la fréquence d'évaluation de la quadratique $q_{st}$ et $\{\theta_j\}$ une suite qui tend vers 0.

La figure 4.4 représente l'évolution de la fonction coût non-linéaire $J$ (courbes continues) et de la quadratique associée $q_{st}$ (courbes pointillées) lors des itérations internes de Gauss-Newton, avec l'algorithme point-selle globalisé, de nouveau pour 10 itérations externes. Cette fois-ci, le nombre d'itération est spécifié suffisamment grand pour que l'algorithme GMRES renvoie une direction $\delta\mathbf{x}_k$ garantissant la convergence vers un point critique de $J$. Pour cette expérience, l'évaluation de la quadratique $q_{st}$ a lieu toutes les 25 itérations. On note ainsi que le minimum est bien atteint en 10 itérations externes pour l'un des préconditionneurs (courbe bleu). Néanmoins, et comme anticipé, ceci se fait au prix d'un nombre conséquent d'itérations, incluant pour certaines l'évaluation de normes $\|\|_{\mathbf{D}^{-1}}$.

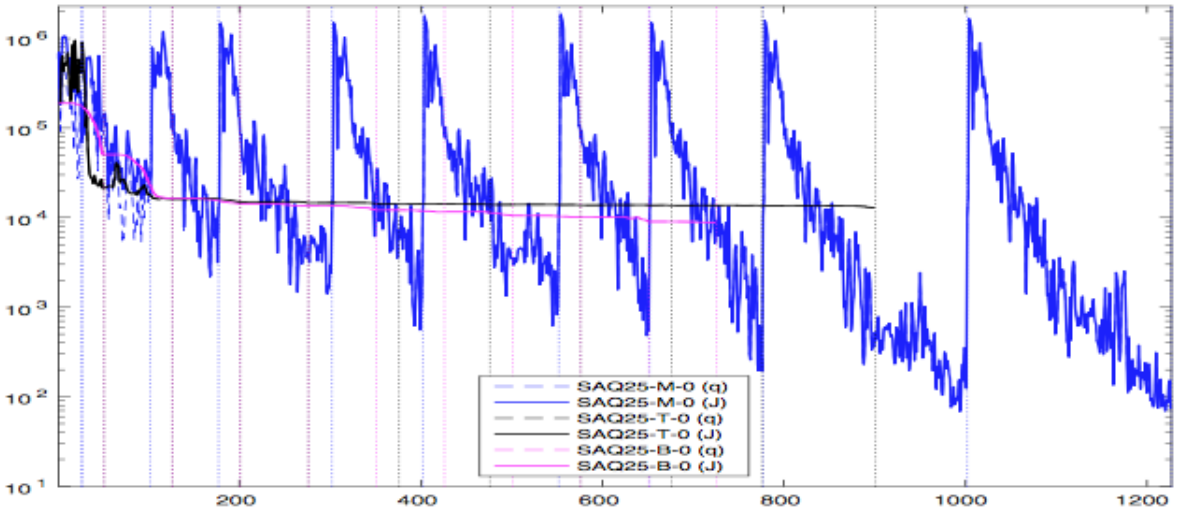



FIGURE 4.4 – Algorithme point-selle régularisé (Algorithm 3). Courbes continues : fonction coût non-linéaire $J$ ; courbes pointillées : approximation quadratique $q_{st}$ (algorithme de Gauss-Newton). Figure tirée de Gratton *et al.* (2018) [66].

Enfin, pour nuancer ces conclusions, la formulation "*point-selle*", même sans présenter de

garantie de convergence, peut fournir des résultats raisonnables selon le problème applicatif rencontré ([54, 53]). Néanmoins, il est naturel de penser que dans ce cas, le surcoût associé à la variante globalisée de l'algorithme soit marginal, dès lors que l'évaluation de l'évolution de $q_{st}$ a lieu peu fréquemment. Même dans ces conditions, il semble préférable d'utiliser un algorithme reposant sur une base théorique certaine.

## 4.2 Des calculs inexacts pour une convergence garantie

Devant la nécessité de devoir résoudre des systèmes linéaires pour l'évaluation de $J$, $q_{st}$ et/ou de leurs dérivées, nous nous sommes intéressés à la possibilité de résoudre ces problèmes de manière approchée, sans pour autant compromettre la convergence de l'algorithme d'optimisation vers un point critique. Ceci a conduit à des collaborations avec S. Gratton, D. Titley-Peloquin et Ph.L. Toint. Dans le cas de l'utilisation d'un algorithme de région de confiance, ceci peut se faire au moyen du contrôle de l'erreur sur les évaluations de la fonction à minimiser, et de ses dérivées, au cours des itérations. Un exemple d'algorithme utilisant une telle stratégie de précision dynamique est décrit dans [32, Section 10.6]. Dans notre cas, nous nous sommes intéressés à la résolution des sous-problèmes des algorithmes de Gauss-Newton (minimisation d'une quadratique convexe par un algorithme de gradient conjugué) et de la variante point-selle de l'algorithme 4D-var à contraintes faibles (résolution d'un système linéaire non symétrique par l'algorithme GMRES), lorsque certaines opérations sont réalisées de manière inexacte.

### 4.2.1 Un algorithme du gradient conjugué inexact pour la minimisation de quadratique convexe

Dans le cas de la minimisation d'une quadratique convexe, telle que $q_{st}$, nous avons développé une variante de l'algorithme du gradient conjugué permettant le choix dynamique de la précision des produits matrice-vecteur effectués au cours des itérations, tout en garantissant la décroissance de la quadratique spécifiée par l'utilisateur [71]. Cette dégradation de la précision des calculs peut traduire à la fois la résolution itérative d'un autre système linéaire (comme pour le modèle quadratique $q_{st}$ de $J$) ou l'utilisation d'une arithmétique multi-précision (passage à la simple et demi-précision).

Soit le problème d'optimisation quadratique

$$\min_{\mathbf{x}\in\mathbb{R}^n} q(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x} \tag{4.24}$$

avec $\mathbf{A} \in \mathbb{R}^{n\times n}$ symétrique définie positive et $\mathbf{b} \in \mathbb{R}^n$. Ce problème est équivalent à celui de résoudre le système linéaire

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{4.25}$$

qui admet une unique solution. La résolution itérative de tels systèmes se fait généralement par l'utilisation de méthodes de Krylov, telles que l'algorithme du gradient conjugué. Dans ce contexte, il est possible de réaliser les produits matrice-vecteur de manière inexacte, tout en contrôlant dynamiquement la précision des calculs au fil des itérations, afin d'obtenir la décroissance voulue sur la norme Euclidienne du résidu [130, 137, 138, 131, 74, 45]. Néanmoins, d'un point de vue optimisation, il peut être plus intéressant de s'intéresser à l'évolution de la valeur d'une fonction non-linéaire, et/ou d'un de ses modèles (quadratiques par exemple), plutôt que celle de la norme de leur gradient. Par exemple, la théorie de convergence d'algorithmes, tels

que ceux issus du cadre des régions de confiance, repose sur des arguments de l'obtention d'une "certaine" décroissance de la fonction objectif ([32, Chapitre 6]). Dans le cas d'une quadratique strictement convexe, cette décroissance peut être mesurée depuis la norme énergie du résidu :

$$\frac{1}{2}\|r(\mathbf{x})\|^2_{\mathbf{A}^{-1}} = q(\mathbf{x}) - q(\mathbf{x}_*) \tag{4.26}$$

avec $\mathbf{x}_*$ le minimum de $q$, et la notation $\|\mathbf{x}\|_{\mathbf{M}} = \|\mathbf{M}^{1/2}\mathbf{x}\|_2$ pour $\mathbf{M}$ une matrice symétrique définie positive. En reprenant le cadre théorique du gradient conjugué avec produits matrice-vecteur inexact, et en modélisant la source d'erreur dans le produit matrice-vecteur à l'itération $k$ par une perturbation $\mathbf{E}_k$ de la matrice $\mathbf{A}$, on obtient l'Algorithme 4.

**Algorithm 4** Algorithme du gradient conjugué inexact en norme énergie

1: Initialisation : $\mathbf{x}_0 = \mathbf{0}$, $\beta_0 = \|\mathbf{b}\|_2^2$, $\mathbf{r}_0 = -\mathbf{b}$, et $\mathbf{p}_0 = \mathbf{b}$
2: **for** $k = 0, 1, \ldots,$ **do**
3: $\quad \mathbf{c}_k = (\mathbf{A} + \mathbf{E}_k)\mathbf{p}_k$
4: $\quad \alpha_k = \beta_k / \mathbf{p}_k^T \mathbf{c}_k$
5: $\quad \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$
6: $\quad \mathbf{r}_{k+1} = \mathbf{r}_k + \alpha_k \mathbf{c}_k$
7: $\quad$ **if** $\|\mathbf{r}_{k+1}\|_{\mathbf{A}^{-1}}$ est suffisamment petit **then**
8: $\quad\quad$ **Stop**
9: $\quad$ **end if**
10: $\quad \beta_{k+1} = \mathbf{r}_{k+1}^T \mathbf{r}_{k+1}$
11: $\quad \mathbf{p}_{k+1} = -\mathbf{r}_{k+1} + (\beta_{k+1}/\beta_k)\mathbf{p}_k$
12: **end for**

Un premier résultat, quantifiant l'écart relatif à la valeur optimale de la quadratique selon les normes énergie du résidu interne $\mathbf{r}_k$ et de l'erreur associée $r(\mathbf{x}_k) - \mathbf{r}_k$ est donné dans le Lemme 4.2.1.

**Lemma 4.2.1.** *Supposons qu'à l'itération $k$ du gradient conjugué inexact (Algorithme 4), le résidu interne et l'erreur sur le résidu vérifient*

$$\max\left[\|r(\mathbf{x}_k) - \mathbf{r}_k\|_{\mathbf{A}^{-1}}, \|\mathbf{r}_k\|_{\mathbf{A}^{-1}}\right] \leq \frac{\sqrt{\epsilon}}{2}\|\mathbf{b}\|_{\mathbf{A}^{-1}} \tag{4.27}$$

*avec $\epsilon > 0$. Alors*

$$|q(\mathbf{x}_k) - q(\mathbf{x}_*)| \leq \epsilon |q(\mathbf{x}_*)|. \tag{4.28}$$

*avec $\mathbf{x}_*$ le minimum de $q$.*

En définissant la norme primale-duale $\|\|_{\mathbf{A}^{-1},\mathbf{A}}$ sur $\mathbb{R}^{n\times n}$ par Eq. (4.29),

$$\|\mathbf{E}\|_{\mathbf{A}^{-1},\mathbf{A}} \stackrel{\text{def}}{=} \sup_{\mathbf{x}\neq 0} \frac{\|\mathbf{E}\mathbf{x}\|_{\mathbf{A}^{-1},\mathbf{A}}}{\|\mathbf{x}\|_{\mathbf{A}}} = \|\mathbf{A}^{-1/2}\mathbf{E}\mathbf{A}^{-1/2}\|_2. \tag{4.29}$$

nous pouvons démontrer le Théorème 4.2.2. Celui-ci fournit une borne supérieure sur la norme primale-duale des perturbations intervenant lors des produits matrice-vecteur au cours des itérations, pour obtenir à l'itération $k$ la borne suivante sur l'erreur sur le résidu

$$\|r(\mathbf{x}_k) - \mathbf{r}_k\|_{\mathbf{A}^{-1}} \leq \frac{\sqrt{\epsilon}}{2}\|\mathbf{b}\|_{\mathbf{A}^{-1}}.$$

**Theorem 4.2.2.** *Soient $\epsilon > 0$ et $\boldsymbol{\phi} \in (\mathbb{R}_+^*)^k$, dont les composantes vérifient*

$$\sum_{j=1}^{k} \frac{1}{\phi_j} \leq 1. \tag{4.30}$$

*Supposons de plus que*

$$\|\mathbf{E}_j\|_{\mathbf{A}^{-1},\mathbf{A}} \leq \omega_j \stackrel{\text{def}}{=} \frac{\sqrt{\epsilon}\,\|\mathbf{b}\|_{\mathbf{A}^{-1}}\|\mathbf{p}_j\|_{\mathbf{A}}}{2\,\phi_{j+1}\|\mathbf{r}_j\|_2^2 + \sqrt{\epsilon}\,\|\mathbf{b}\|_{\mathbf{A}^{-1}}\|\mathbf{p}_j\|_{\mathbf{A}}} \tag{4.31}$$

$\forall j \in \{0, \ldots, k-1\}$. *Alors*

$$\|r(\mathbf{x}_k) - \mathbf{r}_k\|_{\mathbf{A}^{-1}} \leq \frac{\sqrt{\epsilon}}{2}\|\mathbf{b}\|_{\mathbf{A}^{-1}}.$$

*Si de plus,*

$$\|\mathbf{r}_k\|_{\mathbf{A}^{-1}} \leq \frac{\sqrt{\epsilon}}{2}\|\mathbf{b}\|_{\mathbf{A}^{-1}}, \tag{4.32}$$

*alors (4.27) et (4.28) sont vérifiées.*

Quelques remarques peuvent être faites à ce théorème.

1. L'amplitude de la perturbation $\|\mathbf{E}_j\|_{\mathbf{A}^{-1},\mathbf{A}}$ dépend du ratio $\dfrac{\|\mathbf{r}_j\|_2^2}{\|\mathbf{p}_j\|_{\mathbf{A}}}$ : l'erreur sur les produits matrice-vecteur est donc amenée à croître au cours des itérations, au fur et à mesure que ce ratio décroît.
2. L'équation (4.31) renvoie au calcul (ou l'estimation) de la norme primale-duale de $\mathbf{E}_j$, ce qui peut être difficile à réaliser en pratique.
3. L'évaluation des conditions (4.31) et (4.32) est impossible en pratique, car les quantités $\|\mathbf{p}_j\|_{\mathbf{A}}$, $\|\mathbf{b}\|_{\mathbf{A}^{-1}}$, et $\|\mathbf{r}_k\|_{\mathbf{A}^{-1}}$ ne sont pas disponibles au cours des itérations de l'Algorithme 4 ;
4. les valeurs $\phi_j$ servent de budget "d'inexactitude", au sens où les précisions des calculs $\omega_j$ peuvent ne pas correspondre aux précisions disponibles (c'est le cas en arithmétique multi-précision) : ayant été plus précis que nécessaire à une itération, il est ainsi possible de reporter l'écart pour les itérations futures.

Nous avons donc proposé des heuristiques pour évaluer les quantités indisponibles (p.e. $\|\mathbf{p}_j\|_{\mathbf{A}}$, $\|\mathbf{b}\|_{\mathbf{A}^{-1}}$, $\|\mathbf{r}_k\|_{\mathbf{A}^{-1}}$, etc..), afin d'obtenir un algorithme, qui puisse être utilisable en pratique. Plus de détails et une description de celui-ci se trouvent dans [71].

La Figure 4.5 représente l'évolution de différentes quantités d'intérêt, telles que la décroissance de la quadratique au cours des itérations, la norme $\mathbf{A}^{-1}$ du résidu interne $\mathbf{r}_k$ et la borne d'erreur pour les produits matrice-vecteur ($\omega_j$). Le critère d'arrêt est activé selon (4.32), avec $\epsilon = \epsilon_M = 2^{-52}$ la précision machine en double (norme IEEE). Dans ce cas, le Lemme 4.2.1 garantit la décroissance de la quadratique (4.28). La matrice $\mathbf{A}$ correspond à la matrice nos1.mat présente dans la collection de matrices *Matrix Market*[2]. Cette matrice est petite ($n = 237$), mais très mal conditionnée (autour de $10^8$). Enfin, celle-ci est multipliée par une constante, pour lui imposer $\|\mathbf{A}\|_2 = 1$. Nous nous intéressons à deux configurations possibles pour le choix des niveaux de précision des produits matrice-vecteur.

2. https ://math.nist.gov/MatrixMarket/

— Niveau continu de précisions : il est possible de faire varier continûment le niveau de précision des calculs, ce qui rend la borne $\omega_j$ (Eq. 4.31) atteignable en pratique. Ce cas se veut représentatif d'un opérateur produit matrice-vecteur du type $\mathbf{A}^{-1}\mathbf{x}$, résolu itérativement jusqu'à la précision souhaitée.

— Niveau discret de précisions : on ne dispose que d'un nombre fini de précisions pour faire les calculs, et ceux-ci ne correspondent pas nécessairement à la borne $\omega_j$ (Eq. 4.31). Ceux-ci sont donc réalisés avec une précision effective $\hat{\omega}_j$, inférieure à la borne théorique. Ce cas est représentatif du calcul en arithmétique multi-précision (quart, demi, simple et double IEEE précisions par exemple).

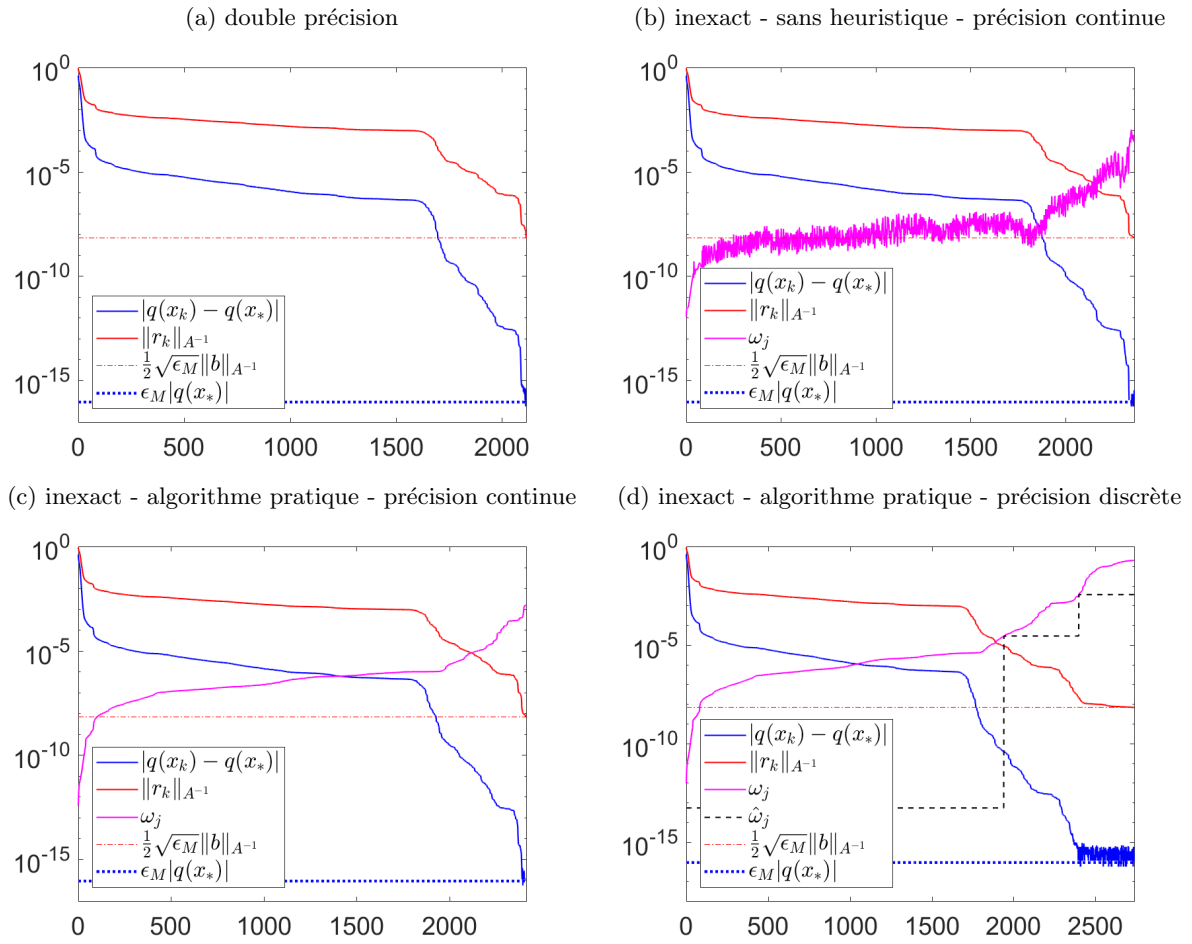



FIGURE 4.5 – Produits matrice-vecteur inexacts : algorithmes du gradient conjugué appliqué à la matrice nos1.mat. Figure tirée de Gratton *et al.* (2021) [71].

La Figure 4.5(a) représente l'algorithme de référence, à savoir le gradient conjugué pour lequel les produits matrice-vecteur sont réalisés en double précision. La Figure 4.5(b) renvoie à l'algorithme inexact théorique, à savoir celui pour lequel aucune heuristique n'est introduite. Celui-ci n'a pas vocation à être utilisé en pratique, mais vise à évaluer l'impact des heuristiques introduites sur la convergence de l'algorithme. L'inexactitude dans les produits matrice-vecteur est simulée par l'introduction d'une perturbation aléatoire satisfaisant la condition (4.31). Pour

cette figure, le scénario correspond à celui du niveau continu de précision. Enfin, les Figures 4.5(c) et 4.5(d) renvoie à l'algorithme pratique, pour les deux scénarios des niveaux continu et discret de précision. Nous notons dans un premier temps, que les variantes inexactes de l'algorithme du gradient conjugué, convergent bien vers la solution du problème d'optimisation, à la précision souhaité, tant pour la norme du résidu interne que pour la décroissance sur la quadratique : les courbes continues bleu et rouge atteignent bien les seuils associés (courbes pointillés de la même couleur). On note que ceci se fait au prix d'une légère augmentation du nombre d'itérations. Néanmoins, un nombre conséquent de celles-ci étant réalisé avec une précision dégradée, et donc un coût de calcul et énergétique réduit, le bilan global reste en faveur de ces approches. On note de surcroît que les bornes théoriques et approximées, sur les erreurs dans les produits matrice-vecteur, augmentent aux cours des itérations, comme le suggère la théorie. Plus on avance dans les itérations, et plus les l'erreur sur les produits matrice-vecteur peut être grande. Nous observons également que l'introduction des heuristiques, afin d'obtenir un algorithme pratique, ne modifie pas fondamentalement le comportement de l'algorithme (cf (c) et (d)). Enfin, on note que la convergence de l'algorithme est similaire pour le scénario du niveau discret de précision (d).

Ceci illustre ainsi la possibilité, qui nous est offerte, de réduire les coût d'opérations telles que les produits matrice-vecteur lors des itérations d'un algorithme de type gradient conjugué, sans pour autant compromettre la convergence de la minimisation de la quadratique associée, et cela grâce au contrôle de l'erreur au fil des itérations.

### 4.2.2 Une variante de GMRES avec produits scalaires inexacts

Outre les produits matrice-vecteur, nous nous sommes également intéressés à un autre processus coûteux en calcul, à savoir l'orthogonalisation d'une famille de vecteurs. Celle-ci peut-être introduite dans des algorithmes de type gradient conjugué, afin d'apporter une meilleure stabilité numérique, pour des matrices mal conditionnées. Il peut également être la base d'algorithmes, tels que ceux reposant sur le processus d'Arnoldi. Nous nous sommes donc intéressés à l'utilisation de produits scalaires inexacts dans le cas de l'algorithme itératif GMRES (systèmes linéaires non-symétriques) [120]. Nous avons proposé une variante de cet algorithme permettant une évolution dynamique de la précison des calculs de ces produits, tout en garantissant la même vitesse de convergence que l'algorithme GMRES originel. Ces travaux sont détaillés dans [72].

Soit le système linéaire

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

avec $\mathbf{A} \in \mathbb{R}^{n\times n}$ inversible et $\mathbf{b} \in \mathbb{R}^n$. Le processus d'Arnoldi est décrit dans l'Algorithme 5, et vise à construire une base orthogonale de l'espace de Krylov associé au couple $(\mathbf{A}, \mathbf{b})$.

Après $k$ étapes de l'Algorithme 5, et supposant travailler en arithmétique exacte, on obtient les matrices $\mathbf{V}_{k+1} = [\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}] \in \mathbb{R}^{n\times(k+1)}$ et $\mathbf{H}_k \in \mathbb{R}^{(k+1)\times k}$ Hessenberg supérieure, telles que

$$\mathbf{v}_1 = \frac{\mathbf{b}}{\beta}, \qquad \mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1}\mathbf{H}_k, \qquad \mathbf{V}_{k+1}^T\mathbf{V}_{k+1} = \mathbf{I}_{k+1}.$$

Les colonnes de $\mathbf{V}_k$ forment alors une base orthonormale du sous-espace de Krylov

$$\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \mathrm{Vect}\big\{\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \ldots, \mathbf{A}^{k-1}\mathbf{b}\big\}.$$

Dans le cas de l'algorithme GMRES, on cherche la solution $\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A}, \mathbf{b})$ qui minimise la norme Euclidienne du résidu sur ce sous-espace. Ceci conduit à trouver $\mathbf{x}_k = \mathbf{V}_k\mathbf{y}_k$, tel que $\mathbf{y}_k \in \mathbb{R}^k$

**Algorithm 5** Processus d'Arnoldi

1: $\beta = \sqrt{\mathbf{b}^T\mathbf{b}}$
2: $\mathbf{v}_1 = \mathbf{b}/\beta$
3: **for** $j = 1, 2, \ldots$ **do**
4: $\quad \mathbf{w}_j = \mathbf{A}\mathbf{v}_j$
5: $\quad$ **for** $i = 1, \ldots, j$ **do**
6: $\quad\quad \mathbf{h}_{ij} = \mathbf{v}_i^T\mathbf{w}_j$
7: $\quad\quad \mathbf{w}_j = \mathbf{w}_j - h_{ij}\mathbf{v}_i$
8: $\quad$ **end for**
9: $\quad h_{j+1,j} = \sqrt{\mathbf{w}_j^T\mathbf{w}_j}$
10: $\quad \mathbf{v}_{j+1} = \mathbf{w}_j/h_{j+1,j}$
11: **end for**

soit la solution du problème aux moindres carrés linéaires suivant

$$\min_{\mathbf{y}\in\mathbb{R}^{\mathbf{k}}} \|\mathbf{b} - \mathbf{A}\mathbf{V}_k\mathbf{y}\|_2 = \min_{\mathbf{y}\in\mathbb{R}^{\mathbf{k}}} \|\mathbf{V}_{k+1}(\beta\mathbf{e}_1 - \mathbf{H}_k\mathbf{y})\|_2 = \min_{\mathbf{y}\in\mathbb{R}^{\mathbf{k}}} \|\beta\mathbf{e}_1 - \mathbf{H}_k\mathbf{y}\|_2.$$

par orthonormalité des colonnes de $\mathbf{V}_{k+1}$. On obtient alors les relations suivantes entre $\mathbf{x}_k$ et les résidus $\mathbf{r}_k$ et $\mathbf{t}_k \stackrel{\text{def}}{=} \beta\mathbf{e}_1 - \mathbf{H}_k\mathbf{y}_k$ :

$$\begin{aligned} \mathbf{x}_k &= \mathbf{V}_k\mathbf{y}_k = \mathbf{V}_k(\mathbf{H}_k^T\mathbf{H}_k)^{-1}\mathbf{H}_k^T(\beta\mathbf{e}_1), \\ \mathbf{r}_k &= \mathbf{b} - \mathbf{A}\mathbf{x}_k = \mathbf{V}_{k+1}(\beta\mathbf{e}_1 - \mathbf{H}_k\mathbf{y}_k) = \mathbf{V}_{k+1}\mathbf{t}_k. \end{aligned} \tag{4.33}$$

Par la suite, supposons uniquement réaliser les produits scalaires de manière inexacte, ce qui se traduit par la modification des instructions 6 : et 9 : de l'Algorithme 5 :

$$6: \quad h_{ij} = \mathbf{v}_i^T\mathbf{w}_j + \eta_{ij}, \qquad i = 1, \ldots, j, \tag{4.34a}$$

$$9: \quad h_{j+1,j} = \sqrt{\mathbf{w}_j^T\mathbf{w}_j + \eta_{j+1,j}}, \tag{4.34b}$$

avec les erreurs $|\eta_{ij}|$ et $|\eta_{j+1,j}|$ majorées par des tolérances à définir. La relation $\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1}\mathbf{H}_k$ reste toujours valide dans ce cas, seule l'orthogonalité des vecteurs colonnes de $\mathbf{V}_{k+1}$ est perdue. On a ainsi les relations

$$[\mathbf{b}, \mathbf{A}\mathbf{V}_k] = \mathbf{V}_{k+1}[\beta\mathbf{e}_1, \mathbf{H}_k], \qquad \mathbf{V}_{k+1}^T\mathbf{V}_{k+1} = \mathbf{I}_{k+1} + \mathbf{F}_{k+1}. \tag{4.35}$$

Nos travaux ont porté sur le contrôle de l'erreur afin de garantir que pour chacune des itérations de l'algorithme GMRES avec produits scalaires inexacts, celui-ci produise des itérés "proches" de celui que l'on obtiendrait avec GMRES en arithmétique exacte. Pour cela, nous nous sommes intéressés à la majoration de la quantité $\|\mathbf{F}_{k+1}\|_2$ traduisant la perte d'orthonormalité des colonnes de $\mathbf{V}_{k+1}$. Nous avons ainsi démontré le Théorème 4.2.3, permettant le contrôle de $\|\mathbf{F}_{k+1}\|_2$ depuis celui des erreurs commises lors des produits scalaires.

**Theorem 4.2.3.** *Supposons que le processus d'Arnoldi s'effectue en utilisant des produits scalaires inexacts, tels que définis par* (4.34).

*Posons*

$$\eta_j \equiv \frac{\phi_j}{\sqrt{2}} \frac{\epsilon\|\mathbf{b}\|_2}{\|\mathbf{t}_{j-1}\|_2} \sigma_{\min}(\mathbf{H}_k), \qquad j = 1, \ldots, k, \tag{4.36}$$

*avec* $\epsilon \in (0,1)$ *et* $\phi_j > 0$ *vérifiant* $\sum_{j=1}^{k} \phi_j^2 \leq 1$.

*Si de plus, les erreurs sur les produits scalaires* (4.34) *vérifient*

$$|\eta_{ij}| \leq \eta_j, \qquad i = 1, \ldots, j, \tag{4.37a}$$

$$|\eta_{j+1,j}| \leq \eta_j h_{j+1,j}, \tag{4.37b}$$

*alors*

$$\|\mathbf{F}_{k+1}\|_2 \leq \frac{2(\sqrt{2}+k)\epsilon\|\mathbf{b}\|_2}{\|\mathbf{t}_k\|_2}. \tag{4.38}$$

Afin d'illustrer ce qu'implique ce théorème, supposons avoir réalisé plus de deux itérations, de sorte que $k > 2$ et $2(\sqrt{2}+k) < 3k$. Il vient alors

$$\|\mathbf{t}_k\|_2 > 6k\epsilon\|\mathbf{b}\|_2, \tag{4.39}$$

ce qui implique que $\|\mathbf{F}_{k+1}\|_2 < \frac{1}{2}$. En combinant (4.35) et la relation $\mathbf{r}_k = \mathbf{V}_{k+1}\mathbf{t}_k$, il vient

$$\kappa_2(\mathbf{V}_{k+1}) < \sqrt{\frac{1+1/2}{1-1/2}} = \sqrt{3} \quad \text{et} \quad \frac{1}{\sqrt{2}} < \frac{\|\mathbf{r}_k\|_2}{\|\mathbf{t}_k\|_2} < \frac{\sqrt{3}}{\sqrt{2}}.$$

Ainsi, la norme du résidu $\|\mathbf{t}_k\|_2 = \|\beta\mathbf{e}_1 - \mathbf{H}_k\mathbf{y}_k\|_2$, évaluée au cours de l'algorithme, fournit une estimation fiable de la norme du résidu $\|\mathbf{r}_k\|_2 = \|\mathbf{b} - \mathbf{A}\mathbf{x}_k\|_2$.

Finalement, le Corollaire 4.2.4 permet de conclure, que la convergence de l'algorithme GMRES n'est pas significativement affectée par l'utilisation de produits scalaires inexacts, dès lors que les erreurs associées vérifient (4.34).

**Corollary 4.2.4.** *Soit* $\mathbf{x}_k^{(e)}$ *le* $k$*-ème itéré de l'algorithme GMRES, exécuté en arithmétique exacte, et* $\mathbf{r}_k^{(e)}$ *le résidu associé. Supposons que le processus d'Arnoldi soit exécuté avec des produits scalaires inexacts vérifiant* (4.34)*. On note* $\mathbf{x}_k$ *et* $\mathbf{r}_k$ *les itéré et résidu de l'algorithme GMRES associé. Si les erreurs* $\eta_{ij}$ *dans les produits scalaires sont majorées selon le Théorème 4.2.3, et ce pour toutes les itérations* $j = 1, \ldots, k$*, alors pour toute itération* $k$ *vérifiant* (4.39)*, on a*

$$1 \leq \frac{\|\mathbf{r}_k\|_2}{\|\mathbf{r}_k^{(e)}\|_2} \leq \sqrt{3}.$$

Une remarque importante est que le Corollaire 4.2.4 ne concerne pas uniquement la précision du dernier itéré $\mathbf{x}_k$ obtenu par GMRES, mais l'ensemble de celles obtenues aux itérations $k$ jusqu'à convergence de l'algorithme pour une précision relative de $6k\epsilon$ sur le résidu, avec $\epsilon$ la tolérance utilisateur.

Quelques remarques pratiques concernent ces résultats.

1. La borne $\eta_j$ (4.36) étant inversement proportionnelle à $\|\mathbf{t}_{j-1}\|_2$, les produits scalaires pourront être effectués de manière de plus en plus inexacte au cours des itérations. On retrouve le résultat similaire observé pour les produits matrice-vecteur dans l'algorithme GMRES [130, 137].
2. Le choix des variables $\phi_j$, associées de nouveau à un budget "d'inexactitude", peut être soit obtenu depuis le nombre d'itération maximum $K_{\max}$ selon $\phi_j = K_{\max}^{-1/2}$, soit de manière plus sophistiquée [71].

3. Le calcul de la borne $\eta_j$ au début de l'itération $j$ requiert la connaissance de $\sigma_{\min}(H_k)$, qui n'est pas disponible avant l'itération $k$. Ce problème est similaire à celui rencontré dans la théorie des produits matrice-vecteur inexacts dans GMRES [130, 137]. En arithmétique exacte, nous avons $\sigma_{\min}(\mathbf{A}) \leq \sigma_{\min}(\mathbf{H}_k)$. Il est donc possible d'approximer la borne $\eta_j$ par la borne "conservative" (4.40).

$$\text{Borne "conservative" :} \qquad \eta_j = \epsilon\, \sigma_{\min}(\mathbf{A}) \frac{\phi_j \|\mathbf{b}\|_2}{\sqrt{2}\|\mathbf{t}_{j-1}\|_2}. \tag{4.40}$$

Dans le cas où $\sigma_{\min}(\mathbf{A})$ ne serait pas disponible, une stratégie "agressive" consiste à faire l'approximation $\sigma_{\min}(\mathbf{A}) \approx 1$. On obtient alors la borne "agressive" (4.41).

$$\text{Borne "agressive" :} \qquad \eta_j = \epsilon \frac{\phi_j \|\mathbf{b}\|_2}{\sqrt{2}\|\mathbf{t}_{j-1}\|_2}. \tag{4.41}$$

Un tel choix peut être vu comme inclure $1/\sigma_{\min}(\mathbf{H}_k)$ dans $\epsilon$ et le Corollaire 4.2.4 reste valide, tant que $\|\mathbf{t}_k\|_2 > 6k\epsilon\|\mathbf{b}\|_2/\sigma_{\min}(\mathbf{H}_k)$. On en déduit que la convergence de l'algorithme ne devrait pas être trop affectée lors des premières itérations, mais qu'une stagnation de la norme relative du résidu devrait apparaître au voisinage de $\epsilon/\sigma_{\min}(H_k)$.

4. Le choix de la précision à utiliser pour les produits scalaires implique l'estimation des perturbations $(\eta_{ij})$ définies en (4.34). Dans le contexte de l'arithmétique en virgule flottante, une analyse d'erreur [81, Section 3.1] des produits scalaires réalisés avec une précision machine $u$ (avec $nu < 1$) conduit aux inégalités

$$\begin{aligned} |\eta_{ij}| &\leq \frac{nu}{1-nu}|\mathbf{v}_i|^T|\mathbf{w}_j|, \approx nu|\mathbf{v}_i|^T|\mathbf{w}_j| \qquad i = 1, \ldots, j, \\ |\eta_{j+1,j}| &\leq \frac{nu}{1-nu}\mathbf{w}_j^T\mathbf{w}_j \approx nu\|\mathbf{w}\|_2 h_{j+1,j}. \end{aligned} \tag{4.42}$$

Le choix de la précision des calculs va se faire en comparant les perturbations (4.42) depuis une précision machine $u$ fixée (p.e. demi, simple ou double) aux bornes (4.40) et (4.41). Ceci doit naturellement se faire sans calculer explicitement les perturbations (4.42), qui font intervenir des produits scalaires. Ceci se fait au prix d'heuristiques décrites dans [72].

Dans la suite sont présentés les résultats d'expériences numériques, réalisées en arithmétique flottante avec précision multiple (demi, simple et double précisions). Pour cela, nous avons utilisé le langage Julia[3], permettant de choisir le type des variables (Float16, Float32 et Float64) au cours des calculs. L'algorithme GMRES avec produits scalaires inexacts est appliqué à la matrice 494_bus[4] avec les bornes conservative et agressive. Cette matrice est de taille $n = 497$, présente un conditionnement de l'ordre de $10^6$, et sa plus petite valeur singulière vaut $\sigma_{\min}(\mathbf{A}) \approx 10^{-2}$. La borne agressive surestime donc celle-ci d'un facteur 100. La tolérance utilisateur, sur la décroissance relative du résidu $\|\mathbf{t}_k\|_2$, est choisie égale à $\epsilon = 10^{-6}\|\mathbf{A}\|_2$ afin d'illustrer le potentiel de l'algorithme pour une précision modérée. Les résultats pour une précision élevée ($\epsilon = 10^{-12}\|\mathbf{A}\|_2$) sont disponibles dans [72].

Nous observons sur la Figure 4.6, que la borne "conservative" conduit rapidement à un changement de précision pour calculer les produits scalaires : la simple précision commence à être utilisée au bout de 40 itérations, et son emploi devient prédominant jusqu'à convergence.

3. https ://julialang.org/
4. https ://math.nist.gov/MatrixMarket/

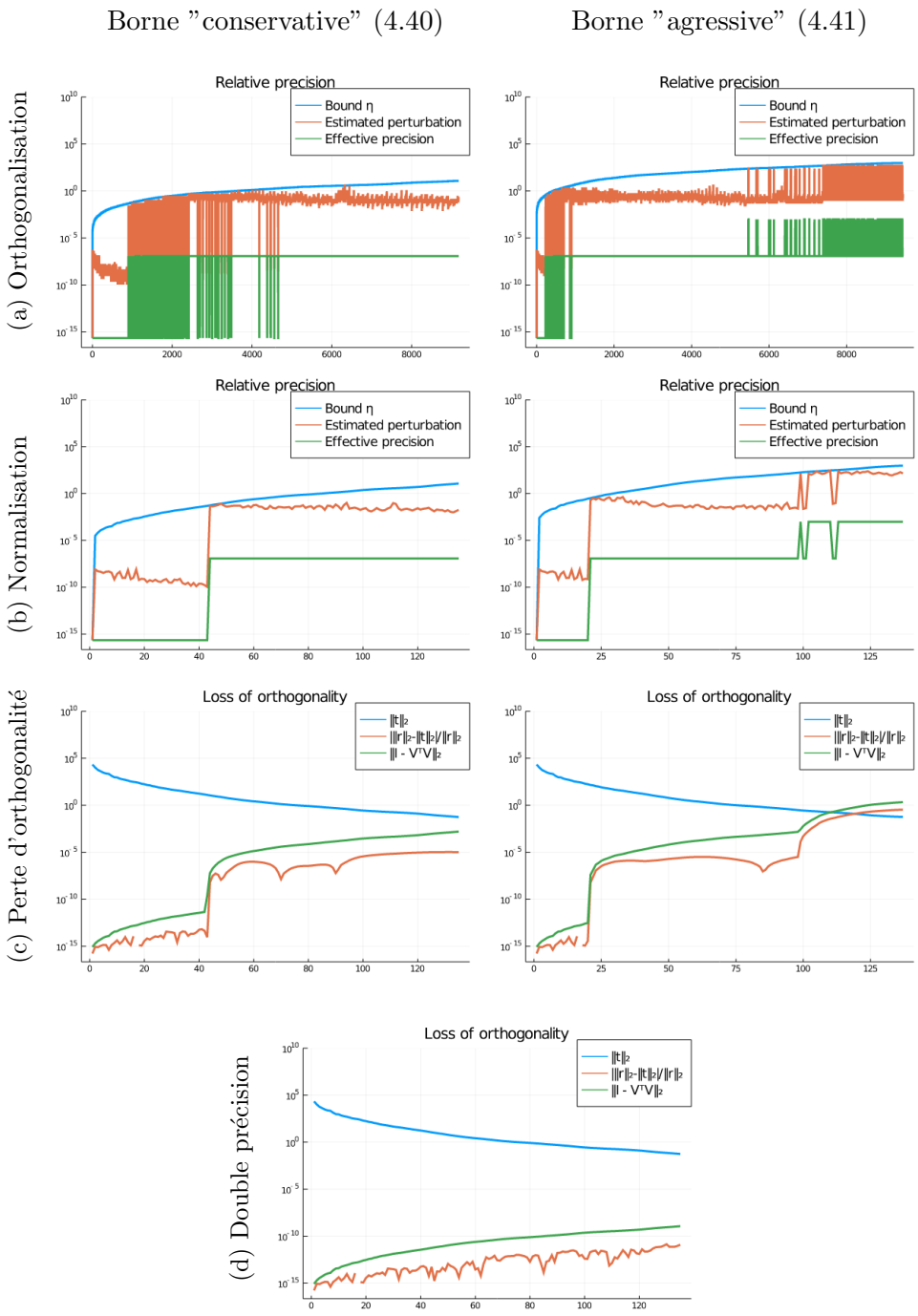



FIGURE 4.6 – GMRES avec produit scalaire en multi-précision : expériences avec la matrice 494_bus. (a) Produits scalaires dans l'étape d'orthogonalisation : la courbe bleu représente les bornes (4.40) et (4.41), la rouge l'estimation de la perturbation associée au produit scalaire (4.42), et la verte la précision choisie pour les calculs. L'axe des abscisses représente le nombre de produits scalaires effectués. (b) Produit scalaire dans l'étape de normalisation : idem (a). L'axe des abscisses correspond au nombre d'itérations $k$. (c) Perte d'orthogonalité : la courbe verte représente $\|\mathbf{I}_{k+1} - \mathbf{V}_{k+1}^T \mathbf{V}_{k+1}\|_2$, la rouge $\frac{|\|\mathbf{r}_k\|_2 - \|\mathbf{t}_k\|_2|}{\|\mathbf{r}_k\|_2}$, et la bleu $\|\mathbf{t}_k\|_2$. (d) Perte d'orthogonalité pour GMRES en double precision : idem (c). Figure tirée de Gratton *et al.* (2022) [72].

Ceci se traduit par une perte d'orthogonalité marquée lors du premier changement de précision. On observe également que la quantité $\|\mathbf{t}_k\|_2$ est un bonne approximation de la norme du résidu $\|\mathbf{r}_k\|_2$, comme la théorie le suggère. Enfin, à l'exception des itérations associées à un changement de précision, les évolutions de la perte d'orthogonalité et de l'erreur relative sur les normes de résidus suivent une tendance similaire à celles obtenues depuis GMRES en double précision.

L'utilisation de la borne agressive a pour conséquence d'activer plus tôt l'emploi de la simple précision (autour de l'itération 20), et même de la demi précision sur les 20 dernières itérations. On note que la précision des produits scalaires oscille entre la simple et la demi, selon l'amplitude des vecteurs $\mathbf{w}_j$. Ceci conduit à une perte d'orthogonalité plus sévère qu'avec la borne "conservative", sans pour autant compromettre la convergence de l'algorithme sur cet exemple.

De nouveau, ceci illustre la possibilité, qui nous est offerte, de réduire les coût de opérations telles que les produits scalaires lors des itérations de l'algorithme GMRES, sans pour autant compromettre sa convergence. Ceci est d'autant plus intéressant que la résolution du système non symétrique est associé à un sous-problème d'un algorithme itératif plus complexe (cf la formulation point-selle de l'algorithme 4D-Var contraintes faibles en section 4.1.2, et l'Algorithme 3, qui y est proposé).

## 4.3 Des normes non-standards

Les formulations classiques du problème d'assimilation de données reposent sur la résolution de problèmes de moindres carrés en norme Euclidienne, potentiellement non-linéaires, pondérés par des matrices de covariance d'erreur (erreur de prévision/background, erreur d'observation, erreur modèle). Selon les variantes considérées, la fonctionnelle à minimiser s'écrit comme la somme d'un terme mesurant l'écart entre la trajectoire/solution du modèle et les observations et d'un terme mesurant l'écart au background, à savoir la solution estimée a priori. Ce dernier terme peut être interprété comme un terme de régularisation de Tikhonov généralisée. Il peut conduire à des biais d'assimilation dus au lissage des solutions qu'il induit, ainsi qu'aux difficultés pour maintenir et/ou garantir des propriétés de parcimonie des solutions estimées [77]. Ces problèmes peuvent être de surcroît amplifiés selon la nature des opérateurs définissant les matrices de covariance d'erreur (opérateur de diffusion [144]). Une stratégie, issue de la théorie du traitement du signal, consiste à modifier la fonctionnelle à minimiser afin de prendre en compte l'information a priori sur la parcimonie de la solution. Ceci se traduit typiquement par l'ajout d'un terme de pénalisation portant sur la norme $\|\|_p$ de la solution exprimée dans la base pour laquelle les propriétés de parcimonie s'expriment. En effet, une manière naturelle d'introduire cette information a priori se fait au travers de l'emploi de la distribution normale généralisée [19, 104]

$$p_p(\mathbf{x}) \sim \exp\left(-\|\mathbf{\Phi}\mathbf{x}\|_p^p\right) \tag{4.43}$$

avec $\|\mathbf{z}\|_p^p = \sum_{j=1}^{n} |z_j|^p$, $\Phi$ la base pour laquelle les propriétés de parcimonie de $\mathbf{x}$ s'expriment et $p \in ]0, 2]$. Le choix de $p$ est important : les grandes valeurs tendent à prévenir l'apparition de discontinuité, tandis que les petites les encouragent [36].

### 4.3.1 Régularisation en norme $\|\|_p$ en assimilation variationnelle de données

Dans le cadre de la thèse d'Antoine Bernigaud, que j'ai co-encadrée avec S. Gratton, nous nous sommes intéressés à la formulation et la résolution numérique du problème d'assimilation

variationnelle de données, quand un terme de régularisation en norme $p$, avec $p > 1$, est introduit dans la fonctionnelle à minimiser. Ces travaux font suite à ceux initiés durant le post-doctorat de F. Lenti, et le stage d'O. Sohab. Une partie de ces travaux est détaillée dans [8].

La formulation de l'algorithme 4DVar avec régularisation en norme $\|\|_p$ peut s'écrire, sous forme concaténée, comme la minimisation de la fonctionnelle

$$J(\mathbf{x}) = \frac{1}{2} \parallel \mathbf{y} - \hat{\mathcal{H}}(\mathbf{x}) \parallel^2_{\mathbf{R}^{-1}} + \frac{1}{2} \parallel \mathbf{x} - \mathbf{x}^0_b \parallel^2_{\mathbf{B}^{-1}} + \frac{\lambda}{p} \parallel \mathbf{\Phi x} \parallel^p_p \tag{4.44}$$

avec $\hat{\mathcal{H}}(\mathbf{x}) = [\mathcal{H}_1(\mathcal{M}_{1,0}(\mathbf{x})), ..., \mathcal{H}_k(\mathcal{M}_{k,0}(\mathbf{x}))]^T$ et $\lambda > 0$.

Suivant [57], en supposant l'opérateur $\hat{\mathcal{H}}$ linéaire et en posant

$$\mathbf{A} = \begin{pmatrix} \mathbf{R}^{-\frac{1}{2}}\hat{\mathbf{H}} \\ \mathbf{B}^{-\frac{1}{2}} \end{pmatrix} \mathbf{\Phi}^{-1}; \quad \mathbf{b} = \begin{pmatrix} \mathbf{R}^{-\frac{1}{2}}\mathbf{y} \\ \mathbf{B}^{-\frac{1}{2}}\mathbf{x}^0_b \end{pmatrix}, \tag{4.45}$$

avec $\boldsymbol{\xi} = \mathbf{\Phi x}$, la fonctionnelle (4.44) à minimiser se réécrit

$$\Omega_p(\boldsymbol{\xi}, \lambda) = \frac{1}{2} \parallel \mathbf{A}\boldsymbol{\xi} - \mathbf{b} \parallel^2_2 + \frac{\lambda}{p} \parallel \boldsymbol{\xi} \parallel^p_p, \tag{4.46}$$

avec $\Omega_p$ strictement convexe et dérivable pour $1 < p < \infty$.

Devant le peu d'efficacité, de la méthode de la descente de gradient pour minimiser $\Omega_p$, observée lors d'expériences réaliser avec un modèle d'advection linéaire 1D, nous nous sommes intéressés aux méthodes de descente dans les espaces de Banach, dont les itérations sont réalisées dans le dual de l'espace de contrôle, et plus particulièrement aux itérations suivantes

$$\begin{cases} \boldsymbol{\xi}^*_k = j_p(\boldsymbol{\xi}_{k-1}) - \mu_k \Omega'_p(\boldsymbol{\xi}_{k-1}, \lambda). \\ \boldsymbol{\xi}_k = j_q(\boldsymbol{\xi}^*_k). \end{cases} \tag{4.47}$$

avec $p$ et $q$ liés par $p^{-1} + q^{-1} = 1$, $j_p : \mathbb{R}^n \to \mathbb{R}^n$ définie par

$$\forall i = 1 : n, \quad [j_p(\boldsymbol{\xi})]_i = |\xi_i|^{p-1}\text{sign}(\xi_i), \tag{4.48}$$

et

$$\text{sign}(\xi) = \begin{cases} 1 & \xi > 0 \\ 0 & \xi = 0 \\ -1 & \xi < 0 \end{cases} \tag{4.49}$$

Différentes stratégies sont possibles pour le choix de la longueur de pas $\mu_k$ conduisant à la convergence de l'algorithme vers le minimum de $\Omega_p$ [18]. Une difficulté importante tient au choix de la valeur du paramètre $\lambda$ à spécifier. Dans le cas, où la variable $\boldsymbol{\xi}$ suit une distribution normale généralisée, $\lambda$ peut être directement calculé depuis l'hyper-paramètre d'échelle de la distribution, pour peu que celui-ci soit connu. Dans le cas contraire, la recherche d'un valeur adéquate se fait au moyen d'heuristiques, telles que le principe de Morozov [2] ou la méthode "*L-curve*" [78, 143].

Dans un premier temps, nous nous sommes intéressés à la faisabilité d'une telle approche, ainsi que la sensibilité des résultats de l'algorithme 4DVar avec régularisation en norme $\|\|_p$ au choix de $p$. Les expériences numériques ont été réalisées avec un modèle d'advection linéaire 1D, pour lequel le schéma de discrétisation permet un contrôle de la diffusion numérique. Ceci est important afin d'éviter une perte de parcimonie, liée à une erreur modèle "diffusive". Un

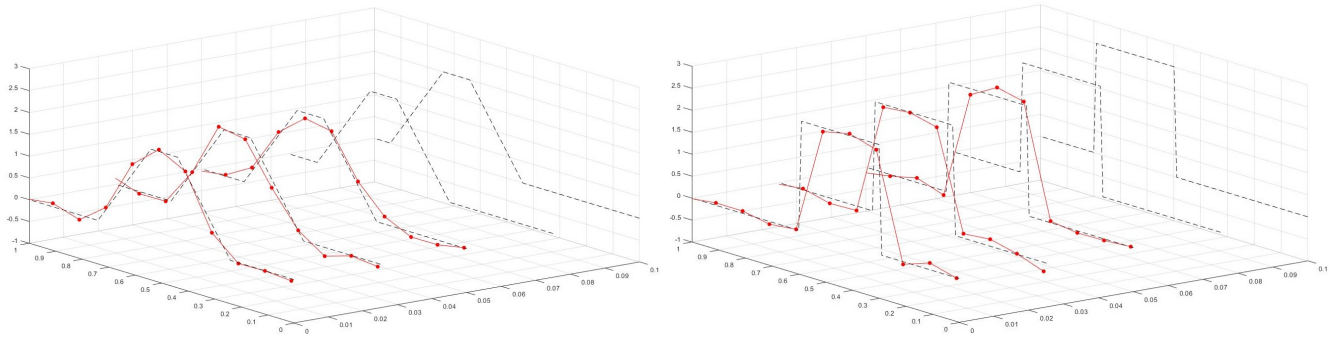


FIGURE 4.7 – Solution exacte (courbe pointillée noir) et observations assimilées (points rouge) à différents instants, pour une expérience avec un modèle d'advection linéaire 1D. A gauche : cas "presque" parcimonieux ; à droite : cas parcimonieux. Figure tirée de Bernigaud *et al.* (2021) [8]

exemple de configurations expérimentales est disponible à la figure (4.7). Nous envisageons deux scénarios : un cas parcimonieux (la référence est une fonction porte) pour lequel la valeur de $p = 1$ a déjà fourni de bon résultats [57], et un cas "presque" parcimonieux, pour lequel la solution de référence ne présente plus de discontinuité.

La figure (4.8) représente les nuages de points "(RMSE, RMAE)" sur l'état initial, à savoir la variable optimisée, pour les deux configurations parcimonieuse et "presque" parcimonieuse. Ces diagnostics sont définis par

$$\begin{aligned} RMSE(\mathbf{u}_t^a, \mathbf{u}_t^{tr}) &= \frac{\| \mathbf{u}_t^a - \mathbf{u}_t^{tr} \|_2^2}{\| \mathbf{u}_t^{tr} \|_2^2}, \\ RMAE(\mathbf{u}_t^a, \mathbf{u}_t^{tr}) &= \frac{\| \mathbf{u}_t^a - \mathbf{u}_t^{tr} \|_1}{\| \mathbf{u}_t^{tr} \|_1}. \end{aligned} \tag{4.50}$$

avec $\mathbf{u}_t^{tr}$ et $\mathbf{u}_t^a$ respectivement l'état de référence et la solution optimisée au temps $t$. Les performances de l'algorithme 4DVar "classique" (points gris), sont comparées à celles du 4DVar avec régularisation en norme $\|\|_p$, pour des valeurs de $p$ allant de 1 (points rouge) à 2 (points noir). On note tout d'abord que l'algorithme 4DVar classique obtient les pires résultats, suivi par l'al-

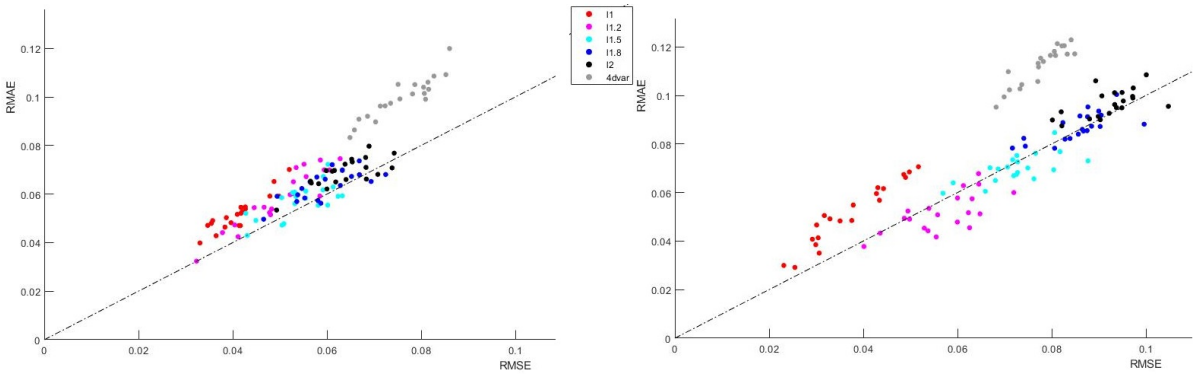



FIGURE 4.8 – Valeurs des RMSE et RMAE pour 20 expériences réalisées avec un modèle d'advection linéaire 1D. A gauche : cas "presque" parcimonieux ; à droite : cas parcimonieux. Figure tirée de Bernigaud *et al.* (2021) [8]

gorithme 4DVar avec régularisation en norme $\|\|_2$. Comme déjà observé en traitement du signal,

la norme Euclidienne n'est pas adaptée pour ce type de solutions. Les meilleurs résultats sont obtenus pour des valeurs de $p$, soit égale à 1 ; soit proche de 1 (1.2, points magenta). Suivant les valeurs spécifiées pour les variances des erreurs d'ébauche et d'observation (elles sont constantes pour les 20 expériences de la figure 4.8), les valeurs 1 et 1.2 conduisent aux meilleurs résultats en terme d'erreur. On note que la norme 1 tend à produire de meilleur résultats quand l'erreur d'ébauche est plus faible ou équivalente à l'erreur d'observation, et que la norme $\|\|_{1.2}$ conduit à de meilleurs résultats quand l'erreur d'ébauche domine l'erreur d'observation.

Dans le cas plus réaliste où le modèle est imparfait, et introduit notamment de la diffusion numérique, les algorithmes 4DVar avec régularisation en norme $\|\|_p$ sont également à même de fournir une condition initiale présentant la plus faible erreur, notamment pour des valeurs de $p$ légèrement supérieures à 1, au contraire de l'algorithme 4DVar classique, et de celui régularisé en norme $\|\|_1$ (cf Table 4.1) En effet, ceux-ci tendent à fournir des solutions pour lesquelles les

| $\mathbf{u}_t^a$ | $t=0$ | | $t=0.025$ | | $t=0.05$ | | $t=0.075$ | |
|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| $\mathbf{u}_t^{tr}$ | 0 | 0 | 0.1526 | 0.0820 | 0.1841 | 0.1175 | 0.2047 | 0.1445 |
| Ebauche | 0.2943 | 0.4313 | 0.1947 | 0.2393 | 0.2112 | 0.2430 | 0.2261 | 0.2539 |
| $\mathbf{u}_{4DVar}$ | 0.2746 | 0.3981 | 0.1738 | 0.1991 | **0.1930** | 0.2035 | **0.2099** | 0.2162 |
| $\mathbf{u}_{4DVar,1}$ | 0.2431 | 0.3486 | **0.1734** | 0.1973 | 0.1931 | **0.2032** | 0.2101 | **0.2161** |
| $\mathbf{u}_{4DVar,1.2}$ | **0.1195** | **0.1331** | 0.1876 | **0.1808** | 0.2119 | 0.2066 | 0.2290 | 0.2272 |
| $\mathbf{u}_{4DVar,1.5}$ | 0.1454 | 0.1617 | 0.1946 | 0.1953 | 0.2154 | 0.2161 | 0.2305 | 0.2334 |
| $\mathbf{u}_{4DVar,2}$ | 0.1754 | 0.2003 | 0.2037 | 0.2120 | 0.2202 | 0.2267 | 0.2334 | 0.2407 |

Table 4.1 – Modèle imparfait : RMSE et RMAE dans le cas parcimonieux. En gras : le meilleur résultat pour chaque instant. Table tirée de Bernigaud *et al.* (2021) [8].

erreurs sont plus faibles en fin de fenêtre d'assimilation, là où celles-ci sont plus sensibles à l'effet de l'erreur modèle, au détriment de la qualité de la condition initiale. On observe le contraire pour les valeurs de $p > 1$, confirmant l'intérêt d'une telle régularisation sur la solution au temps initial. Plus de détails sont disponibles dans [8].

Les travaux d'Antoine ont ensuite porté sur le développement d'algorithmes d'optimisation plus efficaces en coût et temps de calcul, pour la résolution de problèmes aux moindres carrés non-linéaires régularisés en norme $\|\|_p$, tels que (4.44). A l'instar de l'algorithme du gradient conjugué dans le cas Hilbertien qui ne suit plus la direction de plus grande pente après la première itération, l'objectif est de construire de nouvelles directions, de nouveau dans l'espace dual, en vu d'accélérer la convergence de l'algorithme. Outres des résultats préliminaires prometteurs sur un modèle de Saint-Venant 2D, les travaux ont porté sur la preuve de convergence globale de l'algorithme vers une solution annulant la dérivée première de la fonctionnelle à minimiser, ainsi qu'à la définition d'heuristiques permettant le calcul des directions dans l'espace dual. Plus de détails sur cet algorithme et l'évaluation de ses performances sont disponibles dans [9].

### 4.3.2 Vers l'introduction de normes non différentiables

L'introduction de normes non-différentiables dans la définition de la fonction coût nécessite de repenser les algorithmes utilisés pour la minimisation de celle-ci. Pour la norme $\|\|_1$, une stratégie consiste en la réécriture du problème comme un problème d'optimisation avec contraintes (cf [54] pour le cas particulier de l'algorithme 4DVar avec régularisation en norme $\|\|_1$). Dans le cadre

d'une collaboration avec S. Gratton et Ph.L. Toint, nous nous sommes intéressés au problème plus général de la minimisation de fonctions composites

$$\min_{\mathbf{x}\in\mathbb{R}^n} \psi(\mathbf{x}) \stackrel{\text{def}}{=} f(\mathbf{x}) + h(c(\mathbf{x})), \tag{4.51}$$

avec $f : \mathbb{R}^n \to \mathbb{R}$ à gradient Lipschitzien, $c : \mathbb{R}^n \to \mathbb{R}^m$ à Jacobienne Lipschitzienne, et $h : \mathbb{R}^m \to \mathbb{R}$ convexe Lipschitzienne, éventuellement non-dérivable. Pour cette classe de problème, nous avons proposé un algorithme de régularisation avec précision dynamique, celle-ci portant sur les évaluations de $f$, $c$ et leurs dérivées premières respectives (nous supposons que le coût de l'évaluation de $h$ est marginal devant ceux de $f$ et $c$, ce qui est le cas des normes $\|\|_1$ et $\|\|_\infty$). Nous avons démontré sa convergence vers un minimum au premier ordre de $\psi$, et fourni un majorant, dans le pire cas, de la complexité requise pour l'atteindre à la précision $\epsilon$ prêt. Ces travaux sont détaillés dans [73].

Supposons donc être en mesure de pouvoir évaluer $f$, $c$ et leurs dérivées premières de manière inexacte, et selon une précision définie par l'utilisateur. Pour la fonction $f$, ceci peut s'écrire

$$\overline{f}(\mathbf{x}) \stackrel{\text{def}}{=} \overline{f}(\mathbf{x}, \varepsilon_f) \text{ avec } |\overline{f}(\mathbf{x}, \varepsilon_f) - f(\mathbf{x})| \leq \varepsilon_f, \tag{4.52}$$

avec $\varepsilon_f$ la précision souhaitée et $\overline{f}(x, \varepsilon_f)$ l'évaluation inexacte de $f$ en $\mathbf{x}$. A l'itération $k$, les évaluations inexactes de $f$ en $\mathbf{x}_k$, de son gradient $\mathbf{g}_k$, de $c$ et de sa Jacobienne $\mathbf{J}_k$ sont notées

$$\overline{f}_k = \overline{f}(\mathbf{x}_k, \varepsilon_f) \text{ and } |\overline{f}_k - f_k| \leq \varepsilon_f, \tag{4.53}$$

$$\overline{\mathbf{g}}_k = \overline{g}(\mathbf{x}_k, \varepsilon_g) \text{ and } \|\overline{\mathbf{g}}_k - \mathbf{g}_k\| \leq \varepsilon_g, \tag{4.54}$$

$$\overline{\mathbf{c}}_k = \overline{c}(\mathbf{x}_k, \varepsilon_c) \text{ and } \|\overline{\mathbf{c}}_k - \mathbf{c}_k\| \leq \varepsilon_c, \tag{4.55}$$

$$\overline{\mathbf{J}}_k = \overline{J}(\mathbf{x}_k, \varepsilon_J) \text{ and } \|\overline{\mathbf{J}}_k - \mathbf{J}_k\| \leq \varepsilon_J, \tag{4.56}$$

De même, on définit la fonction inexacte

$$\overline{\psi}(\mathbf{x}) = \overline{f}(\mathbf{x}) + h(\overline{c}(\mathbf{x})),$$

ainsi que son "linéarisé" et le modèle associé par

$$\overline{\ell}_k(\mathbf{s}) \stackrel{\text{def}}{=} \overline{f}_k + \overline{\mathbf{g}}_k^T\mathbf{s} + h(\overline{\mathbf{c}}_k + \overline{\mathbf{J}}_k\mathbf{s}) \text{ et } \overline{m}_k(\mathbf{s}) \stackrel{\text{def}}{=} \overline{\ell}_k(\mathbf{s}) + \frac{\sigma_k}{2}\|\mathbf{s}\|^2.$$

Dans la suite, nous nous intéresserons à leur variations définies par

$$\overline{\Delta\ell}_k(\mathbf{s}) \stackrel{\text{def}}{=} -\overline{\mathbf{g}}_k^T\mathbf{s} + h(\overline{\mathbf{c}}_k) - h(\overline{\mathbf{c}}_k + \overline{\mathbf{J}}_k\mathbf{s}) \tag{4.57}$$

et

$$\overline{\Delta m}_k(\mathbf{s}) \stackrel{\text{def}}{=} -\overline{\mathbf{g}}_k^T\mathbf{s} + h(\overline{\mathbf{c}}_k) - h(\overline{\mathbf{c}}_k + \overline{\mathbf{J}}_k\mathbf{s}) - \frac{\sigma_k}{2}\|\mathbf{s}\|^2. \tag{4.58}$$

Enfin, nous utilisons la mesure d'optimalité $\phi_k$ définie par

$$\phi_k \stackrel{\text{def}}{=} \ell_k(\mathbf{0}) - \min_{\|\mathbf{d}\|\leq 1} \ell_k(\mathbf{d}) = \max_{\|\mathbf{d}\|\leq 1} \Delta\ell_k(\mathbf{d}). \tag{4.59}$$

Nous avons en effet que si $\mathbf{x}_k$ est un minimum local de $\psi$, définie par (4.51), alors $\phi_k = 0$ [145]. Cette quantité sera également approximée par

$$\overline{\phi}_k \stackrel{\text{def}}{=} \overline{\ell}_k(\mathbf{0}) - \min_{\|\mathbf{d}\|\leq 1} \overline{\ell}_k(\mathbf{d}) = \max_{\|\mathbf{d}\|\leq 1} \overline{\Delta\ell}_k(\mathbf{d}).$$

**Algorithm 4.3.1: Algorithme ARLDA**

**Etape 1 : évaluation de la mesure d'optimalité et test d'arrêt.**

— Calculer les quantités $\overline{f}_k$, $\overline{\mathbf{g}}_k$, $\overline{\mathbf{c}}_k$ et $\overline{\mathbf{J}}_k$ satisfaisant (4.53)–(4.56), si celles-ci ne sont pas déjà disponibles.

— Appliquer l'Algorithme 4.3.2, pour évaluation du critère d'arrêt, avec $\mathbf{x}_k$ et $\overline{\psi}(\mathbf{x}_k) = \overline{f}_k + h(\overline{\mathbf{c}}_k)$, ou obtenir $\overline{\phi}_k > \epsilon/(1+\omega_k)$ si l'arrêt n'a pas été déclenché.

**Etape 2 : calcul du pas.**

Appliquer l'Algorithme 4.3.3 pour minimiser approximativement $\overline{m}_k(\mathbf{s})$ et obtenir un pas $\mathbf{s}_k$ tel que

$$\overline{\Delta\ell}_k(\mathbf{s}_k) \geq \frac{1}{4} \min\left\{1, \frac{\overline{\phi}_k}{\sigma_k}\right\} \overline{\phi}_k. \tag{4.60}$$

**Etape 3 : mise-à-jour de l'itéré et de la tolérance sur $f$.**

— Réduire (éventuellement) $\varepsilon_f$ pour garantir

$$\varepsilon_f \leq \omega_k \overline{\Delta\ell}_k(\mathbf{s}_k). \tag{4.61}$$

Si $\varepsilon_f$ a été modifié, recalculer $\overline{f}_k(\mathbf{x}_k, \varepsilon_f)$ pour garantir (4.53).

— Calculer $\overline{f}_k(\mathbf{x}_k + \mathbf{s}_k, \varepsilon_f)$.

— Calculer

$$\rho_k = \frac{\overline{\psi}(x_k) - \overline{\psi}(x_k + s_k)}{\overline{\Delta\ell}_k(s_k)}. \tag{4.62}$$

Si $\rho_k \geq \eta_1$, mettre à jour $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$ ; sinon $\mathbf{x}_{k+1} = \mathbf{x}_k$, avec $0 < \eta_1 < 1$.

**Etape 4 : mise-à-jour du paramètre de régularisation.**

Choisir

$$\sigma_{k+1} \in \begin{cases} [\max(\sigma_{\min}, \gamma_1\sigma_k), \sigma_k] & \text{si } \rho_k \geq \eta_2, \\ [\sigma_k, \gamma_2\sigma_k] & \text{si } \rho_k \in [\eta_1, \eta_2), \\ [\gamma_2\sigma_k, \gamma_3\sigma_k] & \text{si } \rho_k < \eta_1, \end{cases} \tag{4.63}$$

avec $0 < \gamma_1 < 1 < \gamma_2 < \gamma_3$ et $0 < \eta_1 \leq \eta_2 < 1$.

**Etape 5 : mise-à-jour des tolérances.**

Définir

$$\omega_{k+1} = \min\left[\kappa_\omega, \frac{1}{\sigma_{k+1}}\right] \tag{4.64}$$

avec $\kappa_\omega \in (0, \frac{1}{3}\alpha\eta_1]$ et $\alpha \in (0,1)$.

Redéfinir $\varepsilon_f \leq \varepsilon_f^{\max}$, $\varepsilon_g \leq \varepsilon_g^{\max}$, $\varepsilon_c \leq \varepsilon_c^{\max}$ et $\varepsilon_J \leq \varepsilon_J^{\max}$ telles que $\varepsilon_f + L_h\varepsilon_c \leq \omega_{k+1}$, où $L_h$ est la constante de Lipschitz de $h$.

Avec ces notations, l'algorithme ARLDA, pour *Adaptive Regularization Algorithm using Dynamic Accuracy*, est décrit par l'Algorithme 4.3.1. Celui-ci fait intervenir les Algorithmes 4.3.2 et 4.3.3, qui visent respectivement à déclencher l'arrêt de l'algorithme ARLDA et proposer un pas à chaque itération.

L'algorithme ARLDA possède naturellement la structure classique des algorithmes de régularisation adaptative, à savoir le calcul d'un pas candidat (étape 2), le test d'adéquation entre la décroissance du modèle et la variation de la fonction obtenue en suivant ce pas, afin de déterminer son acceptation (étape 3), puis une mise à jour du paramètre de régularisation selon ce même ratio (étape 4). De surcroît, il intègre la gestion dynamique de la précision sur les évaluations des différentes fonctions et leurs dérivées. Ceci se traduit par de possibles raffinements des précisions et/ou tolérances (étapes 3 et 5), et dans ce cas, de nouvelles évaluations de fonctions aux précisions souhaitées (étapes 1 et 3). La mise à jour de la précision sur la fonction $f$ (4.61) vise à s'assurer du bon fonctionnement du mécanisme d'acceptation du pas, basé sur la décroissance relative de $\bar{\psi}$ vis-à-vis de celle du modèle régularisé bruité $\bar{m}_k$ dans le cas d'itérations "réussies". En effet, les erreurs sur les valeurs de $\bar{\psi}(\mathbf{x}_k)$ et $\bar{\psi}(\mathbf{x}_k + \mathbf{s}_k)$, peuvent conduire à accepter des pas pour lesquels les variations de $\psi$ ne seraient pas acceptables (décroissance trop faible, voire croissance). La condition (4.61) permet de contrôler l'erreur relative sur les évaluations de $\bar{\psi}$ vis-à-vis de $\psi$, afin de garantir la convergence de $(\mathbf{x}_k)$ vers une solution du premier ordre pour une précision fixée. De la même manière, le choix des précisions sur les évaluations de $f$, $c$ et de leurs dérivées visent à garantir que l'erreur absolue sur $\bar{\psi}$ vis-à-vis de $\psi$, reste inférieure à la tolérance $\omega_k$ prescrite dynamiquement.

Dans le cas de l'Algorithme 4.3.2, visant à arrêter l'algorithme ARLDA, la mise à jour des précisions vise à contrôler les erreurs relative (4.66) et absolue (4.67) sur la variation du modèle linéarisé inexact $(\overline{\Delta\ell}_k(\mathbf{d}_k))$. En effet, cette dernière quantité permet l'évaluation de l'optimalité (au premier ordre) de l'itéré en cours (4.59). Il est donc important de majorer les erreurs vis-à-vis de son évaluation exacte $(\Delta\ell_k(\mathbf{d}_k))$, malheureusement non disponible, afin de garantir que la sortie de l'algorithme conduise bien à une solution "optimale" pour la fonction non bruitée $\psi$.

L'évolution des précisions sur les évaluations des dérivées dans l'Algorithme 4.3.3 vise à s'assurer que le candidat $s_k$ à être pas, obtenu en minimisant le modèle régularisé bruité $\bar{m}_k$, soit pertinent pour la minimisation de la fonctionnelle $\psi$. Pour cela, on s'assure que l'erreur relative sur la variation du modèle linéarisé inexact $\overline{\Delta\ell}_k(\mathbf{d}_k)$ soit bien en dessous de la tolérance $\omega_k$ prescrite dynamiquement.

Outre la convergence de cet algorithme, nous avons étudié sa complexité dans le pire cas, à savoir le nombre d'évaluations des fonctions présentes et de leurs dérivées, pour atteindre une précision $\epsilon$ sur la mesure d'optimalité (4.59). Ceci nous conduit, à la borne

$$O\Big(|\log(\epsilon)|\,\epsilon^{-2}\Big) \text{ évaluations .} \tag{4.70}$$

Ce résultat est similaire à celui obtenu pour l'algorithme inexact de régularisation AR1DA [6]. Nous retrouvons également, au facteur $|\log(\epsilon)|$ près, la borne optimale connue pour la résolution de problèmes d'optimisation lisses et non-convexes avec évaluations inexactes de la fonctionnelle et ses dérivées. Nous avons également proposé une variante de l'algorithme ARLDA, conduisant

**Algorithm 4.3.2: Arrêt de l'Algorithme 4.3.1 (ARLDA)**

**Etape 1.1.** Résolution de

$$\max_{\|\mathbf{d}\|\leq 1} \overline{\Delta\ell}_k(\mathbf{d}) \tag{4.65}$$

afin d'obtenir un maximum global $\mathbf{d}_k$ et la valeur maximale $\overline{\Delta\ell}_k(\mathbf{d}_k)$.

**Etape 1.2.**

— Si

$$\varepsilon_g + L_h\varepsilon_J + 2L_h\varepsilon_c \leq \omega_k\overline{\Delta\ell}_k(\mathbf{d}_k), \tag{4.66}$$

alors

— Définir $\overline{\phi}_k = \overline{\Delta\ell}_k(\mathbf{d}_k)$ ;

— Si $\overline{\phi}_k \leq \epsilon/(1+\omega_k)$, arrêt de l'algorithme ARLDA ;

— Sinon aller à l'étape 2 de l'algorithme ARLDA.

— Sinon, si

$$\overline{\Delta\ell}_k(\mathbf{d}_k) \leq \frac{1}{2}\epsilon \text{ et } \varepsilon_g + L_h\varepsilon_J + 2L_h\varepsilon_c \leq \frac{1}{2}\epsilon, \tag{4.67}$$

arrêt de l'algorithme ARLDA.

**Etape 1.3 :** Multiplication des tolérances $\varepsilon_g$, $\varepsilon_c$ et $\varepsilon_J$ par $\gamma_\varepsilon \in (0,1)$ et retourner à l'Etape 1 de l'algorithme ARLDA.

**Algorithm 4.3.3: Calcul du pas $\mathbf{s}_k$ pour l'Algorithme 4.3.1 (ARLDA)**

**Etape 2.1 :** Résolution approximative de

$$\min_{\mathbf{s}\in\mathbb{R}^n} \overline{m}_k(\mathbf{s}) \tag{4.68}$$

pour obtenir le pas $\mathbf{s}_k$ et les valeurs des modèles associés $\overline{\Delta m}_k(\mathbf{s}_k)$ et $\overline{\Delta\ell}_k(\mathbf{s}_k)$.

**Etape 2.2 :** Si

$$(\varepsilon_g + L_h\varepsilon_J)\|\mathbf{s}_k\| + 2L_h\varepsilon_c \leq \omega_k\overline{\Delta\ell}_k(\mathbf{s}_k), \tag{4.69}$$

aller à l'étape 3 de l'algorithme ARLDA.

**Etape 2.3 :** Sinon, multiplication des tolérances $\varepsilon_g$, $\varepsilon_c$ et $\varepsilon_J$ par $\gamma_\varepsilon \in (0,1)$ et retourner à l'Etape 1 de l'algorithme ARLDA.

à une complexité dans le pire cas de l'ordre de $O\Big(|\log(\epsilon)| + \epsilon^{-2}\Big)$ évaluations. Plus de détails se trouvent dans [73].

La suite logique de ses travaux se trouveraient dans l'analyse des performances de l'algorithme ARLDA dans le cadre d'un algorithme de type 4D-Var avec régularisation en norme $\|\|_1$. La minimisation se ferait directement sur la fonction $\psi$ non-linéaire et non-convexe. De nouveau, les évaluations inexactes de $f$ et son gradient pourraient renvoyer à la résolution tronquée des systèmes linéaires associés aux matrices de covariance d'erreurs, ou l'utilisation d'une arithmétique multi-précision lors des calculs.

## Publications associées

[65] S. Gratton, S. Guröl, E. Simon, Ph.L. Toint : Issues in making the weakly-constrained 4DVar formulation computationally efficient, *Oberwolfach Reports*, 13, 2726-2731, 2016 ;

[66] S. Gratton, S. Guröl, E. Simon, Ph.L. Toint : A note on preconditioning weighted linear squares, with consequences for weakly-constrained variational data assimilation, *Q. J. R. Meteorol. Soc.*, 144, 934-940, 2018 ;

[67] S. Gratton, S. Guröl, E. Simon, Ph.L. Toint : Guaranteeing the convergence of the saddle formulation for weakly-constrained 4D-VAR data assimilation, *Q. J. R. Meteorol. Soc.*, 147, 2067-2081, 2018 ;

[71] S. Gratton, E. Simon, D. Titley-Peloquin, Ph. L. Toint : Minimizing convex quadratics with variable precision conjugate gradients, *Numer. Linear Algebra Appl.*, 28 :e2337, 2021 ;

[73] S. Gratton, E. Simon, Ph.L. Toint : An algorithm for the minimization of nonsmooth and nonconvex functions using inexact evaluations and its worst-case complexity, *Math. Program.*, 187, 1-24, 2021 ;

[8] A. Bernigaud, S. Gratton, F. Lenti, E. Simon, O. Sohab : $L_p$-norm regularization in variational data assimilation, *Q. J. R. Meteorol. Soc.*, 147, 2067-2081, 2021 ;

[72] S. Gratton, E. Simon, D. Titley-Peloquin, Ph.L. Toint : A note on inexact inner products in GMRES, *SIAM J. Matrix Anal. Appl.*, 43 (3), 2022.

[9] A. Bernigaud, S. Gratton, E. Simon : A nonlinear conjugate gradient algorithm for $p$-norm regularized least squares with application to variational data assimilation, soumis.

# Chapitre 5

# Quelques perspectives

**Sommaire**


**5.1 Vers une assimilation de données ensembliste multi-fidèle . . . . . 76**
**5.2 Assimilation de données et non-linéarité : les méthodes à noyaux . 81**
**5.3 Des problèmes issus de l'océanographie et plus généralement de la modélisation du système climatique terrestre . . . . . 85**


*Ce chapitre présente les axes de recherche, que je souhaite développer à court et moyen termes. Certains de ces axes se traduisent notamment par des co-encadrements de thèse en cours. Ils sont en lien direct avec mes travaux précédents, tant sur les approches multi-fidèles en optimisation, algèbre linéaire numérique (via notamment l'emploi d'arithmétiques multi-précision) et assimilation de données (méthodes multi-niveaux et multi-grilles explorées notamment durant ma thèse), que sur la possibilité d'utiliser des transformations ad-hoc des données afin d'améliorer la qualités des résultats des méthodes d'assimilation de données usuelles sur des problèmes fortement non-linéaires et/ou non-gaussiens. Enfin, à plus long terme, les développements méthodologiques en modélisation, et notamment la mise en place de couplages mathématiquement rigoureux, entre les différentes composantes du systèmes Terre, rendent l'assimilation de données dans ce type de système plus complexe, de part l'emboîtement de processus itératifs dans le système d'optimalité associé à la minimisation d'une fonction de coût de type variationnelle. Se pose donc la question d'une reformulation du problème d'estimation dans sa globalité, dans l'espoir d'en minimiser sa complexité. Un dernier axe de recherche est associé à l'essor croissant de l'apprentissage machine en complément de la modélisation numérique du système Terre, ou de certaines de ses composantes.*

## 5.1 Vers une assimilation de données ensembliste multi-fidèle

Dans la continuité de mes travaux antérieurs visant à réduire les coûts et temps de calcul des algorithmes nécessaires à la résolution d'un problème d'assimilation de données, sans compromettre significativement la qualité de la solution obtenue, nous nous intéressons à l'emploi de stratégies dites multi-fidèles pour le cas des méthodes variationnelles d'ensemble. Comme déjà mentionné, la propagation d'un ensemble de simulations afin d'estimer les incertitudes notamment via la matrice de covariance d'erreur de prévision, est confrontée à la taille du domaine

et/ou la résolution spatio-temporelle de la grille. La taille de cet ensemble reste bien souvent très faible (d'une dizaine à une centaine de membres) devant les dimensions du problème, conduisant notamment à des problèmes associés au rang très faible de la matrice estimée, ainsi que des erreurs d'echantillonnage pouvant être conséquentes. La question, qui nous intéresse alors, est de savoir si pour un budget de calculs fixé pour la simulation d'ensemble, et cherchant à estimer une certaine quantité (scalaire, vecteur, matrice, etc..) depuis cet ensemble, s'il est possible d'augmenter la taille de celui-ci, via l'emploi de simulations dégradées mais aux coûts de calculs plus faibles, tout en réduisant la variance d'erreur de l'estimateur obtenu.

Un moyen naturel de réduire le coût de calcul d'une simulation numérique, potentiellement au détriment de la qualité de la solution, consiste en l'augmentation de la taille des mailles de la grille spatiale, avec la possibilité de réduire le pas de temps, si les conditions de stabilité du schéma numérique l'y autorise. Nous parlerons de grilles "grossières", relativement à la grille originelle dite "fine". Supposons disposer d'opérateurs permettant de transférer de l'information entre les différents niveaux de grilles, il est possible d'envisager l'utilisation d'approches multi-niveaux selon différentes stratégies : les approches multi-grilles [22, 76, 24, 136], traditionnellement utilisées en algèbre linéaire numérique, optimisation et pour la résolution numérique de certaines équations aux dérivées partielles, les approches Monte Carlo multi-niveaux (MLMC) [62, 63], ainsi que les liens possibles entre ces cadres méthodologiques.

Suivant [62, 63], soit $\xi \in \mathbb{R}$ une variable aléatoire, dont on cherche à estimer l'espérance mathématique. Supposons disposer d'un échantillon de taille $N$ de cette variable, $(\xi_i)_{i=1:N}$, il est possible d'estimer cette quantité via l'estimateur Monte Carlo

$$E[\xi] \approx \frac{1}{N}\sum_{i=1}^{N} \xi_i.$$

Supposons disposer de $m$ différents modèles, de complexité et coût de calculs croissant, permettant de simuler la variable $\xi$. Il est alors possible de générer des échantillons, depuis ces différents modèles, en vue de l'estimation de $E[\xi]$. L'estimateur Monte Carlo multi-niveaux s'écrit alors

$$\hat{\xi} = \frac{1}{N_0}\sum_{i=1}^{N_0} \xi_{0,i}^0 + \sum_{l=1}^{m} \frac{1}{N_l}\sum_{i=1}^{N_l} (\xi_{l,i}^l - \xi_{l-1,i}^l),$$

avec $(\xi_{l,i}^p)_{i=1:N_l}$ un échantillon de $\xi$ obtenu par des simulations du modèle $l$ depuis une variable aléatoire adaptée aux caractéristiques du modèle $p \in \{l-1, l\}$, avec $l = 0$ le modèle le plus grossier (et moins coûteux). Par exemple, soit le cas d'un modèle ayant différentes résolutions spatiales, pour le niveau de grille $l$, les variables $\xi_{l,i}^l$ et $\xi_{l-1,i}^l$ peuvent s'écrire

$$\xi_{l,i}^l = M_l(\omega_i^l), \quad \xi_{l-1,i}^l = I_{l-1}^l(M_{l-1}(I_l^{l-1}(\omega_i^l))),$$

avec $M_l$ et $M_{l-1}$ les modèles pour les grilles $l$ et $l-1$, $I_l^{l-1}$ et $I_{l-1}^l$ des opérateurs de restriction et prolongation entre les grilles $l$ et $l-1$, et $\omega_i^l$ la réalisation $i$ d'une variable aléatoire sur la grille $l$. Sous des hypothèses portant sur les vitesses de décroissance de la moyenne de l'erreur selon les niveaux $l$, de la variance des corrections $\xi_l - \xi_{l-1}$, et la croissance du coup du calcul d'un échantillon sur les niveaux $l$, il est possible de construire un estimateur multi-niveaux de $\xi$, tel que l'erreur moyenne quadratique de celui-ci soit bornée par une quantité fixée a priori. Ceci s'énonce par le Théorème 5.1.1. [63] :

**Theorem 5.1.1.** *Soit $\xi$ une variable aléatoire. On note $\xi_l$ une approximation de $\xi$ évaluée sur le niveau de grille $l$. Supposons les estimateurs $z_l = \frac{1}{N_l}\sum_{i=1}^{N_l}(\xi_{l,i}^l - \xi_{l-1,i}^l)$ indépendants avec $l = 0 : m$, avec la convention $\xi_{-1}^0 = 0$, et notons respectivement $C_l$ et $V_l$ le coût de calcul et la variance de $z_l$.*

*S'il existe des constantes strictement positives $\alpha, \beta, \gamma, c_1, c_2, c_3$ telles que $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ et pour tout $l = 1 : m$*

*$(H_1)$ $|E[\xi_l^l - \xi]| \leq c_1 2^{-\alpha l}$,*

*$(H_2)$ $E[z_l] = E[\xi_0^0]$ si $l = 0$, $E[z_l] = E[\xi_l^l - \xi_{l-1}^l]$ sinon,*

*$(H_3)$ $V_l \leq c_2 2^{-\beta l}$,*

*$(H_4)$ $C_l \leq c_3 2^{\gamma l}$,*

*alors il existe une constante strictement positive $c_4$ telle que pour tout $\epsilon \leq \frac{1}{e}$, il existe $m \in \mathbb{N}$ et $(N_l)_{l=1:m} \in \mathbb{N}^m$ tels que l'estimateur $z = \sum_{l=0}^{m} z_l$ vérifie les propriétés suivantes :*

- *$E[(z - E[\xi])^2] \leq \epsilon^2$,*
- *$E[C] \leq c_4 \epsilon^{-2} f(\alpha, \beta, \gamma, \epsilon)$*

*avec $C$ le coût moyen du calcul de $z$ et $f$ définie par : $f(\alpha, \beta, \gamma, \epsilon) = 1$, si $\beta > \gamma$, $f(\alpha, \beta, \gamma, \epsilon) = \log(\epsilon)^2$, si $\beta = \gamma$, et $f(\alpha, \beta, \gamma, \epsilon) = \epsilon^{-\frac{\gamma-\beta}{\alpha}}$, si $\beta < \gamma$.*

Si les hypothèses $(H_1)$-$(H_4)$ sont vérifiées, il est alors possible de construire un estimateur multi-niveaux de l'espérance de $\xi$, pour lequel la variance est bornée par une quantité cible. Néanmoins, les hypothèses peuvent ne pas être vérifiées en pratique, selon le problème considéré, et notamment la décroissance exponentielle de la variance de l'estimateur $z_l$ (terme de correction sur le niveau de grille $l$) en fonction des niveaux de grilles (hypothèse $H_3$).

Ce cadre méthodologique est notamment à l'origine des filtres de Kalman d'ensemble multi-niveaux [82, 31], qui exploitent la possibilité de disposer de grilles temporelles ou spatio-temporel les de différentes résolutions. Ceux-ci se basent sur l'application des méthodes Monte Carlo multi-niveaux à l'estimation des moyenne et matrice de variance-covariance des vecteurs d'état des étapes de prévision et d'analyse. Néanmoins, dans le cadre d'expériences réalisées pour des problèmes d'estimation de propriétés géologiques de réservoirs, il a été illustré [56] les possibles difficultés à satisfaire les hypothèses du Théorème 5.1.1, et notamment l'hypothèse $H_3$ portant sur la décroissance des variances des termes de correction entre deux niveaux de grilles. Ceci peut alors conduire à une dégradation des performances des filtres multi-niveaux, compromettant ainsi leur utilisation, comparativement aux filtre de Kalman d'ensemble, à budget de calculs équivalent.

Dans le cadre de la thèse de J. Briant, co-encadrée avec S. Gratton, P. Mycek, M. Destouches, A. Weaver et S. Gürol, nous nous intéressons à l'emploi des méthodes MLMC pour l'estimation de quantités vectorielles et matricielles, avec pour objectif l'estimation de la matrice de covariance d'erreur de prévision, ou de certains paramètres vectoriels intervenant dans sa définition. Dans ce cadre, les quantités estimées dépendent du niveau de grille le plus fin, que nous sommes en mesure d'utiliser pour les simulations d'ensemble, et ainsi des fréquences représentables sur cette grille. Se pose donc la question de savoir si le gain de variance espéré, sous les hypothèses du Théorème 5.1.1, se vérifie pour toutes les fréquences représentables sur la grille la plus fine, ou si les corrections associées aux différents niveaux de grilles tendent à favoriser les basses fréquences, notamment du fait de l'emploi d'ensembles de plus grandes tailles sur les grilles les plus grossières (coût d'une simulation réduit par rapport aux simulations haute résolution).

Dans le cadre d'expériences numériques sur un problème simplifié, à savoir l'estimation de la matrice de variance-covariance d'un vecteur aléatoire, définie depuis un noyau gaussien dont les paramètres furent prescrits a priori pour définir la référence, il a été observé que l'application directe des méthodes de Monte Carlo multi-niveaux, avec un choix simpliste des opérateurs de transferts inter-grilles, conduisait à une réduction de l'erreur quadratique moyenne de la variable aléatoire estimée, comparativement à l'emploi de l'approche Monte Carlo sur la grille la plus fine (à budget de calculs équivalent). Néanmoins, l'étude de cette même erreur pour chacun des modes de Fourier associés à la grille la plus fine, nuance ces résultats : alors que la réduction de l'erreur fut la plus forte sur les modes associés aux plus basses fréquences, les erreurs associées aux fréquences non représentables sur les grilles grossières, pouvaient augmenter significativement selon les fréquences, comparativement à l'estimateur de Monte Carlo haute résolution. Ceci n'est évidemment pas acceptable dans le cadre de la prévision d'écoulements de fluides géophysiques, où un des objectif du raffinement des grilles spatiales est la meilleure représentation des phénomènes non-linéaires de petites échelles, ainsi que leurs interactions avec ceux de plus grandes échelles. Ce phénomène, peu étudié à notre connaissance dans la communauté Monte Carlo multi-niveaux, est quant à lui, bien connu du cadre des approches multi-grilles géométriques, et renvoie à la définition des opérateurs de transferts inter-grilles en lien avec l'opérateur différentiel approximé sur les différents niveaux de grilles. Suivant [76, Note 3.5.1], les opérateurs d'interpolation et de restriction sont supposés vérifier

$$m_i + m_r > 2m_D,$$

avec $m_i$ l'ordre de l'interpolation, $m_r$ l'ordre de l'interpolation dont l'opérateur de restriction est l'adjoint, et $m_D$ l'ordre de l'opérateur différentiel. Pour les opérateurs d'interpolation, ceux-ci sont d'ordre $m_i$, dès lors qu'ils interpolent exactement les polynômes de degrés $m_i - 1$. Les travaux de Jérémy, tant théorique sur une analyse de Fourier locale simplifiée (bi-grille) des approches Monte Carlo multi-niveaux avec des opérateurs de transferts d'ordre plus élevés (filtres de Shapiro), que pratique dans un cadre idéalisé, illustrent l'importance du bon choix de ces opérateurs de transferts sur les performances de l'approche Monte Carlo multi-niveaux.

Néanmoins, de nombreuses pistes de recherche restent à explorer en lien avec ces travaux, en vue d'appliquer ce types de stratégies sur des systèmes réalistes, que je liste par la suite.

1. *Apport des méthodes multi-grilles au cadre méthodologique des méthodes Monte Carlo multi-niveaux ?*
   Le cadre méthodologique des approches multi-grilles, notamment géométriques, me semble naturel pour interpréter les conditions de convergence et anticiper les performances des approches Monte Carlo multi-niveaux, dès lors que la réduction des coûts de calcul du modèle est lié à une dégradation de la grille spatio-temporelle utilisée dans la discrétisation du problème. C'est ce à quoi nous nous sommes intéressés dans la thèse de Jérémy. Des travaux récents [90] ont également suggéré la combinaison des approches multi-grilles et Monte Carlo multi-niveaux pour des problèmes de quantification d'incertitudes pour des équations aux dérivées partielles elliptiques avec coefficients aléatoires. Tandis qu'il est possible de construire formellement des approches multi-grilles satisfaisant l'hypothèse $(H_3)$ du Théorème 5.1.1, la vérification de l'hypothèse $(H_4)$ s'avère numérique et spécifique au problème étudié. Or celui-ci renvoie à la résolution d'équations aux dérivées partielles elliptiques.

   Cette notion d'ellipticité est importante dans la théorie des méthodes multi-grilles géométri ques. Celle-ci peut concerner le problème continu et ses discrétisations (il est alors question

de "h-ellipticité" [22]), et traduit le fait que les hautes fréquences sont locales : des perturbations hautes fréquences du second membre du système associé au problème considéré, engendreront des perturbations hautes fréquences de la solution au voisinage de celles du second membre. Pour ce type de problème, il est possible de construire des lisseurs itératifs performants sur les hautes fréquences de l'erreur, qu'il est possible de combiner avec des étapes de corrections sur des grilles plus grossières. Se pose donc la question de savoir dans quelle mesure cette notion de h-ellipticité joue un rôle dans les bonnes performances des approches Monte Carlo multi-niveaux. Est-ce une condition suffisante pour satisfaire l'hypothèse $(H_4)$ sur la décroissance des variances des corrections sur les différents niveaux de grilles ? Si impact fort il y a, comment traiter des problèmes non-elliptiques ? A l'instar des approches multi-grilles pour l'assimilation variationnelle de données [37], ne faudrait-il pas plutôt appliquer les approches Monte Carlo multi-niveaux à la construction de préconditioneurs efficaces ?

2. *Des estimateurs de matrices de covariance d'erreur possiblement indéfinis*
   Ces approches Monte Carlo multi-niveaux appliquées à l'estimation de matrices de covariance d'erreur peuvent conduire à un estimateur matriciel indéfini, à savoir pour lequel les valeurs propres sont de signe quelconque. La somme télescopique de termes de correction sur les niveaux de grille, introduit des différences de matrices semi-définies positives, dont la somme n'est pas garantie être semi-définie positive. Ce problème est ainsi anticipé dans [31, Remark 3], où il est proposé de ne garder que les couples propres associés aux valeurs propres positives ou nulles de $HP^f_{MLMC}H^T$. Des expériences numériques réalisées par M. Destouches avec un modèle quasi-géostrophique, ont illustré ce phénomène et soulèvent de nombreuses questions, qui sont liées aux ordres de grandeurs de ces valeurs propres négatives, eux-mêmes similaires à ceux des valeurs propres positives : il semble en effet difficile de les associer au noyau "numérique" de la matrice, contrairement à des valeurs négatives de très faible amplitude relative. Se pose donc les questions de savoir ce que représentent les sous-espaces propres associés à ces valeurs propres négatives, pourquoi ils apparaissent, et comment y remédier, soit dans la construction de l'estimateur, soit dans la résolution du système linéaire associé au gain de Kalman.

3. *Vers une adaptabilité de ces approches en cadre avec les contraintes de l'assimilation de données*
   Les approches Monte Carlo multi-niveaux requièrent la spécification du nombre de niveaux disponibles $(m+1)$, ainsi que de la taille des échantillons $(N_l)_{l=0:m}$ nécessaires au calcul des estimateurs $(z_l)_{l=0:m}$. Ceci peut être réalisé de manière adaptative, en visant une variance d'erreur cible sur l'estimateur Monte Carlo multi-niveaux pour un coût de calcul minimal, ou bien un coût de calcul cible pour une variance minimale [62, 63, 103]. Dans ces deux cas, étant en mesure d'évaluer empiriquement les coûts et variances des correction à chaque niveau, et sous certaines hypothèses, il est possible d'estimer itérativement ces paramètres "optimaux".

   Dans le cas de système d'assimilation de données opérationnels, tels que ceux rencontrés en géosciences, ces approches semblent plus difficilement applicables. D'une part, car le nombre de niveaux de grilles possibles est limité, de par la complexité de la mise en place d'une telle hiérarchie pour une configuration réaliste, et les coûts de calculs élevés engendrés, sans raffinement local. L'approche la plus naturelle semble, dans un premier temps, de fixer les coûts de calculs envisageables, sur quelques niveaux de grilles fixés a priori, et estimer les tailles optimales des différents échantillons, afin de minimiser la

variance de l'estimateur. La question qui se pose à nous est la pertinence de ces hyper-paramètres initiaux (taille des ensembles) après de nombreux cycles d'assimilation : les variances des corrections estimées lors du premier cycle d'assimilation, nécessaire pour le calcul des tailles des ensembles, sont-elles de bonnes approximations des corrections futures (impact des observations, changements locaux de la dynamique, apparitions d'évenements rares, etc..) ? Dans le cas contraire, les tailles des ensembles seraient sub-optimales, et un gain en terme de réduction de la variance de l'estimateur pourrait être espéré. Le calcul des nouvelles tailles de l'ensemble pouvant être coûteux, s'il est nécessaire de générer de nouveaux membres sur les différents niveaux de grilles, se pose donc la question de la définition de critères objectifs, indiquant le besoin de réévaluer la pertinence du choix des tailles d'ensemble.

## 5.2 Assimilation de données et non-linéarité : les méthodes à noyaux

Dans la continuité de mes travaux antérieurs sur les extensions non-gaussiennes des filtres et lisseurs de Kalman d'ensemble, nous étudions la possibilité d'étendre ces algorithmes à des problèmes non-linéaires, possiblement non-gaussiens, via l'introduction de méthodes dites à noyaux. Ces méthodes, très populaires dans le cadre de l'analyse de données et des problématiques d'apprentissage machine [83], permettent notamment d'étendre certaines classes de méthodes, originellement définies ou supposant implicitement des modèles linéaires entre les variables (analyse en composantes principales, séparateur à vaste marge, régression ridge, etc..) à des modèles non-linéaires.

Pour cela, il est possible d'introduire les espaces de Hilbert à noyau auto-reproduisant, ou RKHS pour *reproducing kernel Hilbert spaces*. Ceux-ci peuvent être définis de la manière suivante :

**Definition 5.2.1** (RKHS)**.** Soit $\mathcal{X}$ un ensemble et notons $H$ un espace de Hilbert de fonctions de $\mathcal{X}$ à valeurs dans $\mathbb{R}$.
$H$ est un espace de Hilbert à noyau auto-reproduisant (RKHS) si $\forall x \in \mathcal{X}$, l'évaluation ponctuelle en $x$, notée $\delta_x : H \to \mathbb{R}$ et définie par $\forall f \in H \ \delta_x(f) = f(x)$, est une forme linéaire continue sur $H$.

Outre le fait que l'évaluation ponctuelle est bien définie en tout élément de $\mathcal{X}$, il en résulte également pour ce type d'espace de Hilbert que celle-ci s'écrit comme un produit scalaire avec un élément de $H$, d'après le théorème de représentation de Riesz :

$$\forall x \in \mathcal{X}, \exists! k_x \in H \text{ t.q. } \forall f \in H, f(x) =< f, k_x >_H, \tag{5.1}$$

avec $< .,. >_H$ le produit scalaire sur H, lui conférant sa structure Hilbertienne. Il est donc possible d'identifier un élément $x \in \mathcal{X}$ à une fonction $k_x \in H$, qui vit dans un espace ayant une structure potentiellement plus riche que l'ensemble $\mathcal{X}$. On appelle noyau reproduisant de H (il est unique) l'application $k$ définie par

$$\begin{array}{rccc} k: & \mathcal{X} \times \mathcal{X} & \to & \mathbb{R} \\ & (x, y) & \mapsto & < k_x, k_y >_H \end{array}$$

avec $(k_x, k_y) \in H^2$ défini depuis (5.1). L'application $k$ est un noyau positif, au sens où $k$ est symétrique, et $\forall (x_i)_{i=1:p} \in \mathcal{X}^p$, la matrice $\mathbf{K} \in \mathcal{M}_p(\mathbb{R})$, telle que $\forall (i, j) \in [\![1, p]\!]^2, K_{i,j} = k(x_i, x_j)$, est semi-définie positive.

D'un point de vu général, il est possible de construire un noyau positif $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, en se donnant une application $\phi : \mathcal{X} \to H$, avec $H$ Hilbert, et en posant

$$\forall (x, y) \in \mathcal{X}^2, k(x, y) =< \phi(x), \phi(y) >_H . \tag{5.2}$$

Réciproquement, le théorème d'Aronszajn [3] garantit que pour tout noyau positif $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, il existe un espace de Hilbert $H$ et une application $\phi : \mathcal{X} \to H$ tels que (5.2) est vérifiée. Il est ainsi possible de réaliser implicitement des produits scalaires sur des données "transformées" $(\phi(x_i))_{i=1:p} \in H^p$, sans même nécessairement connaître la transformation $\phi$, "simplement" en évaluant le noyau sur les données $(x_i)_{i=1:p} \in \mathcal{X}^p$.

Dans le cas de la régression de ridge, le problème vise à estimer le vecteur de paramètre d'un modèle linéaire $\forall \mathbf{x} \in \mathbb{R}^n, f(\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{x}$ depuis des données, via la minimisation d'un terme de moindres carrés linéaires régularisés :

$$(\mathcal{P}) \quad \min_{\boldsymbol{\beta} \in \mathbb{R}^n} J(\boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2^2 + \frac{\lambda}{2} \|\boldsymbol{\beta}\|_2^2,$$

avec $\mathbf{X} \in \mathcal{M}_{p,n}(\mathbb{R})$ et $\mathbf{y} \in \mathbb{R}^p$ associés aux données respectivement d'entrée et sortie du modèle, et $\lambda > 0$. Avec $\mathcal{X} = \mathbb{R}^n$ et le noyau linéaire, défini par $\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^2$ $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$, le problème s'ecrit

$$(\mathcal{P}_{RKHS}) \quad \min_{f \in H} \tilde{J}(f) = \frac{1}{2} \sum_{i=1}^{p} (f(\mathbf{x}_i) - y_i)^2 + \frac{\lambda}{2} \|f\|_H^2,$$

avec $H$ le RKHS de noyau reproduisant le noyau linéaire, qui correspond, dans ce cas, aux formes linéaires continues sur $\mathbb{R}^n$. Il est alors possible d'étendre la régression de ridge à n'importe quel RKHS, soit en choisissant directement l'espace $H$, soit un noyau positif $k$, et implicitement un RKHS par le théorème d'Aronszajn. Le théorème de représentation (ou du représentant) conclut que les solutions du problème $(\mathcal{P}_{RKHS})$ sont à chercher dans le sous-espace vectoriel de $H$ engendré par les fonctions $(k(., \mathbf{x}_i))_{i=1:p} \in H^p$, qui est ainsi de dimension finie (même si $H$ ne l'est pas). Les solutions de $(\mathcal{P}_{RKHS})$ s'écrivant $f = \sum_{i=1}^{p} \alpha_i k(., \mathbf{x}_i)$, avec $\boldsymbol{\alpha} \in \mathbb{R}^p$, le problème s'écrit alors comme la résolution du problème

$$(\tilde{\mathcal{P}}_{RKHS}) \quad \min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \bar{J}(\boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{K}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 + \frac{\lambda}{2} \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha},$$

avec $\mathbf{K} \in \mathcal{M}_p(\mathbb{R})$ contenant l'évaluation du noyau $k$ sur les données $(\mathbf{x}_i)_{i=1:p} \in \mathcal{X}^p$. Celui-ci correspond à un problème aux moindres carrés linéaires, similaire au problème originel $(\mathcal{P})$, mais de dimension $p$ à savoir le nombre d'observations dont on dispose, et non plus de la dimension du vecteur de paramètre. Le noyau $k$ ou son RKHS associé, apparaît alors comme un hyper-paramètre, et permet d'étendre la méthode à des modèles $f$ plus complexes, typiquement non-linéaires.

Des travaux récents de [98] vont dans cette direction en proposant une première formulation d'un algorithme d'assimilation de données ensembliste basé sur les noyaux "RBF" (radial-basis-function ; [25]). Il est alors proposé d'approximer le terme d'innovation par une combinaison linéaire de noyaux RBF, dont les paramètres sont estimés au cours du processus d'assimilation. L'introduction de RKHS est également proposée pour modéliser l'évolution temporelle de système dynamique en vu d'appliquer des méthodes d'ensemble en assimilation de données [64, 85]. De plus, l'étape d'analyse de l'algorithme *Random Feature Map Data Assimilation* [64] peut être vue comme l'application d'un filtre de Kalman d'ensemble dans un RKHS particulier.

Dans le cadre de la thèse de S. Mauran, débutée en octobre 2021, et co-encadrée par moi-même, S. Mouysset et L. Bertino, nous nous intéressons à reformuler certains algorithmes ensemblistes d'assimilation de données sous la forme d'un problème d'optimisation sur un RKHS quelconque, afin d'étendre l'étape d'analyse de ces approches, optimales dans un cadre linéaire gaussien, à des problèmes non-linéaires. J'énumère par la suite différents axes de recherche en lien avec ces travaux.

1. *Une extension de l'étape d'analyse des filtres de Kalman d'ensemble par méthodes à noyaux*
   Les méthodes de Kalman d'ensemble, filtres, lisseurs, itératifs ou non, sont associées à la minimisation d'une fonctionnelle de type moindres carrés régularisés. Par exemple, pour le cas d'un opérateur d'observation linéaire, il est possible de dériver l'ETKF depuis la résolution du problème d'optimisation suivant

   $$\min_{\mathbf{w}\in\mathbb{R}^N} J(\mathbf{w}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f - \mathbf{H}\mathbf{A}^f\mathbf{w}\|^2_{\mathbf{R}^{-1}} + \frac{N-1}{2}\|\mathbf{w}\|^2_2,$$

   avec $N$ la taille de l'ensemble, $\bar{\mathbf{x}}^f$ la moyenne empirique de l'ensemble de prévision, et $\mathbf{A}^f$ la matrice d'anomalies de prévision. A l'instar du problème $(\mathcal{P})$, il est naturel d'introduire le noyau linéaire, afin de réécrire l'ETKF comme un problème d'optimisation sur le RKHS associé au noyau linéaire, puis de changer de noyau selon la nature des données. Néanmoins, le terme résiduel de $J$ étant dans l'espace des observations, le théorème de représentation implique que les solutions seront fonctions des variables observées uniquement. Quid de la correction des variables d'état non observées ? Il est donc nécessaire de reformuler le problème d'optimisation originel afin d'être en mesure d'introduire les variables d'état dans le sous-espace vectoriel du RKHS dans lequel sont construits les solutions. Ceci peut notamment être fait, par augmentation du vecteur résiduel, comme il a été proposé dans le cadre de l'extension de l'EnKF aux opérateurs d'observation non-linéaires [48, Chapter A.1] Ceci conduit à la résolution d'un problème aux moindres carrés similaire à $(\tilde{\mathcal{P}}_{RKHS})$, permettant le calcul de l'état analysé moyen. Une stratégie doit ensuite être définie pour la génération de l'ensemble d'analyse, centré sur l'état analysé obtenu lors de la résolution de $(\tilde{\mathcal{P}}_{RKHS})$. Plusieurs stratégies sont envisageables : exploiter la pseudo-inversion de la Hessienne de la fonctionnelle à l'optimum (ou d'un bloc de celle-ci) en vue d'une factorisation de la matrice $\mathbf{P^a}$ ou définir une approche d'ensemble d'analyse sur des données perturbées, la question étant alors de savoir quelles données et comment les perturber (quelle distribution de probabilité) ?

2. *Des difficultés prévisibles liées à la dimension du problème*
   Une première difficulté est liée à la dimension du problème $(\tilde{\mathcal{P}}_{RKHS})$ : alors que l'ETKF peut être vu sous l'angle d'un problème de minimisation, de dimension celle de l'ensemble $(N)$ petite en pratique, la dimension de $(\tilde{\mathcal{P}}_{RKHS})$ va correspondre aux données disponibles, soit de l'ordre de la taille du vecteur d'observation plus celle du vecteur d'état. En pratique, ceci conduit à résoudre des systèmes linéaires de la taille du vecteur d'observation, et chercher les couples propres dominants d'une matrice pleine de taille celle du vecteur d'état. Il va donc falloir repenser le problème pour ne pas avoir à construire de telles matrices. Ce problème des méthodes à noyau n'est pas nouveau et est bien connu dans le domaine de l'apprentissage machine. Il y est notamment proposé des approches visant à approximer la matrice de Gram par une matrice de rang faible [5]. Nous y retrouvons des algorithmes déterministes ou stochastiques notamment basés sur la méthode de Nyström [44, 91, 102]. Il serait donc intéressant d'évaluer la faisabilité de ces approches pour les

problèmes d'assimilation de données en très grande dimension, ainsi que les performances de ce type d'approximation sur la stabilité du filtre au cours des cycles d'assimilation.

3. *Choix du noyau et estimation de ses hyper-paramètres, lien possible avec l'anamorphose gaussienne ?*
   Une seconde difficulté réside dans le choix de noyaux pertinents vis-à-vis du problème applicatif, ainsi que l'estimation des hyper-paramètres associés. Le cas le plus simple, pour le choix du noyau, est celui pour lequel nous disposons de la connaissance du modèle régissant les données, ou de certaines de ses caractéristiques, et pouvons le relier à un RKHS, nous donnant ainsi le noyau reproduisant de ce RKHS comme candidat possible. Néanmoins, ce cas semble peu probable pour l'assimilation de données pour les fluides géophysiques. Dans ces conditions, un axe de recherche possible est l'emploi de méthodes à noyaux adaptatives [98, 149]. L'idée est alors d'utiliser un mélange de noyaux, souvent gaussiens, pour approximer une variable d'état, un résidu dans l'espace des observations ou une densité de probabilité, et d'en estimer les paramètres au cours du processus d'assimilation de données, dans l'étape d'analyse et/ou de prévision.

   Une autre question d'intérêt me semble être l'analyse du lien possible entre les extensions des filtres de Kalman d'ensemble par méthodes à noyaux et par anamorphose gaussienne [10]. L'anamorphose gaussienne, appliquée à une variable d'état, un paramètre, et/ou une observation peut être interprétée sous l'angle de l'apprentissage machine comme une application permettant de passer de l'espace des données vers un nouvel espace (appelé celui des *features*). Dans ces conditions, il est alors possible de définir un noyaux depuis cette application, et ainsi obtenir un RKHS associé. L'extension par anamorphose gaussienne des filtres de Kalman d'ensemble apparaît comme une extension par méthodes à noyaux dans un RKHS particulier. La question se pose donc de savoir quel est ce RKHS de noyau reproduisant défini depuis la fonction anamorphose. De surcroît, dans quelle mesure la caractérisation de celui-ci pourrait-elle nous apporter des informations pour la définition de fonctions anamorphoses multivariées ?

4. *Un cadre unificateur pour les extensions par méthodes à noyaux de l'assimilation de données ensembliste ?*
   Les travaux récents, tant sur l'utilisation de noyaux pour l'étape de prévision ensembliste [64, 85, 149], que sur celle d'analyse [98, 149] (ainsi que la thèse de S. Mauran) permettent d'envisager le développement d'approches par méthodes à noyaux pour le processus complet d'estimation bayésienne, à savoir dans les phases de prévision et d'analyse. Un premier algorithme [149] propose ainsi d'utiliser des noyaux dans ces deux étapes pour approximer la densité de probabilité de filtrage au cours de l'évolution du système. L'approche repose sur l'introduction d'un mélange de noyaux gaussiens, à la fois pour la densité de probabilité de prévision, et celle d'analyse. En pratique, ceci conduit à minimiser des fonctionnelles de type moindres carrés pour estimer certains paramètres des noyaux, tels que leur poids dans le mélange et leur matrice de variance-covariance. Les expériences numériques, réalisées sur des modèles de petites dimensions dont le modèle Lorenz-96, montrent que cette approche permet de réduire significativement l'erreur RMS comparativement à l'utilisation d'un filtre particulaire. Comparativement à un EnKF, les performances sont similaires sur le système Lorenz-96 avec observation directe de quelques variables (100 membres pour l'EnKF), et très nettement meilleures dans un problème de suivi de trajectoire avec un opérateur d'observation non-linéaire. Néanmoins, pour ces dernières expériences, les temps de calcul de l'approche à noyaux proposée sont proches de ceux d'un EnKF avec 10000

membres (problème de dimension 4). Ceci laisse penser que cette approche, telle que proposée, ne pourra pas traiter des problèmes de grande dimension, tels que ceux rencontrés dans les géosciences. En effet, celle-ci requiert l'estimation d'autant de matrices de variance-covariance, de dimension celle du vecteur d'état, que de noyaux présents dans le mélange, et ceux par la minimisation de problèmes aux moindres carrés. Il apparaît nécessaire de revisiter la modélisation par méthodes à noyaux du problème d'estimation bayésienne, en introduisant les contraintes associées à la dimension des problèmes envisagés.

## 5.3 Des problèmes issus de l'océanographie et plus généralement de la modélisation du système climatique terrestre

Enfin, à l'avenir, j'aimerais pouvoir m'intéresser à des problèmes issus de l'océanographie, notamment ceux découlant des limitations de la puissance de calculs. A court et moyen termes, l'utilisation de grilles spatio-temporelles, aussi fines que souhaitées, reste impossible. Il en est de même pour l'utilisation de méthodes de couplage avancées, dans les systèmes complexes, tels que ceux visant à modéliser le système climatique terrestre. Cette dernière section est moins développée, et renvoie à des réflexions sur un plus long terme, même si un premier axe de recherche est déjà en phase d'exploration, via une collaboration avec des chercheurs du LEGOS et de l'IRIT, qui se structure. D'autres collaborations avec des experts en modélisation numérique et couplage sont naturellement envisagées pour l'avancée de ces travaux.

1. *Océan numérique : apport de l'apprentissage machine pour l'estimation et la simulation de certaines inconnues ?*
   L'essor de l'apprentissage machine, et plus récemment des réseaux de neurones, permet d'envisager l'emploi de nouvelles approches, complémentaires de celles issues du formalisme de l'assimilation de données, à l'estimation voire l'émulation de certaines quantités inconnues, telles que l'erreur modèle, et notamment sa composante liée à la non représentation des échelles fines sur des niveaux de grille fixés. En effet, la difficulté, voire l'impossibilité, de l'usage de grilles à suffisamment haute résolution pour représenter certaines fines échelles sur de vastes domaines spatio-temporels, peut conduire à une mauvaise représentation de certaines structures à plus grandes échelles, pourtant représentables sur ces niveaux de grilles. Il en va tout aussi bien de l'océan physique, que de sa composante biogéochimique. De plus, ces modèles intègrent des paramétrisations visant à modéliser les interactions aux bords du domaine (par exemple, celles avec l'atmosphère) et des schémas de fermeture du système, pour lesquels des incertitudes subsistent, tant sur leurs formulations, que sur les paramètres qu'elles introduisent. Avec les nombreuses simulations numériques opérées à différentes résolutions, avec différents modèles, incluant ou non des résultats de systèmes d'assimilation de données (réanalyses), nous disposons ainsi d'une masse de données conséquente permettant d'envisager l'usage d'approches d'apprentissage machine. Ainsi, différentes approches combinant apprentissage machine et assimilation de données ont été récemment proposées pour émuler la dynamique d'un système en lieu et place de la résolution de systèmes d'équations différentielles [20, 15], pour estimer et corriger en partie l'erreur modèle [51, 52], pour modéliser de nouvelles paramétrisations sous-mailles [21], voire reformuler certaines étapes de l'algorithmie de l'assimilation de données depuis des réseaux de neurones [115]. Des travaux préliminaires suggèrent qu'il est également possible d'estimer certaines paramétrisations, soit numériquement par réseaux de neurones convolutifs, soit par l'apprentissage d'un opérateur différentiel, sans nécessairement disposer d'un

système d'assimilation [146, 147].

Dans le cadre d'une collaboration avec S. Zhang (IRIT), L. Renault (LEGOS) et R. Benshila (LEGOS), nous nous intéressons par la possibilité d'utiliser de telles approches, dans le but d'apprendre des paramétrisations de fermeture, dites *scale aware*, pour la composante océanique de systèmes climatiques afin d'émuler l'impact des fines échelles, non représentables sur les grilles spatio-temporelles utilisées dans les prévisions climatiques, sur les plus larges échelles. L'idée serait d'apprendre une paramétrisation, depuis des données issues de simulations numériques à différents niveaux de grilles, qui puisse s'adapter à la résolution de la grille utilisateur, et corriger la dynamique du système représentable sur celle-ci, sans aide de l'utilisateur. A ce stade, nous aimerions nous passer d'un système d'assimilation de données, afin de pouvoir déployer, les approches proposées, à des systèmes n'en disposant pas. Outre des difficultés et interrogations quant à la nature et la structure de la paramétrisation (des réseaux de neurones ? Si oui, lesquels ? etc..), notre capacité d'apprentissage (quel critère minimiser ?), se posent également les difficultés liées au couplage de cette paramétrisation avec un modèle d'océan réaliste, tant d'un point de vue logiciel que numérique (source possible d'instablités numériques ?). Ces travaux sont complémentaires de ceux qui vont être réalisés sur la possibilité d'estimer l'erreur modèle par apprentissage machine, en collaboration avec S. Gratton (IRIT) dans le cadre du PPR Mediation, qui vient d'être lancé. De nouveau, se pose la question de l'exploitation des masses de données produites dans le cadre de simulations numériques à différentes échelles.

2. *Assimilation de données et décomposition de domaines pour les modèles couplés*

   Un autre axe de recherche porte sur le développement d'algorithmes d'assimilation de données pour des modèles couplés via des méthodes de Schwarz [58], tels que ceux envisagés pour l'océan et et l'atmosphère. Le développement de modèles couplés océan-atmosphère, tant pour la prévision à court terme du temps, que pour les simulations climatiques, a conduit naturellement à s'interroger sur la pertinence des stratégies de couplages envisagées [92], et au développement de systèmes d'assimilation de données spécifiques à ces modèles et aux nouvelles difficultés rencontrées (cf [148] pour une veille scientifique et les références associées). Les stratégies de couplage les plus usitées sont asynchrones de part leur faible coût de calcul : les composantes océanique et atmosphérique communiquent séquentiellement au cours du temps, via le calcul de flux à l'interface depuis des moyennes temporelles des variables d'état de l'autre composante. Ceci a pour effet d'engendrer un décalage temporel des échanges à l'interface des deux composantes, l'une étant en avance sur l'autre au cours du temps, et ne permet donc pas de résoudre "proprement" le problème du couplage. Pour ce type de systèmes, différentes stratégies d'assimilation ont été proposées, allant du faiblement couplé, à savoir la résolution de l'étape d'analyse dans chacune des composantes et ce de manière séparée, au fortement couplé pour lequel l'analyse se fait sur l'ensemble des composantes du système. Outre des difficultés pouvant relever du développement logiciel (adjoint d'un modèle couplé, HPC, ..), la modélisation et/ou l'estimation de la matrice de covariance d'erreur de prévision (ainsi que celle de l'erreur modèle) est complexe et soumise aux erreurs d'echantillonage, dues à un ensemble de faible rang devant les dimensions du problème, ce qui compromet notamment les covariances d'erreur entre les variables des différentes composantes du système. L'emploi de méthodes d'ensemble pour ce type de modèle requiert donc la définition de stratégies de localisation de l'étape d'analyse adaptées aux composantes couplées du système, ce qui est loin d'être trivial dans le cas de l'océan et de l'atmosphère, pour lesquels les dynamiques d'évolution se font à des échelles temporelles très différentes selon les composantes. Enfin,

les déséquilibres des états, entre les différentes composantes du système, issus de l'analyse peuvent conduire à des chocs d'assimilation, et à la génération d'ondes "parasites" pouvant se propager dans tout le système au fil du temps.

Plutôt que de continuer à adapter les algorithmes d'assimilation de données aux modèles couplés actuels, il pourrait être judicieux de modéliser conjointement les algorithmies du couplage de modèles et de l'assimilation de données. En effet, les stratégies de couplage de modèles basées sur la convergence de processus itératifs de quantités définies à l'interface des domaines, tels que les méthodes de Schwarz, semblent prometteuses pour la résolution du problème de couplage [92]. Néanmoins, elles peuvent induire une augmentation conséquente des coûts et temps de calculs, si la convergence de la méthode est lente. Pour ce type de modèle couplé, plusieurs variantes de l'algorithme du 4D-Var incrémental ont été proposées [114, 113], exploitant le cadre des méthodes de Gauss-Newton perturbées [68], pour introduire des simplifications du modèle couplé dans la résolution du sous-problème de minimisation quadratique. Il me semble intéressant de poursuivre ces travaux, et notamment la formulation du problème d'optimisation à résoudre, en vue de pouvoir construire des algorithmes réalisant à la fois le couplage et l'étape d'analyse, en un seul processus itératif et non pas deux imbriqués, ce qui réduirait leur complexité calcul.

# Annexes

# Annexe A

# Curriculum Vitae

## A.1 Etudes et expériences professionnelles

### A.1.1 Parcours universitaire

**10-2004 - 11-2007** Doctorat en Mathématiques Appliquées
Université Joseph Fourier - Grenoble I
Directeurs de thèse : E. Blayo et L. Debreu.

**09-2003 - 09-2004** Master 2 Recherche en Mathématiques Appliquées
Université Joseph Fourier - Grenoble I
Responsables de stage : E. Blayo et C. Robert

**09-2001 - 09-2004** Diplôme ingénieur ENSIMAG.

**09-1998 - 06-2001** CPGE, Lycée Clémenceau, Nantes.

### A.1.2 Déroulement de carrière

**Depuis 09-2017** Maître de conférence, section 26
Toulouse INP - ENSEEIHT - Département Sciences du Numérique
Institut de Recherche en Informatique de Toulouse (IRIT) - UMR CNRS 5505
Equipe Algorithmes Parallèles et Optimisation (APO).

**09-2013 - 08-2017** Maître de conférence, section 26
Toulouse INP - ENSEEIHT - Département Informatique et Mathématiques Appliquées
Institut de Recherche en Informatique de Toulouse (IRIT) - UMR CNRS 5505
Equipe Algorithmes Parallèles et Optimisation (APO).

**02-2011 - 07-2013** Chercheur sur projets
Nansen Environmental and Remote Sensing Center, Bergen, Norvège
Mohn-Sverdrup Center for Global Ocean Studies and Operational Oceanography

**02-2008 - 01-2011** Post-doctorant
Nansen Environmental and Remote Sensing Center, Bergen, Norvège
Mohn-Sverdrup Center for Global Ocean Studies and Operational Oceanography
Responsabilité : L. Bertino.

**10-2004 - 11-2007** Doctorant en Mathématiques Appliquées de l'Université Joseph Fourier - Grenoble I
Laboratoire Jean Kuntzmann (LJK) - UMR CNRS 5224
Equipe-projet INRIA MOISE
Directeurs de thèse : E. Blayo et L. Debreu.

## A.2 Liste des communications

Mes travaux de recherche portent sur les développements méthodologiques en assimilation de données, optimisation, algèbre numérique, et leurs applications aux fluides géophysiques. Ceci m'a amené à suivre différentes règles relatives à l'ordonnancement des auteurs selon les collaborations : l'ordre alphabétique propre aux revues de mathématiques appliquées, et un ordre dit significatif pour les revues orientés vers les géosciences. Dans ce dernier, les auteurs sont classés par ordre d'implication dans la conduite des travaux, avec usage de l'ordre alphabétique si implication similaire. Afin d'indiquer ces différences de pratique, les publications suivant l'ordre alphabétique seront référencées par [a ? ?], tandis que les autres le seront par [s ? ?]. Enfin, le soulignement d'un nom indiquera un co-encadrement de ma part - stages de Master, doctorants ou post-doctorant - en lien avec la publication.

### A.2.1 Revues internationales

*Articles soumis :*

**[a ? ?]** A. Bernigaud, S. Gratton, **E. Simon** : A nonlinear conjugate gradient algorithm for $p$-norm regularized least squares with application to variational data assimilation.

**[s ? ?]** T.H. Nguyen, S. Ricci, A. Piacentini, **E. Simon**, R. Rodriguez-Suquet, S. Peña-Luque : Gaussian anamorphosis for ensemble Kalman filter analysis of SAR-derived wet surface ratio observations.

**[s ? ?]** A. Rouvière, L. Pascal, F. Méry, **E. Simon**, S. Gratton : Neural prediction model for transition onset of a boundary layer in presence of 2D surface defect.

*Articles acceptés ou publiés :*

**[a18]** S. Gratton, **E. Simon**, D. Titley-Peloquin : Covariance matrix estimation for ensemble-based Kalman filters with multiple ensembles, *Mathematical Geosciences*, accepté.

**[a17]** S. Gratton, **E. Simon**, D. Titley-Peloquin, Ph. L. Toint : A note on inexact inner products in GMRES, *SIAM Journal on Matrix Analysis and Applications*, 43 (3), 1406-1422, 2022.

**[a16]** A. Bernigaud, S. Gratton, F. Lenti, **E. Simon**, O. Sohab : $L_p$-norm regularization in variational data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 147, 2067-2081, 2021 ;

**[a15]** S. Gratton, **E. Simon**, Ph. L. Toint : An algorithm for the minimization of nonsmooth and nonconvex functions using inexact evaluations and its worst-case complexity, *Mathematical Programming*, 187, 1-24, 2021.

**[a14]** S. Gratton, **E. Simon**, D. Titley-Peloquin, Ph. L. Toint : Minimizing convex quadratics with variable precision conjugate gradients, *Numerical Linear Algebra with Applications*, 28 :e2337, 2021.

**[s13]** R. Benshila, G. Thoumyre, M. Al Najar, G. Abessolo, R. Almar, E. Bergsma, G. Hugonnard, L. Labracherie, B. Lavie, T. Ragonneau, **E. Simon**, B. Vieuble, D. Wilson : A deep learning approach for estimation of the nearshore bathymetry, *Journal of Coastal Research*, 95(sp1), 1011-1015, 2020.

**[a12]** S. Gratton, S. Gürol, **E. Simon**, Ph. L. Toint : Guaranteeing the convergence of the saddle formulation for weakly-constrained 4D-VAR data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 144, 2592-2602, 2018.

**[a11]** S. Gratton, S. Gürol, **E. Simon**, Ph. L. Toint : A note on preconditioning weighted

linear squares, with consequences for weakly-constrained variational data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 144, 934-940, 2018.

[**s10**] M. E. Gharamti, A. Samuelsen, L. Bertino, **E. Simon**, A. Korosov, U. Daewel : Online Tuning of Ocean Biogeochemical Parameters using Ensemble Estimation Techniques : Application to a one-dimensional Model in the North Atlantic, *Journal of Marine Systems*, 168, 1-16, 2017.

[**s9**] L. Debreu, E. Neveu, **E. Simon**, F.-X. Le Dimet, A. Vidard : Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems, *Quarterly Journal of the Royal Meteorological Society*, 142, 515-528, 2016.

[**s8**] **E. Simon**, A. Samuelsen, L. Bertino, S. Mouysset : Experiences in multiyear combined state-parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using the Ensemble Kalman Filter, *Journal of Marine Systems*, 152, 1-17, 2015.

[**a7**] S. Gratton, M. Rincon-Camacho, **E. Simon**, Ph. L. Toint : Observation Thinning in Data Assimilation Computations. *EURO Journal of Computational Optimization*, 3, 31-51, 2015.

[**s6**] M. Gehlen, R. Barciela, L. Bertino, P. Brasseur, M. Butenschön, F. Chai, A. Crise, Y. Drillet, D. Ford, D. Lavoie, P. Lehodey, C. Perruche, A. Samuelsen, **E. Simon** : Building the capacity for forecasting marine biogeochemistry and ecosystems : recent advances and future developments, *Journal of Operational Oceanography*, 8(S1), s168-s187, 2015.

[**s5**] J. S. Pelc, **E. Simon**, L. Bertino, G. El Serafy, A. Heemink : Application of model reduced 4D-Var to a 1D ecosystem model, *Ocean Modelling*, 57-58, 43-58, 2012.

[**s4**] **E. Simon**, A. Samuelsen, L. Bertino, D. Dumont : Estimation of positive sum-to-one constrained zooplankton grazing preferences with the DEnKF : a twin experiment, *Ocean Science*, 8, 587-602, 2012.

[**s3**] **E. Simon**, L. Bertino : Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation : application to a 1D ocean ecosystem model, *Journal of Marine Systems*, 89, 1-18, 2012.

[**s2**] **E. Simon**, L. Debreu, E. Blayo : 4D Variational Data Assimilation for Locally Nested Models : complementary theoretical aspects and application to a 2D shallow water model, *International Journal for Numerical Methods in Fluids*, 66, 135-161, 2011.

[**s1**] **E. Simon**, L. Bertino : Application of the Gaussian anamorphosis to assimilation in a 3D coupled physical-ecosystem model of the North Atlantic with the EnKF : a twin experiment, *Ocean Science*, 5, 495-510, 2009.

### A.2.2 Actes de conférences, chapitres de livres, vulgarisation scientifique

*Conférences internationales avec actes :*

[**s4**] S. Mauran, S. Mouysset, **E. Simon**, L. Bertino : A kernel extension of the Ensemble Transform Kalman Filter, *International Conference on Computational Science (ICCS 2023)*, Prague, 2023.

[**s3**] A. Rouvière, L. Pascal, F. Méry, **E. Simon**, S. Gratton : Neural prediction model for transition onset of a boundary layer in presence of 2D surface defect, *AIAA SCITECH forum*, San Diego, 2022.

[**a2**] S. Gratton, S. Gürol, **E. Simon**, Ph. L. Toint : Issues in making the weakly-constrained 4DVar formulation computationally efficient, *Oberwolfach Reports*, 13(4), 2726-2730, Oberwolfach, 2016.

[**s1**] **E. Simon**, L. Bertino : Joint state-parameter estimation in a 3D coupled physical-

ecosystem model of the North Atlantic : assimilation of SeaWiFS data with a non-Gaussian extension of an ESRF, *ESA Living Planet Symposium*, Bergen, 2010.

*Conférences nationales avec actes :*

[**a2**] S. Gratton, S. Gürol, **E. Simon**, Ph. L. Toint : Les algorithmes et la puissance de calcul dans les techniques de prévision pour les géosciences en grande dimension vus sous l'angle de l'optimisation mathématique, *Colloque Modélisation : succès et limites*, CNRS & Académie des Technologies, Paris, 37-47, 2018.

[**s1**] **E. Simon**, E. Blayo, L. Debreu : Assimilation variationnelle de données pour des modèles emboîtés, *Colloque National sur l'Assimilation de Données*, Toulouse, 2006.

*Chapitre d'ouvrage :*

[**s1**] L. Debreu, E. Neveu, **E. Simon**, F.-X. Le Dimet : Multigrid algorithms and local mesh refinement methods in the context of variational data assimilation, *In : E. Blayo et al. (eds) : Advanced Data Assimilation for Geosciences. Lecture notes of Les Houches summer school 2012*, Oxford University Press, 395-412, 2014.

*Article de vulgarisation scientifique :*

[**s1**] L. Bertino, A. Samuelsen, F. Counillon, **E. Simon**, P. Sakov : Advanced data assimilation in oceanography, *META (ISSN 1890-1956)*, 8-12, 2012.

### A.2.3 Développements technologiques

1. *Réanalyse de la biogéochimie marine de l'océan Arctique pour la période 2007-2010.* Dans le cadre des projets européens MyOcean et MyOcean2 relatifs à la surveillance, l'analyse et la prédiction des océans, j'ai été amené à produire des données de réanalyse (analyse historique d'un phénomène océanique ou climatique) de la biogéochimie de l'océan Arctique sur la période 2007-2010. Ces données contiennent les moyennes mensuelles de différentes variables et paramètres biogéochimiques du modèle couplé océan-écosystème HYCOM-NORWECOM et furent disponibles en accès libre sur le portail web de MyOcean. Le NERSC (employeur jusqu'en juillet 2013) est responsable du centre de prévision pour l'océan Arctique ("Arctic Marine Forecasting Center"). La production de ces données correspond à environ 200 cycles d'assimilation par un filtre de Kalman d'ensemble - ensemble de taille 100 - et a duré un an, nonobstant le temps associé au développement du système d'assimilation.

   `http://marine.copernicus.eu`

2. *GOTM-NORWECOM.* J'ai développé un système d'assimilation de données, par filtrage de Kalman d'ensemble avec anamorphose gaussienne, générique aux différents modèles d'écosystèmes marins 1D (colonne d'eau) embarqués dans le modèle GOTM (General Ocean Turbulence Model). Il correspond à différents modules Fortran90 implémentés dans GOTM, ainsi qu'à quelques modifications du code de base (introduction de l'assimilation de données, parallélisation MPI). Ce système se veut représentatif d'une colonne d'eau (un point de grille) du système TOPAZ-ECO et a pour objectif la validation des développements méthodologiques avant transfert dans le système "opérationnel". Ce système n'est pas distribué au delà de ses développeurs, à savoir Annette Samuelsen

(implémentation du modèle NORWECOM), Dany Dumont (GOTM) et moi-même, ainsi que des instituts impliqués (NERSC, ISMER, et maintenant INPT).

3. *TOPAZ-ECO.* J'ai également été amnené à développer la composante biogéochimique des systèmes d'assimilation TOPAZ3, puis TOPAZ4 dans le cadre de divers projets européens. Ces systèmes correspondent à différents programmes Fortran90 (EnKF, module d'anamorphose), différentes modifications des modéles HYCOM et NORWECOM dans l'optique d'une estimation de paramétres, et différents scripts shell. Le système TOPAZ4-ECO a été utilisé pour produire les données de réanalyse de la biogéochimie marine de l'océan Arctique pour la période 2007-2010. Ce système n'avait pas vocation à être distribué en dehors du NERSC et de son partenaire Met.no.
4. *MICOM-HAMOCC (NorESM).* J'ai développé un système d'assimilation par méthodes d'ensemble pour la composante biogéochimique océanique (MICOM-HAMOCC) du Norwegian Earth System Model (NorESM). Ce système correspond à différents programmes Fortran90 (EnKF, module d'anamorphose), différentes modifications des modéles MICOM et HAMOCC dans l'optique d'une estimation de paramétres, et différents scripts shell. Ce système n'a pas vocation à être distribué au delà de la communauté NorESM.

### A.2.4 Communications orales et posters

Les conférences mentionnées n'éditent pas de proceedings associés aux présentations, sauf exceptions. Le processus de sélection se base généralement sur des résumés d'une page (format A4). Ne sont pas listées les conférences pour lesquelles j'étais co-auteur, mais ne présentais pas les travaux. Les présentations associées à l'encadrement de doctorants sont spécifiées dans la section A.3. Le terme "présentation sollicitée" traduit une invitation à présenter mes travaux en ouverture d'une session de conférence (avec appel ouvert à soumission d'abstracts), sans pour autant financer ma venue et participation à la dite conférence.

1. S. Gratton, **E. Simon**, D. Titley-Peloquin, Ph. L. Toint : Computing Inexact Inner Products in GMRES, *Preconditioning 2022*, 8-10 juin 2022, Chemnitz, Allemagne. [Oral]
2. S. Gratton, **E. Simon**, D. Titley-Peloquin, Ph. L. Toint : Computing Inexact Inner Products in GMRES, *SIAM Conference on Applied Linear Algebra*, 17-21 mai 2021, Nouvelle-Orléans, USA. [Oral, virtuel]
3. S. Gratton, **E. Simon**, D. Titley-Peloquin : Towards the estimation of forecast error covariance matrices in multi-ensemble data assimilation, *4th OceanPredict Data Assimilation Task Team Meeting*, 20-23 janvier 2020, Toulouse, France. [Oral]
4. S. Gratton, **E. Simon**, Ph. L. Toint : Minimizing convex quadratics with variable precision Krylov methods, *ICIAM 2019*, 14-19 juillet 2019, Valence, Espagne. [Oral]
5. S. Gratton, **E. Simon**, Ph. L. Toint : Minimizing convex quadratics with variable precision Krylov methods, *Optimization Days 2019*, 13-15 mai 2019, Montréal, Canada. [Oral]
6. S. Gratton, S. Gürol, **E. Simon**, Ph. L. Toint : On the use of the saddle formulation in weakly-constrained 4D-Var, *Workshop on sensitivity analysis and data assimilation in meteorology and Oceanography 2018*, 1-6 juillet 2018, Aveiro, Portugal. [Oral]
7. S. Gratton, S. Gürol, **E. Simon**, Ph. L. Toint : On the use of the saddle formulation in weakly-constrained 4D-Var, *EGU General Assembly 2018*, 8-13 avril 2018, Vienne, Autriche. [Oral, Présentation sollicitée].

8. S. Gratton, **E. Simon**, D. Titley-Peloquin : Towards the estimation of forecast error covariance matrices in multi-ensemble data assimilation, *7th WMO symposium on data assimilation*, 11-15 septembre 2017, Florianopolis, Brésil. [Poster]
9. S. Gratton, **E. Simon**, D. Titley-Peloquin : Estimating optimal covariance matrices in multi-ensemble data assimilation : towards preconditioning EnVar algorithms, *20th ILAS Conference 2016*, 11-15 juillet 2016, Louvain, Belgique. [Oral]
10. S. Gratton, M. Rincon-Camacho, **E. Simon**, Ph. L. Toint : Observation thinning in data assimilation computations, *Dynamics Days 2015*, 6-10 septembre 2015, Exeter, UK. [Oral, Présentation sollicitée].
11. S. Gratton, M. Rincon-Camacho, **E. Simon**, Ph. L. Toint : Dual space multigrid strategies for variational data assimilation, *Workshop on Sensitivity Analysis and Data Assimilation in Meteorology and Oceanography*, 1-5 juin 2015, Roanoke, USA. [Oral]
12. S. Gratton, M. Rincon-Camacho, **E. Simon**, Ph. L. Toint : Dual space multigrid strategies for variational data assimilation, *Optimization 2014 Conference*, Guimaraes, Portugal, 2014. [Oral]
13. **E. Simon**, A. Samuelsen, L. Bertino : Towards the multiyear combined state parameter estimation with ensemble-based Kalman filters in ocean ecosystem models : a 2007-2010 hindcast of the North Atlantic and Arctic biology. *EGU General Assembly*, Vienne, Autriche, 2013. [Poster]
14. **E. Simon**, A. Samuelsen, L. Bertino, D. Dumont : Estimation of positive sum-to-one constrained parameters with ensemble-based Kalman filters : Application to an ocean ecosystem model, *International Conference on Ensemble Methods in Geophysical Sciences*, Toulouse, France, 2013. [Oral]
15. **E. Simon**, A. Samuelsen, L. Bertino : Towards a reanalysis of the North Atlantic and Arctic Ocean Biology : A multi-year assimilation of satellite ocean color data with the deterministic ensemble Kalman filter, *2012 ICES Annual Science Conference*, Bergen, Norvège, 2012. [Poster]
16. **E. Simon**, A. Samuelsen, L. Bertino, D. Dumont : Estimation of positive sum-to-one constrained zooplankton grazing preferences with the DEnKF : a twin experiment, *7th International EnKF Workhsop*, Os, Norvège, 2012. [Oral]
17. **E. Simon**, A. Samuelsen, L. Bertino, D. Dumont : Estimation of positive sum-to-one constrained zooplankton grazing preferences with the DEnKF : a twin experiment, *ART-APECS Workshop*, Sopot, Pologne, 2012. [Poster]
18. **E. Simon**, L. Bertino, A. Samuelsen : Estimation of sum-to-one constraint parameters with non-Gaussian extensions of ensemble Kalman filters : application to a 1D ocean biogeochemical model. *AGU Fall meeting 2011*, San Francisco, USA, 5-9 décembre 2011. [Poster]
19. **E. Simon**, L. Bertino : Joint state-parameter estimation in a 3D coupled physical-ecosystem model of the North Atlantic : assimilation of SeaWiFS data with a non-Gaussian extension of an ESRF, *ESA Living Planet Symposium*, Bergen, 28 juin - 2 juillet 2010. [Oral]
20. **E. Simon**, L. Bertino : Non-Gaussianity and biased parameter estimation of an ocean biological system with the EnKF. $5^{th}$ *international workshop on Ensemble Kalman filter for model updating*, Bergen, Norvège, 18-20 mai 2010. [Oral, donné par L. Bertino]

21. **E. Simon**, L. Bertino : Application of the Gaussian anamorphosis to assimilation in a 3D coupled physical-ecosystem model of the North Atlantic with the EnKF : a twin experiment. *EGU General Assembly*, Vienne, Autriche, 19-24 avril 2009. [ Poster]
22. **E. Simon**, L. Debreu, E. Blayo : 4D-Variational Data Assimilation for Locally Nested Numerical Ocean Models. *EGU General Assembly*, Vienne, Autriche, 19-24 avril 2009. [ Poster]
23. **E. Simon**, L. Bertino : EnKF with Gaussian Anamorphosis for a 3D Ocean Ecosystem Model. Inital Results. *GODAE Final Symposium*, Nice, France, 12-15 novembre 2008. [Poster]
24. **E. Simon**, L. Debreu, E. Blayo : 4D-Variational Data Assimilation for Locally Nested Numerical Ocean Models. *GODAE Final Symposium*, Nice, 12-15 novembre 2008. [ Poster]
25. **E. Simon**, L. Debreu, E. Blayo : Assimilation variationnelle de données pour des modèles localement emboîtés, *Congrès National de Mathématiques Appliquées et Industrielles*, Praz sur Arly, France, 4-8 juin 2007. [Oral]
26. **E. Simon**, L. Debreu, E. Blayo : Assimilation variationnelle de données pour des modèles emboîtés, *Colloque National sur l'Assimilation de Données*, Toulouse, France, 9-10 mai 2006. [Poster]
27. **E. Simon**, L. Debreu, E. Blayo : Variational data assimilation for locally nested models, *EGU General Assembly*, Vienne, Autriche, 2-7 avril 2006. [Oral]
28. **E. Simon**, L. Debreu, E. Blayo, Y. de Vismes : Variational data assimilation for locally nested models, *4th WMO International Symposium - Assimilation of Observations in Meteorology and Oceanography*, Prague, République Tchèque, 18-22 avril 2005. [Poster]

## A.3 Encadrements d'étudiants et chercheurs

Les encadrements mentionnés ont été réalisés dans le cadre de projets, de bourses MESR, contrat CIFRE, visite internationale pour les doctorants et post-doctorant mentionnés. La liste des communications associées à mes encadrements utilise la nomenclature suivante : [a ?] pour article dans une revue internationale, [ca ?] pour conférence internationale avec actes, et [c ?] pour conférence nationale ou internationale sans acte, [mt] pour manuscrit de thèse.

### A.3.1 Post-doctorant

— **Flavia Lenti** : co-encadrement (50%) avec Serge Gratton (INPT, IRIT) de février 2015 à juillet 2016. Les travaux de Flavia ont été réalisés dans le cadre du projet AVENUE, financé par la la fondation STAE (Toulouse). Ses travaux ont visé en partie à explorer l'introduction de la norme $\|\|_p$, avec $p > 1$, en assimilation variationnelle de données. A la suite de son post-doctorat, Flavia a été recrutée comme Ingénieure consultante par CLC space auprès d'EUMETSAT (Darmstadt).

j1- A. Bernigaud, S. Gratton, F. Lenti, E. Simon, O. Sohab : $L_p$-norm regularization in variational data assimilation, *Q. J. R. Meteorol. Soc.*, 147, 2067-2081, 2021.

### A.3.2 Doctorants

*Thèses soutenues :*

1. **Adrien Rouvière** : co-encadrement (15%) de la thèse avec Serge Gratton (INPT, IRIT), Fabien Mery (ONERA) et Pascal Lucas (ONERA), d'octobre 2019 jusqu'au 4 avril 2023, date de soutenance. Cette thèse fut financée par l'ONERA et le programme CleanSky2.

   [**a ?**] A. Rouvière, L. Pascal, F. Méry, E. Simon, S. Gratton : Neural prediction model for transition onset of a boundary layer in presence of 2D surface defect, soumis.

   [**mt**] A. Rouvière : Amélioration des modèles de tolérance de surface pour les couches limites en s'appuyant sur des outils d'intelligence artificielle, *Doctorat de l'Université de Toulouse*, délivré par l'Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), 4/04/2023 ;

   [**ca2**] A. Rouviere, L. Pascal, F. Méry, E. Simon, S. Gratton : Neural prediction model for transition onset of a boundary layer in presence of 2D surface defect, *22th ONERA-DLR Aerospace Symposium*, Hambourg, Allemagne, 1-3 juin 2022. [Oral]

   [**ca1**] A. Rouviere, L. Pascal, F. Méry, E. Simon, S. Gratton : Neural prediction model for transition onset of a boundary layer in presence of 2D surface defect, *AIAA SCITECH forum*, San Diego, USA, 3-7 janvier 2022. [Oral, virtuel]

2. **Antoine Bernigaud** : co-encadrement (70%) de la thèse avec Serge Gratton (INPT, IRIT), d'octobre 2019 jusqu'au 16 décembre 2022, date de soutenance. Cette thèse fut financée par une bourse MESR. Depuis début 2023, Antoine Bernigaud est post-doctorant au NERSC à Bergen (Norvège).

   [**a ?**] A. Bernigaud, S. Gratton, E. Simon : A nonlinear conjugate gradient algorithm for $p$-norm regularized least squares with application to variational data assimilation, soumis.

   [**mt**] A. Bernigaud : Introduction de la régularisation en norme $L_p$, avec $1 < p < 2$, pour la prise en compte de la parcimonie en assimilation de données, *Doctorat de l'Université de Toulouse*, délivré par Toulouse INP, 16/12/2022 ;

   [**c4**] A. Bernigaud, S. Gratton, E. Simon : Régularisation en norme $p$ en assimilation de données avec $1 < p < 2$, bénéfices et minimisation de l'algorithme du 4DVar avec pénalisation, *Journées SMAI-MODE 2022*, Limoges, France, 30 mai - 3 juin 2022 ; [Oral]

   [**c3**] A. Bernigaud, S. Gratton, E. Simon : $L_p$-norm regularization in variational data assimilation with $1 < p < 2$, benefits and minimization of the penalized 4DVar, *8th International Symposium on Data Assimilation (ISDA)*, Fort Collins, USA, 6-10 juin 2022 ; [Oral]

   [**c2**] A. Bernigaud, S. Gratton, E. Simon : Utilisation de la norme $p$ en assimilation de données et méthode de minimisation de la fonctionelle sous-jacente, *Congrès des jeunes chercheuses et chercheurs en mathématiques appliquées (CJC-MA)*, Palaiseau, France, 27-29 octobre 2021 ; [Poster]

   [**a1**] A. Bernigaud, S. Gratton, F. Lenti, E. Simon, O. Sohab : $L_p$-norm regularization in variational data assimilation, *Q. J. R. Meteorol. Soc.*, 147, 2067-2081, 2021 ;

   [**c1**] A. Bernigaud, S. Gratton, F. Lenti, E. Simon, O. Sohab : $p$-norm regularization in variational data assimilation, *EGU General Assembly*, Vienne, Autriche, 4-9 mai 2020. [Oral, virtuel]

3. **Dimitri Mottet** : co-encadrement (40%) de la thèse avec Serge Gratton (INPT, IRIT) et

Jean-Philippe Argaud (EDF), de novembre 2017 au 12 janvier 2021, date de sa soutenance. Cette thèse fut financée par un contrat CIFRE (EDF R&D). A la suite de sa thèse, Dimitri a été recruté comme Ingénieur de Recherche par EDF R&D.

[**mt**] D. Mottet : Raffinement adaptatif du processus d'assimilation de données par méthodes de Kalman d'ensemble pour des problèmes non-linéaires, *Doctorat de l'Université de Toulouse*, délivré par Toulouse INP, 12/01/2021 ;

[**c3**] J.-F. Argaud, S. Gratton, D. Mottet, E. Simon : Interaction between ensemble filters and model's dynamics to improve forecast for nuclear reactor monitoring, *14th EnKF Workshop*, Voss, Norvège, 3-6 juin 2019 ; [Oral]

[**c2**] J.-F. Argaud, S. Gratton, D. Mottet, E. Simon : Interaction between Ensemble filter/smoother and model dynamics for stiff ODEs, *Colloque National d'Assimilation de Données*, Rennes, France, 26-28 septembre 2018 ; [Poster]

[**c1**] J.-F. Argaud, S. Gratton, D. Mottet, E. Simon : Interaction between ensemble filter/smoother and model dynamics for stiff ODEs, *11th Workshop on Sensitivity Analysis and Data Assimilation in Meteorology and Oceanography*, Aveiro, Portugal, 1-6 juillet 2018. [Poster]

4. **Joanna Pelc** : j'ai co-supervisé la visite scientifique de Joanna Pelc au NERSC avec Laurent Bertino (NERSC) à l'automne 2010 (2-3 mois). Joanna était alors doctorante à TU Delft (Pays-bas) et encadrée par Ghada El Serafy et Arnold Hemmink.

   [**a1**] J.S. Pelc, E. Simon, L. Bertino, G. El Serafy, A. Heemink : Application of model reduced 4D-Var to a 1D ecosystem model, *Ocean Model.*, 57-58, 43-58, 2012 ;

*Thèses en cours :*

1. **Sophie Mauran** : direction de la thèse (50%) avec Sandrine Mouysset (UT3, IRIT) et Laurent Bertino (NERSC, Bergen, Norvège) depuis octobre 2021. Cette thèse est financée par une bourse MESR.

   [**ca1**] S. Mauran, S. Mouysset, E. Simon, L. Bertino : A kernel extension of the Ensemble Transform Kalman Filter, *International Conference on Computational Science (ICCS 2023)*, Prague, 2023.

   [**c2**] S. Mauran, S. Mouysset, E. Simon, L. Bertino : A kernel extension of the Ensemble Transform Kalman Filter, *18th international EnKF workshop*, Norheimsund, Norvège, 2-5 mai 2023. [Oral]

   [**c1**] S. Mauran, E. Simon, S. Mouysset, L. Bertino : Introduction of kernel methods in data assimilation, *Sparse days in Saint-Girons IV*, Saint-Girons, France, 20-22 juin 2022. [Poster]

2. **Jérémy Briant** : co-encadrement (20%) de la thèse avec Serge Gratton (INPT, IRIT), Paul Mycek (Cerfacs), Selime Gürol (Cerfacs) et Anthony Weaver (Cerfacs), depuis octobre 2020. Cette thèse est financée par l'appel à projets 80|Prime 2020 du CNRS.

   [**c2**] J. Briant, M. Destouches, S. Gratton, S. Gürol, P. Mycek, E. Simon, A. Weaver : Spectral analysis of multivariate multilevel Monte Carlo methods, *15th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC)*, Linz, Autriche, 17-22 juillet 2022 ; [Oral]

   [**c1**] J. Briant, M. Destouches, S. Gratton, S. Gürol, P. Mycek, E. Simon, A. Weaver : Background error covariance matrix estimation from multifidelity ensembles, *International Symposium on Data Assimilation - Online*, 2 juillet 2021. [Oral, virtuel]

### A.3.3 Stage de Master 2

— **Sophie Mauran** : co-encadrement (40%) avec Sandrine Mouysset (UPS, IRIT) et Serge Gratton (INPT, IRIT) d'un stage de 3ème année ENSEEIHT, de mars à mi-septembre 2021. Ce stage porte sur l'introduction de méthodes à noyaux pour l'assimilation de données ensembliste ;
— **Antoine Bernigaud** : co-encadrement (50%) avec Serge Gratton (INPT, IRIT) d'un stage de 3ème année ENSEEIHT, de mars à mi-septembre 2019. Ce stage portait sur l'introduction de stratégie de régularisation parcimonieuse de l'algorithme 4D-Var en norme $p$, et faisait suite aux travaux de F. Lenti et O. Sohab.

#### A.3.3.1 Stage de Master 1

— **Oumaima Sohab** : co-encadrement (50%) avec Serge Gratton (INPT, IRIT) d'un stage de 2ème année ENSEEIHT, de fin juin à mi-septembre 2018. Ce stage portait sur l'introduction de stratégie de régularisation parcimonieuse de l'algorithme 4D-Var en norme $p$ ;
— **Kéran Asseko Nguema** : co-encadrement (25%) avec Serge Gratton (INPT, IRIT), Philippe L. Toint (Université de Namur) et Selime Gürol (Cerfacs) d'un stage de 2ème année ENSEEIHT, de fin juin à mi-septembre 2017. Ce stage portait sur l'étude d'une méthode itérative pour la résolution de système de type point selle ;
— **Auguste Caen** : co-encadrement (50%) avec Serge Gratton (INPT, IRIT) d'un stage de 2ème année ENSEEIHT, de fin juin à mi-septembre 2014. Ce stage portait sur l'étude de quelques problèmes en data mining.

## A.4 Coordination de projets

— P.I du projet *High-performance ensemble-variational data assimilation using multi-fidelity ensembles for Earth system modelling applications*, 2020-2022.
Ce projet est financé par le CNRS via son programme 80|PRIME. Il porte sur des développements algorithmiques pour l'assimilation de données multi-fidélité (échelles, précision,..). Ce projet est conjoint avec des membres du CECI (Cerfacs, Toulouse) et finance la bourse de thèse de Jérémy Briant, ainsi que les missions et le petit matériel associés.

— P.I. du projet *Introduction de la norme p pour la prise en compte de propriétés de parcimonie en assimilation variationnelle de donnés*, 2019-2020.
Ce projet fut financé par l'INSU sur l'appel à projet Les Enveloppes Fluides et l'Environnement - Méthodes mathématiques et numériques (LEFE/MANU). Il avait pour but de financer les missions et le petit matériel en lien avec la thèse d'Antoine Bernigaud.

— Co P.I, puis P.I. du projet *Assimilation Variationnelle-ENsemble UnifiéE (AVENUE)*, 2014-2017.
Ce projet fut financé par la Fondation de Coopération Scientifique Sciences et Technologies pour l'Aéronautique et l'Espace (Toulouse). Il portait sur l'étude d'approches unifiées de lissage hybride pour l'assimilation de données dans les applications majeures des géosciences que constituent les systèmes océaniques, atmosphériques et gravimétriques.

Ce projet a financé notamment quatre contrats de post-doctorat de 18 mois, répartis entre l'IRIT (Flavia Lenti), le Cerfacs, Météo-France et l'OMP-CNES.

— Responsable adjoint pour le Nansen Environmental and Remote Sensing Center (NERSC) au sein du projet *GeoCarbon* (période 2011-2013).
Ce projet, financé dans le cadre du programme européen FP7, portait sur des thématiques Climat et proposait la définition et l'implémentation de systèmes d'observations et d'analyse - notamment par assimilation de données - du cycle du carbone au niveau terrestre global. La tache principale du NERSC portait sur l'étude des flux de carbone entre l'océan et l'atmosphère, via le développement d'un système d'assimilation de données par méthodes d'ensemble pour la composante biogéochimique océanique de NorESM et la production de simulations d'ensemble sur une décennie.

— Responsable adjoint pour le Nansen Environmental and Remote Sensing Center (NERSC) au sein du projet *Ocean Colour - Climate Change Initiative (OC-CCI)* (période 2012-2013).
Ce projet, financé par l'Agence Spatiale Européenne (ESA), portait sur la production d'une série temporelle, d'une duré de 15 ans, d'observations de couleur de l'eau (incluant notamment des données de concentration de chlorophylle de surface) depuis les données satellitaires récoltées par les instruments MERIS, SeaWiFS et MODIS. Le NERSC, et moi-même, utilisateurs de ce type d'observations, avons joué un rôle d'expertise en assimilation de données auprès des équipes de recherche en charge de la production et validation de cette série temporelle, ainsi qu'un role de beta testeurs (via le système GOTM-NORWECOM).

## A.5 Responsabilités collectives

### A.5.1 Enseignement

— Responsable de la deuxième année du parcours HPC & Big Data du département Sciences du Numérique de l'ENSEEIHT depuis septembre 2020 (niveau M1). Ceci m'amène notamment à être membre du Conseil de Recherche et Formation de ce même département, conseil où est discuté la politique d'affectation des postes (priorisation des profils de postes) du département ;

— Co-responsable du Certificat Sciences des données et Big Data, organisé par Toulouse Tech, regroupant les écoles d'ingénieurs du site Toulousain, de septembre 2019 à juin 2022. L'objectif de ce certificat est la sensibilisation aux sciences des données pour les étudiants de M1/M2 des établissements membres, n'ayant pas choisi ces thématiques comme spécialité. Cette formation a lieu sur 7 mois, pour un total de 90 heures.

### A.5.2 Recherche

— Co-responsable de l'équipe Algorithmes Parallèles et Optimisation de l'IRIT depuis mai 2022. De nouveau, ceci m'amène à être membre du Conseil de Recherche et Formation du département Sciences du Numérique de l'ENSEEIHT ;

— Membre du conseil scientifique de l'action LEFE GMMC de l'INSU [1], dédié à l'océanographie opérationnelle, depuis janvier 2022. Ce mandat est d'une durée de quatre ans, et j'occupe actuellement la fonction de vice-président. Les missions consistent notamment à préparer l'appel à projets, qui a lieu annuellement, à évaluer les projets soumis, à préparer les journées GMMC, ainsi qu'à répondre aux différentes sollicitations du CS en lien avec l'océanographie opérationnelle française (par exemple, la prospective INSU en 2022).

1. https://programmes.insu.cnrs.fr/lefe/cs_actions/gmmc/

# Bibliographie


[1] S. Anzengruber and R. Ramlau. Morozov's discrepancy principle for Tikhonov-type functionals with non-linear operators. *Johann Radon Institue for Computational and Applied Mathematics Austrian Academe of Sciences*, RICAM-Report, 2009-2013.

[2] S.W. Anzengruber. *The discrepancy principle for Tikhonov regularization in Banach spaces*. PhD thesis, Johannes Kepler Universitat Linz, 2011.

[3] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68 :337–340, 1950.

[4] M. Asch, M. Bocquet, and M. Nodet. *Data Assimilation - Methods, Algorithms and Applications*. SIAM, 2016.

[5] F. Bach. Sharp analysis of low-rank kernel matrix approximations. *Journal of Machine Learning Research*, 30 :1–25, 2013.

[6] S. Bellavia, G. Gurioli, B. Morini, and Ph.L. Toint. Adaptive regularization algorithms with inexact evaluations for nonconvex optimization. *SIAM Journal on Optimization*, 29(4) :2881–2915, 2020.

[7] R. Benshila, G. Thoumyre, M. Al Najar, G. Abessolo, R. Almar, E. Bergsma, G. Hugonnard, L. Labracherie, B. Lavie, T. Ragonneau, E. Simon, B. Vieuble, and D. Wilson. A deep learning approach for estimation of the nearshore bathymetry. *Journal of Coastal Research*, 95(sp1) :1011–1015, 2020.

[8] A. Bernigaud, S. Gratton, F. Lenti, E. Simon, and O. Sohab. $l_p$-norm regularization in variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 147 :2067–2081, 2021.

[9] A. Bernigaud, S. Gratton, and E. Simon. A nonlinear conjugate gradient algorithm for $p$-norm regularized least squares with application to variational data assimilation. soumis.

[10] L. Bertino, G. Evensen, and H. Wackernagel. Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71 :223–242, 2003.

[11] C.H. Bishop, B.J. Etherton, and S.J. Majumdar. Adaptive sampling with the ensemble transform Kalman filter. Part I : theoretical aspects. *Monthly Weather Review*, 129 :420–436, 2001.

[12] A. Björck. Solving linear least squares problems by Gram-Schmidt orthogonalization. *BIT Numerical Mathematics*, 7 :1–21, 1967.

[13] E. Blayo and L. Debreu. Nesting ocean models. In Eric P. Chassignet and Jacques Verron, editors, *Ocean Weather Forecasting : An Integrated View of Oceanography*, pages 127–147. Springer, 2006.

[14] M. Bocquet. Localization and the iterative ensemble Kalman smoother. *Quartely Journal of the Royal Meteorological Society*, 142 :1075–1089, 2016.

[15] M. Bocquet, J. Brajard, A. Carrassi, and L. Bertino. Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization. *Foundations of Data Sciences*, 2 :55–80, 2020.

[16] M. Bocquet and P. Sakov. Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems. *Nonlinear Processes in Geophysics*, 19 :383–399, 2012.

[17] M. Bocquet and P. Sakov. An iterative ensemble Kalman smoother. *Quartely Journal of the Royal Meteorological Society*, 140 :1521–1535, 2014.

[18] T. Bonesky, K.S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster. Minimization of Tikhonov functional in Banach spaces. *Abstract and Applied Analysis*, 2008 :1–18, 2007.

[19] C. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Transactions on Image Processing*, 2 :296–310, 1993.

[20] J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations : a case study with the Lorenz 96 model. *Journal of Computational Science*, 44 :101171, 2020.

[21] J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino. Combining data assimilation and machine learning to infer unresolved scale parametrisation. *Philosophical Transactions of the Royal Society A*, 379 :20200086, 2021.

[22] A. Brandt. *Multigrid Techniques : Guide with applications to fluid dynamics.* GMD-Studien, 1984.

[23] J.-M. Brankart, C.-E Testut, D. Béal, M. Doron, C. Fontana, M. Meinvielle, P. Brasseur, and J. Verron. Towards an improved description of ocean uncertainties : effect of local anamorphic transformations on spatial correlations. *Ocean Science*, 8 :121–142, 2012.

[24] W.L. Briggs. *A Multigrid Tutorial.* SIAM, 1987.

[25] D.S. Broomhead and D. Lowe. *Radial basis functions, multi-variable functional interpolation and adaptive networks.* Royal Signals and Radar Establishment Malvern, 1988.

[26] D. Béal, P. Brasseur, J.-M. Brankart, Y. Ourmières, and J. Verron. Characterization of mixing errors in a coupled physical biogeochemical model of the north atlantic : implications for nonlinear estimation using Gaussian anamorphosis. *Ocean Science*, 6 :247–262, 2010.

[27] C. Cardinali, S. Pezzuli, and E. Anderson. Influence matrix diagnostic of a data assimilation system. *Quartely Journal of the Royal Meteorological Society*, 130(603) :2767–2786, 2004.

[28] Carla Cardinali. Observation influence diagnostic of a data assimilation system, 2013.

[29] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimiliation in the Geosciences : An overview of methods, issues and perspectives. *WIREs CLimate Change*, 9 :e535, 2018.

[30] B. Chapnik. *Réglage des statistiques d'erreur en assimilation variationnelle.* Phd thesis, Université Toulouse III, 2005.

[31] A. Chernov, H. Hoel, K. Law, F. Nobile, and R. Tempone. Multilevel ensemble Kalman filtering for spatio-temporal processes. *Numerische Mathematik*, 147 :71–125, 2021.

[32] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods.* MPS-SIAM Series on Optimization, SIAM, Philadelphia, USA, 2000.

[33] P. Courtier. Dual formulation of four-dimensional variational assimilation. *Quartely Journal of the Royal Meteorological Society*, 123 :2449–2461, 1997.

[34] P. Courtier, J. N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quartely Journal of the Royal Meteorological Society*, 120 :1367–1387, 1994.

[35] J.D. Crouch, V.S. Kosorygin, and M.I. Sutanto. Modeling gap effects on transition dominated by Tollmien-Schlichting instability. In *AIAA Aviation Forum 2020*, 2020.

[36] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communication on Pure and Applied Mathematics*, 57 :1413–1457, 2004.

[37] L. Debreu, E. Neveu, E. Simon, F.-X. Le Dimet, and A. Vidard. Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems. *Quarterly Journal of the Royal Meteorological Society*, 142 :515–528, 2016.

[38] G. Desroziers, E. Argobast, and L. Berre. Improving spatial localization in 4DEnVAR. *Quartely Journal of the Royal Meteorological Society*, 142 :3171–3185, 2016.

[39] F.-X. Le Dimet. *A general formalism of variational analysis.* CIMMS Report, 1982.

[40] F.-X. Le Dimet and V. Shutyaev. On deterministic error analysis in variational data assimilation. *Nonlinear Processes in Geophysics*, 12 :481–490, 2005.

[41] F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations : Theorotical aspects. *Tellus*, 38A :97–110, 1986.

[42] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM Journal on Numerical Analysis*, 33 :1106–1124, 1996.

[43] M. Doron, P. Brasseur, and J.-M. Brankart. Estimation of biogeochemical parameters of a 3D ocean coupled physical-biogeochemical model with a stochastic data assimilation method : twin experiments. *Journal of Marine Systems*, 87 :194–207, 2011.

[44] P. Drineas and M.W. Mahoney. On the Nyström method for approximating a Gram matrix for improved kernel-based learning. *Journal of Machine Learning Research*, 6 :2153–2175, 2005.

[45] X. Du, E. Haber, M. Karampataki, and D.B. Szyld. Varying iteration accuracy using inexact conjugate gradients in control problems governed by PDE's. In *Proceedings of the 2nd Annual International Conference on Computational Mathematics, Computational Geometry and Statistics (CMCGS 2013).*, pages 29–38, Singapore, 2013. Global and Technology Forum.

[46] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysiscal Research*, 99-C5 :10143–10182, 1994.

[47] G. Evensen. The ensemble Kalman filter : theoritical formulation and practical implementation. *Ocean Dynamics*, 53 :343–367, 2003.

[48] G. Evensen. *Data assimilation - The Ensemble Kalman Filter, 2nd Edition.* Springer, 2009.

[49] G. Evensen, F.C. Vossepoel, and P.J. van Leeuween. *Data Assimilation - A Unified Formulation of the State and Parameter Estimation Problem.* Springer, 2022.

[50] A. Farchi and M. Bocquet. On the efficiency of covariance localisation of the Ensemble Kalman Filter using augmented ensembles. *Frontiers in Applied Mathematics and Statistics*, 5 :3, 2019.

[51] A. Farchi, M. Bocquet, P. Laloyaux, M. Bonavita, and Q. Malartic. A comparison of combined data assimilation and machine learning methods for offline and online model error correction. *Journal of Computational Science*, 55 :101468, 2021.

[52] A. Farchi, P. Laloyaux, M. Bonavita, and M. Bocquet. Using machine learning to correct model error in data assimilation and forecast applications. *Quartely Journal of the Royal Meteorological Society*, 147 :3067–3084, 2021.

[53] M. Fisher, S. Gratton, S. Gürol, Y. Trémolet, and X. Vasseur. Low rank updates in preconditioning the saddle point systems arising from data assimilation problems. *Optimization Methods and Software*, 33 :45–69, 2018.

[54] M. Fisher and S. Gürol. Parallelisation in the time dimension of four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143 :1136–1147, 2017.

[55] M. Fisher, Y. Trémolet, H. Auvinen, D. Tan, and P. Poli. *Weak-constrained and long window 4D-Var.* Technical Report 655 - ECMWF, 2011.

[56] K. Fossum, T. Mannseth, and S. Stordal. Assessment of multilevel ensemble-based data assimilation for reservoir history matching. *Computational Geosciences*, 24 :217–239, 2020.

[57] M.A. Freitag, N.K. Nichols, and C.J. Budd. Resolution of sharp fronts in the presence of model error in variational data assimilation. *Quartely journal of the royal meteorological society*, 139 :742–757, 2013.

[58] M. Gander. Schwarz methods over the course of time. *Electronic Transactions on Numerical Analysis*, 31 :228–255, 2008.

[59] M. Gehlen, R. Barciela, L. Bertino, P. Brasseur, M. Butenschön, F. Chai, A. Crise, Y. Drillet, D. Ford, D. Lavoie, P. Lehodey, C. Perruche, A. Samuelsen, and E. Simon. Building the capacity for forecasting marine biogeochemistry and ecosystems : recent advances and future developments. *Journal of Operational Oceanography*, 8(S1) :s168–s187, 2015.

[60] A. Gelman. Method of moments using Monte Carlo simulation. *Journal of Computational and Graphical Statistics*, 4 (1) :36–54, 1995.

[61] M.E. Gharamti, A. Samuelsen, L. Bertino, E. Simon, A. Korosov, and U. Daewel. Online tuning of ocean biogeochemical parameters using ensemble estimation techniques : Application to a one-dimensional model in the North Atlantic. *Journal of Marine Systems*, 168 :1–16, 2017.

[62] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56 :607–617, 2008.

[63] M.B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24 :259–328, 2015.

[64] G.A. Gottwald and S. Reich. Supervised learning from noisy observations : Combining machine-learning techniques with data assimilation. *Physica D : Nonlinear Phenomena*, 423 :132911, 2021.

[65] S. Gratton, S. Gürol, E. Simon, and Ph.L. Toint. Issues in making the weakly-constrained 4DVar formulation computationally efficients. *Oberwolfach Reports*, 13 :2726–2731, 2016.

[66] S. Gratton, S. Gürol, E. Simon, and Ph.L. Toint. Guaranteeing the convergence of the saddle formulation for weakly-constrained 4d-var data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144 :2592–2602, 2018.

[67] S. Gratton, S. Gürol, E. Simon, and Ph.L. Toint. A note on preconditioning weighted linear squares, with consequences for weakly-constrained variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144 :934–940, 2018.

[68] S. Gratton, A.S. Lawless, and N.K. Nichols. Approximate Gauss-Newton methods for non-linear least squares problems. *SIAM Journal on Optimization*, 18 :106–132, 2007.

[69] S. Gratton, M. Rincon-Camacho, E. Simon, and Ph.L. Toint. Observation thinning in data assimilation computations. *EURO Journal of Computational Optimization*, 3 :31–51, 2015.

[70] S. Gratton, E. Simon, and D. Titley-Peloquin. Covariance matrix estimation for ensemble-based Kalman filters with multiple ensembles. *Mathematical Geosciences*, accepté.

[71] S. Gratton, E. Simon, D. Titley-Peloquin, and Ph.L. Toint. Minimizing convex quadratics with variable precision conjugate gradients. *Numerical Linear Algebra with Applications*, 28 :e2337, 2021.

[72] S. Gratton, E. Simon, D. Titley-Peloquin, and Ph.L. Toint. A note on inexact inner products in GMRES. *SIAM Journal on Matrix Analysis and Applications*, 43(3), 2022.

[73] S. Gratton, E. Simon, and Ph.L. Toint. An algorithm for the minimization of nonsmooth and nonconvex functions using inexact evaluations and its worst-case complexity. *Mathematical Programming*, 187 :1–24, 2021.

[74] S. Gratton, Ph. L. Toint, and J. Tshimanga. Range-space variants and inexact matrix-vector products in Krylov solvers for linear systems arising from inverse problems. *SIAM Journal on Matrix Analysis*, 32(3) :969–986, 2011.

[75] S. Gratton and J. Tshimanga. An observation-space formulation of variational assimilation using a restricted preconditioned conjugates gradient algorithm. *Quartely journal of the royal meteorological society*, 135 :1573–1585, 2009.

[76] W. Hackbusch. *Multigrid Methods and Applications.* Springer, 1985.

[77] P. Hansen. *Rank-deficient and discrete ill-posed problem : numerical aspect of linear inversion.* SIAM, 1987.

[78] P. C. Hansen. Regularization tools version 4.0 for Matlab 7.3. *International Journal for Numerical Methods in Fluids*, 46 :189–194, 2007.

[79] L. Hascoët, R.-M. Greborio, and V. Pascual. Computing adjoints by automatic differentiation with tapenade. In B. Sportisse and F.-X. LeDimet, editors, *Ecole INRIA-CEA-EDF "Problemes non-lineaires appliques"*. Springer, 2005. to appear.

[80] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of the National Bureau of Standards*, 49 :409–436, 1952.

[81] N. Higham. *Accuracy and Stability of Numerical Algorithms.* SIAM, 2002.

[82] H. Hoel, K. Law, and R. Tempone. Multilevel ensemble Kalman filtering. *SIAM Journal on Numerical Analysis*, 54 :1813–1839, 2016.

[83] T. Hoffmann, B. Schölkopf, and A.J. Smola. Kernel methods in machine learning. *The Annals of Statistics*, 36(3) :1171–1220, 2008.

[84] I. Hoteit, D.-T. Pham, and J. Blum. A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical Pacific. *Journal of Marine Systems*, 36 :101–127, 2002.

[85] B. Hug, E. Mémin, and G. Tissot. Ensemble forecast in reproducing kernel Hilbert space family : dynamical systems in Wonderland. *arXiv :2207.14653*, 2022.

[86] B.R. Hunt, E.J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos : a Local Transform Kalman Filter. *Physica D*, 230 :112–126, 2007.

[87] K. Ide, P. Courtier, M. Ghil, and A.C. Lorenc. Unified notation for data assimilation : Operational, sequential and variational. *Journal of the Meteorological Society of Japan*, 75, 1997.

[88] A.H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, New York, 1970.

[89] R. Kalman. A new approach to linear filtering and prediction problems. *Journal of Physical Oceanography*, 23 :2541–2566, 1960.

[90] P. Kumar, C. Rodrigo, F.J. Gaspar, and C.W. Oosterlee. On local Fourier analysis of multrigid methods for PDEs with jumping and random coefficients. *SIAM Journal on Scientific Computing*, 41 (3) :A1385–A1413, 2019.

[91] S. Kumar, M. Mohri, and A. Talwalkar. Sampling methods for the Nyström method. *Journal of Machine Learning Research*, 13 :981–1006, 2012.

[92] F. Lemarié. *Algorithmes de Schwarz et couplage océan-atmosphère.* PhD thesis, Université Joseph Fourier - Grenoble I, 2008.

[93] R.M Lewis and S.G. Nash. Model problems for the multigrid optimization of systems governed by differential equations. *SIAM Journal on Scientific Computing*, 26 :1811–1837, 2005.

[94] J.L. Lions. *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles.* Dunod, Paris, 1968.

[95] A. Longhurst. Seasonal cycles of pelagic production and consumption. *Progress in Oceanography*, 36 :77–167, 1995.

[96] E.N. Lorenz. Determinisitic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20 :130–141, 1963.

[97] E.N. Lorenz and K.A. Emanuel. Optimal sites for supplementary weather observations : Simulation with a small model. *Journal of the Atmospheric Sciences*, 55 :399–414, 1998.

[98] X. Luo. Ensemble-based kernel learning for a class of data assimilation problems with imperfect forward simulators. *PLoS ONE*, 14(7) :e0219247, 2019.

[99] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability"*, volume 14, pages 281–287, 1967.

[100] D. Mottet. *Raffinement adaptatif du processus d'assimilation de données par méthodes de Kalman d'ensemble pour des problèmes non-linéaires.* PhD thesis, Université de Toulouse : Toulouse INP, 2021.

[101] S. Mouysset, J. Noailles, D. Ruiz, and C. Tauber. Spectral clustering : Interpretation and Gaussian parameter. *Data Analysis, Machine Learning and Knowledge Discovery*, pages 153–162, 2014.

[102] C. Musco and C. Musco. Recursive sampling for the nystrom method. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[103] P. Mycek and M. De Lozzo. Multilevel Monte Carlo covariance estimation for the computation of Sobol' indices. *SIAM/ASA Journal of Uncertainty Quantification*, 7(4) :1323–1348, 2019.

[104] S. Nadarajan. A generalized Normal distribution. *Journal of Applied Statistic*, 32 :685–694, 2005.

[105] S.G. Nash. Preconditioning of truncated-Newton methods. *SIAM Journal on Scientific and Statistical Computing*, 6 :599–616, 1985.

[106] S.G. Nash. A multigrid approach to discretized otpimization problems. *Journal of Optimization Methods and Software*, 14 :99–116, 2000.

[107] E. Neveu. *Multigrilles pour l'assimilation variationnelle.* Rapport de stage m2r mathématiques appliquées, Université Joseph Fourier-Grenoble I, 2007.

[108] A.Y. Ng, M.I. jordan, and Y. Weiss. On spectral clustering : Analysis and an algorithm. *Advances in neural information processing systems*, 2 :849–856, 2002.

[109] T.H. Nguyen, S. Ricci, A. Piacentini, E. Simon, R. Rodriguez-Suquet, and S. Peña-Luque. Gaussian anamorphosis for ensemble kalman filter analysis of sar-derived wet surface ratio observations. soumis.

[110] J. Nocedal and S.J. Wright. *Numerical Optimization.* Springer, 2006.

[111] K.J. Nurmela. Constructing spherical codes by global optimization methods, 1995.

[112] J.S. Pelc, E. Simon, L. Bertino, G. El Serafy, and A. Heemink. Application of model reduced 4D-Var to a 1D ecosystem model. *Ocean Modelling*, 57-58 :43–58, 2012.

[113] R. Pellerej. *Etude et développement d'algorithmes d'assimilation de données variationnelles adaptés aux modèles couplés océan-atmosphère.* PhD thesis, Université Grenoble Alpes, 2018.

[114] R. Pellerej, P.A. Vidard, and F. Lemarié. Toward variational data assimilation for coupled models : first experiments on a diffusion problem. In *CARI 2016*, 2016.

[115] M. Peyron, A. Fillion, S. Gürol, V. Marchais, S. Gratton, P. Boudier, and G. Goret. Latent space data assimilation by using deep learning. *Quartely Journal of the Royal Meteorological Society*, 147 :3759–3777, 2021.

[116] D.T. Pham, J. Verron, and M.-C. Roubaud. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16,3-4 :323–340, 1998.

[117] S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation.* Cambridge University Press, 2015.

[118] A. Rouviere, L. Pascal, F. Méry, E. Simon, and S. Gratton. Neural prediction model for transition onset of a boundary layer in presence of 2D surface defect. In *AIAA SCITECH forum*, San Diego, 2022.

[119] A. Rouviere, L. Pascal, F. Méry, E. Simon, and S. Gratton. Neural prediction model for transition onset of a boundary layer in presence of 2D surface defect. *Flow*, soumis.

[120] Y. Saad and M.H. Schultz. GMRES : a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7 :856–869, 1986.

[121] P. Sakov, F. Counillon, L. Bertino, K.A. Lisæter, and A. Korablev. TOPAZ4 : an ocean-sea ice data assimilation system for the North Atlantic. *Ocean Science*, 8 :633–656, 2012.

[122] P. Sakov, G. Evensen, and L. Bertino. An iterative EnKF for strongly nonlinear systems. *Monthly Weather Review*, 140 (6) :1988–2004, 2012.

[123] P. Sakov and P. Oke. A deterministic formulation of the ensemble Kalman filter : an alternative to ensemble square root filters. *Tellus A*, 60 (2) :361–371, 2008.

[124] Y. Sasaki. A fundamental study of the numerical prediction based on the variational principle. *Journal of the Meteorological Society of Japan*, 33 :262–265, 1955.

[125] E. Simon and L. Bertino. Application of the Gaussian anamorphosis to assimilation in a 3D coupled physical-ecosystem model of the North Atlantic with the EnKF : a twin experiment. *Ocean Science*, 5 :495–510, 2009.

[126] E. Simon and L. Bertino. Joint state-parameter estimation in a 3D coupled physical-ecosystem model of the North Atlantic : assimilation of SeaWiFS data with a non-Gaussian extension of an ESRF. In *Proceedings of ESA Living Planet Symposium*. ESA, 2010.

[127] E. Simon and L. Bertino. Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation : application to a 1D ocean ecosystem model. *Journal of Marine Systems*, 89 :1–18, 2012.

[128] E. Simon, A. Samuelsen, L. Bertino, and D. Dumont. Estimation of positive sum-to-one constrained zooplankton grazing preferences with the DEnKF : a twin experiment. *Ocean Science*, 8 :587–602, 2012.

[129] E. Simon, A. Samuelsen, L. Bertino, and S. Mouysset. Experiences in multiyear combined state-parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using the Ensemble Kalman Filter. *Journal of Marine Systems*, 152 :1–17, 2015.

[130] V. Simoncini and D.B. Szyld. Theory of inexact Krylov subspace methods and applications to scientific computing. *SIAM Journal on Scientific Computing*, 25(2) :454–477, 2003.

[131] V. Simoncini and D.B. Szyld. Recent computational developments in Krylov subspace methods for linear systems. *Numerical Linear Algebra with Applications*, 14 :1–59, 2007.

[132] A. Smith and N. Gamberoni. *Transition, Pressure Gradient and Stability Theory*. Douglas Aircraft Company, 1956.

[133] S. Ta'asan. *Multigrid one-shot methods and design strategy*. Lecture note, Carnegie Mellon University, 2001.

[134] Y. Trémolet. Accounting for an imperfect model in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 132 :2483–2504, 2006.

[135] Y. Trémolet. Model error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 133 :1267–1280, 2006.

[136] U. Trottenberg, C. Oosterlee, and A. Schüller. *Multigrid*. Academic Press, 2001.

[137] J. van den Eshof and G. L. G. Sleijpen. Inexact Krylov subspace methods for linear systems. *SIAM Journal on Matrix Analysis and Application*, 26(1) :125–153, 2004.

[138] J. van den Eshof, G. L. G. Sleijpen, and M. B. van Gijzen. Relaxation strategies for nested Krylov methods. *Journal of Computational and Applied Mathematics*, 177 :347–365, 2005.

[139] J. van Ingen. *A suggested semi-empirical method for the calculation of the boundary layer transition region.* Technische Hogeschool Delft, Vliegtuigbouwkunde, 1956.

[140] P. Vermeulen and A. Heemink. Model-reduced variational data assimilation. *Monthly Weather Review*, 134 :2888–2899, 2006.

[141] A. Vidard. *Vers une prise en compte des erreurs modèle en assimilation de données 4D-variationnelle. Application à un modèle réaliste d'océan.* PhD thesis, Université Joseph Fourier-Grenoble I, 2001.

[142] A. Vidard, A. Piacentini, and F.-X. Le Dimet. Variational data analysis with control of the forecast bias. *Tellus*, 56A :117–188, 2004.

[143] Y. Wang, I.M Navon, X. Wang, and Y Cheng. 2D Burgers equation with large Reynolds number using POD/DEIM and calibration. *International Journal for Numerical Methods in Fluids*, 82(12) :909–931, 2016.

[144] A.T. Weaver and P. Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *Quartely Journal of the Royal Meteorological Society*, 127 :1815–1846, 2001.

[145] Y. Yuan. Conditions for convergence of trust region algorithms for nonsmooth optimization. *Mathematical Programming*, 31 :220–228, 1985.

[146] L. Zanna and T. Bolton. Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, 11 :376–399, 2019.

[147] L. Zanna and T. Bolton. Data-driven equation discovery of ocean mesoscale closure. *Geophysical Research Letter*, 47 :e2020GL088376, 2021.

[148] S. Zhang, Z. Liu, X. Zhang X. Wu, G. Han, Y. Zhao, X. Yu, C. Liu, Y. Liu, S. Wu, F. Lu, M. Li, and X. Deng. Coupled data assimilation and parameter estimation in coupled ocean–atmosphere models : a review. *Climate Dynamics*, 54 :5127–5144, 2020.

[149] Z. Zhang, R. Archibald, and F. Bao. A PDE-based adaptive kernel method for solving optimal filtering problems. *arXiv :2203.05031*, 2022.

[150] D. Zupanski. A general weak constraint applicable to operational 4DVAR data assimilation systems. *Monthly Weather Review*, 125, 1997.